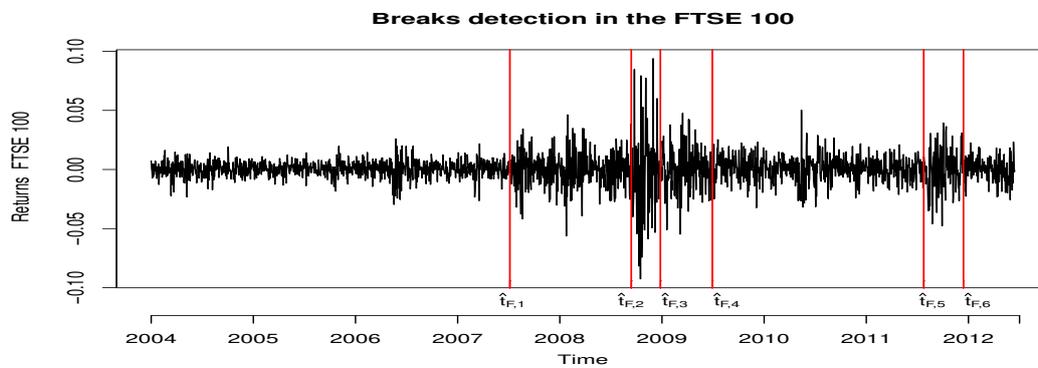


Ecole CIMPA Togo 2018

Plan détaillé du cours “Time Series”

JEAN-MARC BARDET (UNIVERSITÉ PARIS 1, SAMM)



OUTLINE

- (1) Times series: definitions and properties
- (2) Parametric and non-parametric estimation of additive trends
- (3) Common examples of time series: ARMA, GARCH, long memory, ...
- (4) Semi-parametric estimation and tests for stationary time series
- (5) Model selection for time series
- (6) Prediction for times series

REFERENCES

- [1] Amemiya, T. (1985). *Advanced Econometrics*. Cambridge , MA: Harvard University Press.
- [2] Azencott, R. et Dacunha-Castelle, D. (1984) *Séries d'observation irrégulières*. Masson, Paris.
- [3] Barbe P. et Ledoux M. (1998) *Probabilité*. EDP Sciences.
- [4] Brockwell P.J. et Davis R.A. (1991) *Time Series: Theory and Methods*. Wiley.
- [5] Brockwell P.J. et Davis R.A. (2002) *Introduction to Time-Series and Forecasting*. SpringerVerlag.
- [6] Dacunha-Castelle, D. et Duflo, M. (1983) *Probabilités et statistiques. Tome 1: Problèmes à temps fixe et Tome 2: Problèmes à temps mobile*. Collection Mathématiques Appliquées pour la Maîtrise, Masson.
- [7] Gourieroux, C. et Montfort, A. *Séries temporelles et modèles dynamiques*. Economica.
- [8] Hamilton, J.D. (1994). *Time series analysis*. Princeton University Press, Princeton.

Documents accessibles librement sur internet.

- Cours de Paul Doukhan à l'ENSAE: <http://samos.univ-paris1.fr/Teaching>.
- Cours de Xavier Guyon pour STAFVAV: <http://www.stafav.org>.
- Aide-mémoire en économétrie de A. Trognon et J.M. Fournier à l'ENSAE: <http://www.ensae.fr/ParisTech/SEC02/ENSAEEconometrieCursusintegre2006.pdf>.
- Cours de R. Bourdonnais: http://www.dauphine.fr/eurisco/eur-wp/CoursSeriesTemp-Chap*.pdf où on peut remplacer * par 1, 2, 3 ou 4.

QUELQUES SITES INTERNET INTÉRESSANTS

- Le site de Toulouse III: <http://www.lsp.ups-tlse.fr>. Regarder les documents pédagogiques.
- Le site de Paris V: <http://www.math-info.univ-paris5.fr>. Regarder les documents pédagogiques.
- Le site de Paris VI: <http://www.proba.jussieu.fr>. Regarder les documents pédagogiques.
- Le site de la S.M.A.I.: <http://smi.math.fr>. Regarder la rubrique Logiciels dans laquelle de nombreux logiciels de mathématiques peuvent être téléchargés (en particulier, Scilab et Mupad).
- Le site français d'où l'on peut télécharger le logiciel R: <http://cran.cict.fr>.

INTRODUCTION

Exemple.

Processus ou séries chronologiques financières,...

Objectifs. *Les objectifs de ce cours sont:*

- (1) *décrire une série chronologique en général.*
- (2) *modéliser une série financière et tester les modèles proposés.*
- (3) *comprendre et expliquer des phénomènes attendant à une telle série.*
- (4) *prévoir le comportement futur d'une série financière.*

1. PROCESSUS ALÉATOIRES: PREMIÈRES DÉFINITIONS ET PROPRIÉTÉS

Nous allons voir en premier lieu qu'un processus diffère de ce que l'on a jusqu'alors essentiellement rencontré en probabilités et statistiques, c'est-à-dire des suites de v.a.i.i.d. C'est d'une part l'hypothèse d'être identiquement distribuées sur laquelle nous allons revenir mais surtout celle d'indépendance, en proposant des formes de dépendances que l'on pourra caractériser par les covariances, les probabilités conditionnelles,...

1.1. Processus aléatoire.

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

- *On dit que $X = (X_t, t \in T)$ est un processus aléatoire (ou encore stochastique) sur T à valeurs dans \mathbb{R}^k lorsque pour tout $t \in T$, X_t est une variable aléatoire sur (Ω, \mathcal{A}) à valeurs dans \mathbb{R}^k .*
- *Pour $\omega \in \Omega$, $(X_t(\omega), t \in T)$ est appelé une trajectoire du processus X .*
- *On dit que $X = (X_t, t \in T)$ est un processus aléatoire du second ordre lorsque pour tout $t \in T$, X_t est une variable aléatoire appartenant à $\mathbb{L}^2(\Omega, \mathcal{A}, P)$.*
- *On appelle fonction espérance, variance, covariance et corrélation d'un processus aléatoire du second ordre à valeurs réelles, pour $(s, t) \in T^2$, les fonctions $m(t) = \mathbb{E}X_t$, $\sigma^2(t) = \mathbb{E}X_t^2 - m^2(t)$, $\gamma(s, t) = \mathbb{E}(X_s - \mathbb{E}X_s)(X_t - \mathbb{E}X_t)$ et $r(s, t) = \gamma(s, t)/(\sigma(s)\sigma(t))$.*

Exemple.

Fonction réelle; Suite de variables indépendantes; Marche aléatoire; Chaîne de Markov; Mouvement brownien.

Définition. *Une série chronologique (temporelle) est un processus aléatoire réel indicé par \mathbb{Z} ou \mathbb{N} (on dit encore processus à temps discret; quand ce n'est pas le cas, notamment lorsque $T = \mathbb{R}$, on parle de processus à temps continu).*

Définition. *On appelle processus aléatoire (ou une série chronologique) $X = (X_t, t \in T)$ gaussien, un processus aléatoire tel que $\forall k \in \mathbb{N}^*$, $\forall (t_1, \dots, t_n) \in T^n$, $(X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien.*

Remarque.

Rappelons que $(X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien si et seulement si $\forall (u_1, \dots, u_n) \in \mathbb{R}^n$, $u_1 X_{t_1} + \dots + u_n X_{t_n}$ est une variable gaussienne.

Propriété. • *Si X et Y sont des vecteurs gaussiens, alors (Indépendance de X et $Y \iff cov(X, Y) = 0$).*

- *Un processus gaussien est entièrement défini par ses fonctions espérance $m(t) = \mathbb{E}X(t)$ et covariance $\gamma(s, t) = \mathbb{E}X_s X_t$. Réciproquement, la connaissance d'une fonction $\gamma(s, t)$ définie positive et d'une fonction $m(t)$, définit un unique processus gaussien.*

Définition. • *Un bruit blanc (fort) est une suite de variables aléatoires identiquement distribuées indépendantes (v.a.i.i.d.) centrées.*

- Un bruit blanc faible est une suite de variables aléatoires identiquement distribuées centrées non corrélées.
- Un bruit blanc gaussien est une suite de v.a.i.i.d. gaussiennes centrées.

Remarque.

La terminologie de bruit blanc est surtout employée par les spécialistes de traitements du signal. Ceux-ci évoquent aussi parfois des bruits roses pour certaines formes de dépendances entre les données...

1.2. Stationnarité.

Définition. On dit qu'un processus aléatoire $X = (X_t, t \in T)$ est (strictement) stationnaire lorsque X est invariant en distribution par toute translation du temps, c'est-à-dire que $\forall n \in \mathbb{N}^*$, $\forall (t_1, \dots, t_n) \in T^n$, $\forall c \in T$, $(X_{t_1}, \dots, X_{t_n})$ à la même distribution que $(X_{t_1+c}, \dots, X_{t_n+c})$.

Remarque.

Une caractérisation de l'égalité en loi est celle obtenue par la fonction caractéristique. On montrera ainsi que pour tout $(u_1, \dots, u_n) \in \mathbb{R}^n$,

$$\phi_{(X_{t_1}, \dots, X_{t_n})}(u_1, \dots, u_n) = \mathbb{E} \left(e^{i \sum_{j=1}^n u_j X_{t_j}} \right) = \phi_{(X_{t_1+c}, \dots, X_{t_n+c})}(u_1, \dots, u_n).$$

On pourra ainsi utiliser le fait que la fonction caractéristique de la somme de 2 v.a. indépendantes vaut le produit des fonctions caractéristiques...

Exemple.

Suite de v.a.i.i.d., chaîne de Markov homogène.

Propriété. Conséquences de la stationnarité sur les fonctions espérance, variance, covariance d'un processus à temps discret X :

- L'espérance $m(t) = \mathbb{E}(X_t)$ est constante.
- La variance $\sigma^2(t) = \text{var}(X_t)$ est constante.
- La covariance $\gamma(s, t) = \text{cov}(X_s, X_t)$ est une fonction ne dépendant que de $|t - s|$.

Définition. Soit un processus à temps discret $X = (X_t, t \in T)$. On dit que X est:

- Un processus stationnaire d'ordre 2 lorsque: 1/ son espérance $m(t)$ est constante, 2/ sa covariance $\text{cov}(X_s, X_t)$ est une fonction de $|t - s|$.
- Un processus à accroissements stationnaires lorsque le processus $Y = \{Y_t, t \in T\}$ telle que $Y_t = X_{t+1} - X_t$ pour $t \in T$, est stationnaire.

Propriété. • (Stationnarité stricte \implies Stationnarité d'ordre 2), mais la réciproque est fausse.

- Si X est un processus gaussien, alors (Stationnarité stricte \iff Stationnarité d'ordre 2).

Remarque.

On pourrait penser qu'il est plus difficile d'être stationnaire strict que stationnaire d'ordre 2. Cependant la stationnarité d'ordre 2 requiert d'avoir des moments d'ordre 2 ce qui n'est pas demandé par la stationnarité stricte. On peut ainsi montrer que pour un ARCH(p) (on verra plus loin la définition d'un tel processus), les conditions de stationnarité d'ordre 2 sont plus fortes que celles de stationnarité stricte.

1.3. Cas particulier des processus stationnaires. Dans toute la suite, on suppose que $X = (X_k)_{k \in \mathbb{Z}}$ est un processus à temps discret **stationnaire** (donc un processus sans tendance additive ou multiplicative non constantes). Ceci induit en particulier que les X_n sont des variables identiquement distribuées. On considérera également que les processus sont centrés, ce qui s'adapte par simple translation au cas non centré.

Définition. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus du second ordre à temps discret centré stationnaire,

- on appelle $r(k) = \mathbb{E} X_0 X_k = \mathbb{E} X_i X_{i+k}$ pour $k \in \mathbb{Z}$ et $i \in \mathbb{Z}$, la covariance de X . Ainsi, $r(0)$ est la variance de la série.
- on appelle $\rho(k) = r(k)/r(0)$ la corrélation de X . On a $-1 \leq \rho(k) \leq 1$ pour $k \in \mathbb{Z}$.

Définition. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus du second ordre à temps discret centré stationnaire. S'il existe une fonction $f : [-\pi, \pi[\rightarrow \mathbb{C}$ telle que $\forall k \in \mathbb{Z}$, $r(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$, alors on dit que X admet une densité spectrale f .

Exemple.

Montrer que la densité spectrale d'un bruit blanc faible stationnaire à variance finie existe et la calculer.

Propriété. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus du second ordre à temps discret centré stationnaire.

Si la densité spectrale f existe sur $[-\pi, \pi[$ alors f est paire et $f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} r(k) e^{-ik\lambda}$ pour tout $\lambda \in [-\pi, \pi[$.

Propriété. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus du second ordre à temps discret centré stationnaire.

- (1) (Les covariances $r(k)$ vérifient $\sum |r(k)|^2 < \infty$) \iff (f existe et est de carré intégrable sur $[-\pi, \pi[$).
- (2) (Les covariances $r(k)$ sont telles que $\sum |r(k)| < \infty$) \implies (f existe et f continue sur $[-\pi, \pi[$).

Exemple.

Soit $(\varepsilon_k)_{k \in \mathbb{N}}$ une suite de v.a.i.i.d. à variance finie et soit $X_k = \varepsilon_k - \varepsilon_{k-1}$ pour $k \in \mathbb{Z}$. Déterminer la densité spectrale de $(X_k)_{k \in \mathbb{Z}}$.

Définition. Soit $X = (X_k, k \in \mathbb{Z})$ un processus à temps discret. On appelle filtre linéaire une famille de réels $a = (a_i)_{i \in I}$, où $I \subset \mathbb{Z}$, et la série filtrée est $Y = (Y_k, k \in \mathbb{Z})$ telle que $Y_k = \sum_{i \in I} a_i X_{k-i}$.

Exemple.

Filtres des accroissements, moyenne empirique sur des fenêtres mobiles,...

Proposition. Avec les notations de la définition précédente,

- (1) Si I est fini, alors Y dans \mathbb{L}^p lorsque X existe dans \mathbb{L}^p (avec $p > 0$). De plus si X est stationnaire strict (respectivement du second ordre) alors Y est stationnaire strict (respectivement du second ordre).
- (2) Si I est infini, alors la condition $\sum_{i \in I} |a_i| < +\infty$ et $\sup_{t \in \mathbb{Z}} \mathbb{E} |X_t| < \infty$ garantit que Y existe p.s. De plus si X est stationnaire strict alors Y est stationnaire strict. Sous cette condition, si X est gaussien alors Y est gaussien, si X est centré alors Y est centré.
- (3) Si I est infini, alors la condition $\sum_{i \in I} |a_i| < +\infty$ et $\sup_{t \in \mathbb{Z}} \mathbb{E} |X_t|^2 < \infty$ garantit que Y existe dans L^2 . De plus si X est stationnaire strict (respectivement du second ordre) alors Y est stationnaire strict (respectivement du second ordre).
- (4) Si I est infini et X est un bruit blanc fort (respectivement faible) et $\sum_{i \in I} |a_i|^2 < +\infty$ alors Y existe p.s. et dans L^2 , est appelé un processus linéaire et Y est stationnaire strict (respectivement du second ordre).

Proof. (1) L'existence ne pose pas de problème. Pour la stationarité, en supposant $I = -m, \dots, m$, comme X est stationnaire, on sait que pour tout $n \in \mathbb{N}^*$, pour tout t_1, \dots, t_n dans \mathbb{Z} , alors pour tout $c \in \mathbb{Z}$, $(X_{t_1-m}, X_{t_1-m+1}, \dots, X_{t_1+m}, X_{t_2-m}, \dots, X_{t_n-m}, \dots, X_{t_n+m})$

a même loi que $(X_{t_1-m+c}, X_{t_1-m+1+c}, \dots, X_{t_1+m+c}, X_{t_2-m+c}, \dots, X_{t_n+m+c})$. Maintenant, en considérant la fonction $g : \mathbb{R}^{(2m+1)k} \rightarrow \mathbb{R}^k$ telle que $g(X_{t_1-m}, X_{t_1-m+1}, \dots, X_{t_1+m}, X_{t_2-m}, \dots, X_{t_n+m}) = (\sum_{i=-m}^m a_i X_{t_1-i}, \dots, \sum_{i=-m}^m a_i X_{t_n-i})$, g étant continue donc mesurable, on montre bien que $g(X_{t_1-m}, X_{t_1-m+1}, \dots, X_{t_1+m}, X_{t_2-m}, \dots, X_{t_n+m})$ à la même loi que $g(X_{t_1-m+c}, X_{t_1-m+1+c}, \dots, X_{t_1+m+c}, X_{t_2-m+c}, \dots, X_{t_n+m+c})$ donc $(Y_{t_1}, \dots, Y_{t_n})$ a la même loi que $(Y_{t_1+c}, \dots, Y_{t_n+c})$: (Y_t) est bien stationnaire.

Si X est stationnaire d'ordre 2, il est facile voir que $\mathbb{E} Y_t$ est constante. On a $\text{cov}(Y_s, Y_t) = \sum_{i \in I} \sum_{i' \in I} a_i a_{i'} \text{cov}(X_{t-i}, X_{s-i'}) = \sum_{i \in I} \sum_{i' \in I} a_i a_{i'} \gamma(t-s-i+i')$ en notant $\gamma(k) = \text{cov}(X_0, X_k)$. Donc $\text{cov}(Y_s, Y_t)$ est bien une fonction de $t-s$. De plus, on peut intervertir i et i' et du fait de la parité de γ on voit bien que $\text{cov}(Y_s, Y_t)$ est une fonction de $|t-s|$.

(2) On a $\mathbb{E} |Y_t| \leq \mathbb{E} (\sum_{i \in I} |a_i| |X_{t-i}|) \leq (\sum_{i \in I} |a_i|) \sup_{j \in \mathbb{Z}} \mathbb{E} |X_j|$ d'après le Théorème de Lebesgue, donc $\mathbb{E} |Y_t| < \infty$: (Y_t) est bien finie avec une probabilité 1.

Pour la stationarité stricte, on procède comme précédemment en considérant des restrictions $(Y_t^{(m)})_t$ qui sont bien stationnaires. Du fait de l'existence d'une limite, on a $(Y_{t_1}^{(m)}, \dots, Y_{t_n}^{(m)})$ qui tend dans L^1 vers $(Y_{t_1}, \dots, Y_{t_n})$ lorsque m tend vers l'infini, donc on a également convergence en loi. Aussi comme $(Y_{t_1}^{(m)}, \dots, Y_{t_n}^{(m)})$ a la même loi que $(Y_{t_1+c}^{(m)}, \dots, Y_{t_n+c}^{(m)})$, cette égalité a également lieu à la limite: (Y_t) est bien stationnaire.

(3) On a $\mathbb{E} (Y_t - Y_t^{(m)})^2 = \mathbb{E} (\sum_{|i|>m} \sum_{|i'|>m} a_i a_{i'} \mathbb{E} (X_{t-i} X_{t-i'})) \leq (\sum_{|i|>m} |a_i|)^2 \sup_{j \in \mathbb{Z}} \mathbb{E} |X_j|^2$ d'après le Théorème de Lebesgue, donc $\mathbb{E} (Y_t - Y_t^{(m)})^2 \rightarrow 0$ ($m \rightarrow \infty$) car $\sum_{i \in I} |a_i| < \infty$: (Y_t) existe dans L^2 .

La stationarité stricte s'obtient comme dans le cas précédent et la stationarité d'ordre 2 utilise le point (1): on a $(Y_t^{(m)})_t$ qui est stationnaire d'ordre 2 et comme on a convergence dans L^2 donc en loi, on a bien convergence de l'espérance et de la covariance des $(Y_t^{(m)})$: (Y_t) est bien aussi stationnaire d'ordre 2.

(4) Dans le cas d'un processus linéaire, on peut donc obtenir l'existence et la stationarité sous la condition $\sum_{i \in I} a_i^2 < \infty$ qui est plus faible que la condition $\sum_{i \in I} |a_i| < \infty$. En effet, on a $\mathbb{E} (Y_t - Y_t^{(m)})^2 = \mathbb{E} (\sum_{|i|>m} \sum_{|i'|>m} a_i a_{i'} \mathbb{E} (X_{t-i} X_{t-i'})) = \sum_{|i|>m} a_i^2 \mathbb{E} (X_0^2) \rightarrow 0$ ($m \rightarrow \infty$) dès que (X_t) est un bruit blanc. La preuve de la stationarité est immédiate (voir (3)), et pour la stationarité d'ordre 2, on a clairement $\text{cov}(Y_s, Y_t) = \mathbb{E} X_0^2 \sum_{i \in I} a_i a_{i+s-t}$ qui est une fonction dépendant de $|t-s|$ et qui existe (d'après Cauchy-Schwarz). \square

Propriété. Avec les notations de la définition précédente, la composée de 2 filtres linéaires est un filtre linéaire.

Grâce à la notion de stationarité il est possible d'obtenir un résultat théorique très puissant, qui n'a cependant que peu d'intérêt en pratique:

Théorème (Décomposition de Cramér-Wald). Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus à temps discret stationnaire d'ordre 2 tel que $\int_{-\pi}^{\pi} \log(f(\lambda)) d\lambda > -\infty$. Alors il existe un unique bruit blanc faible $(\varepsilon_t)_{t \in \mathbb{Z}}$ (donc $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$ pour $t \neq s$ mais les ε_t ne sont pas forcément indépendantes) et une famille de réels $(a_i)_{i \in \mathbb{N}}$ telle que $a_0 \geq 0$ et $\sum_{k \in \mathbb{N}} |a_k|^2 < \infty$, vérifiant dans L^2

$$(1) \quad X_t = \mathbb{E} X_0 + \sum_{i=0}^{\infty} a_i \varepsilon_{t-i} \quad \text{pour tout } t \in \mathbb{Z}.$$

Ceci permet d'avoir une forme un peu générale d'un processus stationnaire. Cependant nous verrons que l'on peut surtout traiter le cas des processus linéaires (quand le bruit blanc est fort, donc avec l'hypothèse d'indépendance en plus); dans le cas général, la décomposition de Cramér-Wald n'apporte pas beaucoup de renseignements (voir le cas des processus GARCH).

1.4. Tendances et composante saisonnière. Les séries de données réelles que l'on peut rencontrer ne sont que très rarement des séries que l'on peut modéliser par un processus à temps discret stationnaire. En effet, très souvent leurs moyennes varient dans le temps, parfois de manière suffisamment régulière pour être facilement modélisées (typiquement une tendance linéaire), parfois non... Voici pour commencer le type principal de "tendances" dites tendances additives:

Définition. Tout processus à temps discret $X = (X_t)_{t \in T}$ avec $T \subset \mathbb{Z}$ peut s'écrire sous la forme

$$X_t = a(t) + S(t) + \varepsilon_t, \quad \text{pour } t \in T,$$

où $t \mapsto a(t)$ et $t \mapsto S(t)$ sont deux fonctions déterministes telles que $\mathbb{E} X_t = a(t) + S(t)$, avec

$$(1) \quad \text{si } t \mapsto S(t) \text{ est une fonction périodique non nulle de période } r > 0 \text{ telle que } \sum_{i=1}^r S(i) = 0,$$

alors S est la composante saisonnière, saisonnalité de X ,

- (2) si $t \mapsto a(t)$ est non nulle, a est la tendance de X ,
 (3) si $\varepsilon = (\varepsilon_t)_{t \in T}$ est non nulle, ε est une série chronologique centrée appelée souvent le bruit de X .

Exemple.

Tendances polynômiales, saisonnalité annuelle,..

Remarque.

Attention, il n'y a pas unicité de la décomposition précédente. Il y a en revanche unicité du bruit de X puisque $u_t = X_t - \mathbb{E} X_t$ pour tout t .

Définition. Pour $X = (X_t)_t$ un processus aléatoire ayant pour tendance a et pour saisonnalité S . On appelle:

- Série détendancialisée la série $(X_t - a(t))_t$. Si la fonction $a(\cdot)$ n'est pas connue explicitement, ce qui est le plus souvent le cas, $(X_t - \hat{a}(t))_t$, où $\hat{a}(t)$ est un estimateur de $a(t)$ sera la série détendancialisée (même dénomination).
- Série désaisonnalisée (ou série corrigée des variations saisonnières) la série $(X_t - S(t))_t$. Si la fonction $S(\cdot)$ n'est pas connue explicitement, ce qui est le plus souvent le cas, $(X_t - \hat{S}(t))_t$, où $\hat{S}(t)$ est un estimateur de $S(t)$ sera la série désaisonnalisée (même dénomination).

Il est aussi possible que le processus ait une variance qui varie de manière déterministe en fonction du temps. En ce cas on évoquera une tendance multiplicative:

Définition. On dira que X possède une tendance multiplicative si $X_t - m(t) = \sigma(t) u(t)$ pour tout $t \in T$, où $u = (u_t)_{t \in T}$ est une suite de variables aléatoires centrées de variance constante et $\sigma(\cdot)$ est une fonction positive.

Estimation de la tendance et de la saisonnalité. Cette partie est souvent omise ou vite traitée dans les livres consacrés aux processus ou aux séries chronologiques. Pourtant l'estimation de la tendance et de la saisonnalité est essentielle dans la plupart des travaux concrets portant sur les séries chronologiques, en particulier parce qu'elle apporte une information souvent bien plus importante que la partie bruit en vue de prédictions. Les techniques utilisées peuvent être semi-paramétriques (typiquement régression linéaire) ou non-paramétriques (régressions locales, régressions par splines, lissages exponentiels, convolution par un noyau, décomposition dans une base d'ondelettes,...). Cependant, les séries financières de type log-ratio sur lesquelles nous allons essentiellement travailler n'ont ni tendance ni saisonnalité.

2. LES PROCESSUS ARMA, GARCH ET LONGUE MÉMOIRE

2.1. Processus ARMA.

Définition. Pour $(X_k)_{k \in \mathbb{Z}}$ un processus à temps discret, on appelle:

- opérateur retard B l'application linéaire qui à X associe $Y = (Y_n)_{n \in \mathbb{Z}} = B \cdot X$ tel que $Y_n = X_{n-1}$ pour $n \in \mathbb{Z}$ (par abus de notation $B \cdot X_n = X_{n-1}$).
- puissance B^k la k -ième composée de B par elle même qui est donc telle $B^k \cdot X_n = X_{n-k}$.
- avec le polynôme $P(B) = \sum_{j=0}^p a_j B^j$ où $(a_i)_{0 \leq i \leq p} \in \mathbb{R}^{p+1}$, $P(B) \cdot X = (\sum_{j=0}^p a_j X_{n-j})_{n \in \mathbb{Z}}$.
- B^{-1} vérifiant $B^{-1} \cdot X_n = Y_{n+1}$ est tel que $B^{-1}B = BB^{-1} = I_d$ d'où l'extension à B^k pour $k \in \mathbb{Z}$.
- si $\sum_{i=-\infty}^{\infty} |a_i| < \infty$, alors $f(B) = \sum_{i=-\infty}^{\infty} a_i B^i$ transforme X en Y tel que $Y_n = \sum_{i=-\infty}^{\infty} a_i X_{n-i}$.

Du fait de la structure linéaire des applications B , $P(B)$ ou $f(B)$, on a les propriétés suivantes:

- Propriété.**
- Additivité des fonctions de B : $f(B) \cdot (\lambda X + \lambda' X') = \lambda f(B) \cdot X + \lambda' f(B) \cdot X'$.
 - Commutativité du produit de fonctions de B : $f_1(B) f_2(B) = f_2(B) f_1(B)$.

Proposition. Soit $X = (X_n)_{n \in \mathbb{Z}}$ et $Y = (Y_n)_{n \in \mathbb{Z}} = (1 - \lambda B)X$ avec X et Y deux processus stationnaires. Alors, si

- si $|\lambda| < 1$, $X_n = (1 - \lambda B)^{-1} \cdot Y_n = \sum_{i=0}^{\infty} \lambda^i Y_{n-i}$;
- si $|\lambda| > 1$, $X_n = -\lambda^{-1} B^{-1} (1 - \lambda^{-1} B^{-1})^{-1} \cdot Y_n = -\sum_{i=1}^{\infty} \lambda^{-i} Y_{n+i}$;
- si $|\lambda| = 1$, $1 - \lambda B$ n'est pas inversible.

Conséquence.

- (1) Si on suppose que $P(B)$ est un polynôme dont les racines ne sont pas situées sur le cercle trigonométrique (donc les racines sont telles que $|z| \neq 1$), alors il existe une unique série de la forme $(P(B))^{-1} = \sum_{i=-\infty}^{\infty} a_i B^i$ avec $\sum_{i=-\infty}^{\infty} |a_i| < \infty$.
- (2) Si les racines de P sont hors du disque unité (donc les racines sont telles que $|z| > 1$) alors alors il existe une unique série de la forme $(P(B))^{-1} = \sum_{i=0}^{\infty} a_i B^i$ avec $\sum_{i=0}^{\infty} |a_i| < \infty$. On dira alors que la relation entre les deux processus est causale.

Définition. Soit P et Q deux polynômes de degrés respectifs p et q tels que P et Q soient premiers entre eux dans $\mathbb{C}[X]$. On suppose que les racines de P ne sont pas sur le cercle unité. Soit également $\varepsilon = (\varepsilon_n)_{n \in \mathbb{Z}}$ un bruit blanc (fort).

- Un processus $X = (X_n)_{n \in \mathbb{Z}}$ tel que $P(B) \cdot X = \varepsilon$ est appelé un processus AR(p). Ceci revient à écrire qu'il existe $(b_1, \dots, b_p) \in \mathbb{R}^p$ tel que:

$$X_n + b_1 X_{n-1} + \dots + b_p X_{n-p} = \varepsilon_n \text{ pour tout } n \in \mathbb{Z}.$$

- Un processus $X = (X_n)_{n \in \mathbb{Z}}$ tel que $X = Q(B) \cdot \varepsilon$ est appelé un processus MA(q). Ceci revient à écrire qu'il existe $(c_1, \dots, c_q) \in \mathbb{R}^q$ tel que:

$$X_n = \varepsilon_n + c_1 \varepsilon_{n-1} + \dots + c_q \varepsilon_{n-q} \text{ pour tout } n \in \mathbb{Z}.$$

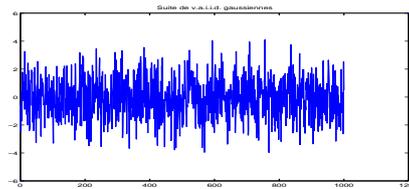
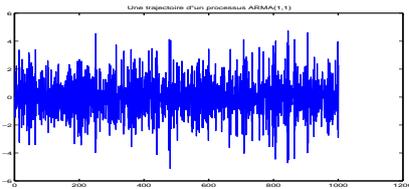
- Un processus $X = (X_n)_{n \in \mathbb{Z}}$ tel que $P(B) \cdot X = Q(B) \cdot \varepsilon$ est appelé un processus ARMA(p, q). Ceci revient à écrire qu'il existe $(b_1, \dots, b_p) \in \mathbb{R}^p$ et $(c_1, \dots, c_q) \in \mathbb{R}^q$ tels que:

$$X_n + b_1 X_{n-1} + \dots + b_p X_{n-p} = \varepsilon_n + c_1 \varepsilon_{n-1} + \dots + c_q \varepsilon_{n-q} \text{ pour tout } n \in \mathbb{Z}.$$

Exemple.

Exemples d'AR(1), de MA(1), d'ARMA(1, 1).

Voici un exemple (à gauche) d'une trajectoire du processus ARMA(1, 1) avec $b_1 = 0.6$ et $c_1 = -0.5$, à comparer avec une trajectoire (à droite) de variables gaussiennes indépendantes et identiquement distribuées:



Propriété. Si les racines de P sont à l'extérieur (strictement) du disque trigonométrique, alors un processus ARMA est stationnaire et causal. Si les racines de P sont à l'intérieur (strictement) du disque trigonométrique, le processus ARMA est stationnaire mais non causal.

Exemple.

Exemples d'ARMA stationnaires et non-stationnaires.

Propriété. Lorsque les racines de P sont à l'extérieur du disque trigonométrique, on peut établir des relations de récurrence permettant de calculer la fonction covariance d'un processus ARMA(p, q) tel que

$$P(B) \cdot X_n = X_{n+p} + b_1 X_{n+p-1} + \dots + b_p X_n = Q(B) \cdot \varepsilon_n,$$

et on obtient

$$r(n+p) + b_1 r(n+p-1) + \dots + b_p r(n) = 0 \text{ pour tout } n \geq p-q.$$

Corollaire. En particulier, on obtient les équations de Yule-Walker, vérifiant le système:

$$\begin{pmatrix} r(q) & r(q-1) & \dots & r(q-p+1) \\ r(q+1) & r(q) & \dots & r(q-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ r(q+p-1) & r(q+p-2) & \dots & r(q) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} = - \begin{pmatrix} r(q+1) \\ r(q+2) \\ \vdots \\ r(q+p) \end{pmatrix}.$$

Conséquence.

- La covariance d'un processus ARMA décroît asymptotiquement exponentiellement vite.
- S'il existe un estimateur convergent de la covariance (ou de la corrélation), l'inversion des équations de Yule-Walker permet d'obtenir des estimateurs des paramètres a_i quand une trajectoire (X_1, \dots, X_N) est connue.

Propriété. Soit $(\varepsilon_n)_{n \in \mathbb{Z}}$ un bruit blanc et soit $(a_i)_{i \in \mathbb{Z}}$ une famille de réels tels que $\sum_{i=-\infty}^{\infty} |a_i| < \infty$.

Alors le processus à temps discret $X = (X_n)_{n \in \mathbb{Z}}$ tel que $X_n = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{n-i}$ pour $n \in \mathbb{Z}$ est stationnaire et appelée processus linéaire (ou processus à moyennes mobiles infinies). Si $a_i = 0$ pour $i < 0$ alors on dira que X est un processus linéaire causal.

Propriété. Soit le processus à temps discret $X = (X_n)_{n \in \mathbb{Z}}$ tel que $X_n = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{n-i}$ pour $n \in \mathbb{Z}$,

avec $(\varepsilon_n)_{n \in \mathbb{Z}}$ un bruit blanc de variance σ^2 et $\sum_{i=-\infty}^{\infty} |a_i| < \infty$. Alors X admet une densité spectrale f telle que

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda} \right|^2, \text{ pour } \lambda \in [-\pi, \pi[.$$

Proof. Le processus X a pour autocovariance $r(n) = \mathbb{E}(X_0 X_n) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_i a_j \mathbb{E}(\varepsilon_{-i} \varepsilon_{n-j}) = \sigma^2 \sum_{j=-\infty}^{\infty} a_{j-n} a_j$ car (ε_i) est un bruit blanc. Donc $f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{n=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_{j-n} a_j e^{-in\lambda}$. Il suffit alors de poser $k = j - n$ et ainsi on obtient que $f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_k a_j e^{-i(j-k)\lambda} = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} a_k e^{ik\lambda} \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda} = \frac{\sigma^2}{2\pi} \left| \sum_{k=-\infty}^{\infty} a_k e^{-ik\lambda} \right|^2$. \square

Propriété. Soit le processus à temps discret $Y = (Y_n)_{n \in \mathbb{Z}}$ tel que $Y_n = \sum_{i=-\infty}^{\infty} a_i X_{n-i}$ pour $n \in \mathbb{Z}$, avec $X = (X_n)_{n \in \mathbb{Z}}$ un processus à temps discret stationnaire de densité spectrale f_X et

$\sum_{i=-\infty}^{\infty} |a_i| < \infty$. Alors Y est un processus à temps discret stationnaire de densité:

$$f_Y(\lambda) = f_X(\lambda) \left| \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda} \right|^2, \text{ pour } \lambda \in [-\pi, \pi[.$$

Proof. Si on note r_X l'autocovariance de X et r_Y celle de Y , on a $r_Y(n) = \mathbb{E}(Y_0 Y_n) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k r_X(n-k+j)$. Donc $f_Y(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k r_X(n-k+j) e^{-in\lambda}$. Les séries pouvant être interverties grâce au Théorème de Fubini, on a donc encore $f_Y(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_j a_k r_X(m) e^{-i(m+k-j)\lambda}$ en posant $m = n - k + j$ et ainsi $f_Y(\lambda) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j e^{ij\lambda} a_k e^{-ik\lambda} f_X(\lambda)$, soit $f_Y(\lambda) = f_X(\lambda) \left| \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda} \right|^2$. \square

Propriété. Lorsque les racines de P sont à l'extérieur du disque trigonométrique, un processus ARMA(p, q) (tel que $P(B)(X) = Q(B)(\varepsilon)$ avec (ε_n) un bruit blanc de variance σ^2) a pour densité spectrale:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{Q(e^{i\lambda})}{P(e^{i\lambda})} \right|^2, \text{ pour } \lambda \in [-\pi, \pi[.$$

Proof. Soit le processus $Z = P(B)(X)$. Alors d'après ce qui précède Z est stationnaire et $f_Z(\lambda) = f_X(\lambda) |P(e^{i\lambda})|^2$. Mais on a également $Z = Q(B)(\varepsilon)$. D'où $f_Z(\lambda) = f_\varepsilon(\lambda) |Q(e^{i\lambda})|^2 = \frac{\sigma^2}{2\pi} |Q(e^{i\lambda})|^2$. En égalisant les 2 valeurs de $f_Z(\lambda)$ on trouve $f_X(\lambda)$. \square

2.2. Les processus GARCH. Les processus ARCH(p) ont été introduits par Engle (1982) et généralisés en GARCH(p, q) par Bollersév (1986) dans le cadre de données financières où la "volatilité" (la variance) dépend conditionnellement du passé.

Définition. Soit $(\varepsilon_k)_{k \in \mathbb{Z}}$ un bruit blanc. Alors on dit que $(X_k)_{k \in \mathbb{Z}}$ est un processus GARCH(p, q), où $p, q \in \mathbb{N}$ s'il existe $(a_0, a_1, \dots, a_p, b_1, \dots, b_q)$ une famille de constantes réelles positives telles que

$$(2) \quad X_k = \sigma_k \varepsilon_k \quad \text{et} \quad \sigma_k^2 = a_0 + \sum_{j=1}^p a_j X_{k-j}^2 + \sum_{i=1}^q b_i \sigma_{k-i}^2 \text{ pour tout } k \in \mathbb{Z}.$$

On peut se demander s'il est possible de trouver des processus stationnaires qui peuvent vérifier une telle équation de récurrence. Nous commençons par étudier le cas des processus GARCH(1, 1) pour lesquels un

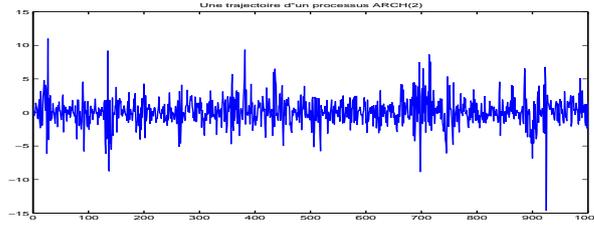
Propriété. Un processus GARCH(1, 1) vérifiant l'équation (2) est stationnaire si et seulement si $\mathbb{E} [\log(a_1 \eta_0^2 + b_1)] < 0$.

Proof. \square

Nous contenterons ici de donner une réponse de la cas de la stationnarité d'ordre 2:

Propriété. Un processus ARCH(p) vérifiant l'équation (2) est stationnaire d'ordre 2 si et seulement si $\mathbb{E} [\varepsilon_0^2] \sum_{k=1}^p a_k < 1$.

Ce résultat est très récent et nous n'en donnerons pas la preuve (en particulier la condition nécessaire n'a été montrée qu'en 2000...). Voici un exemple d'une trajectoire du processus ARCH(2) avec $a_0 = 1$, $a_1 = 0.3$ et $a_2 = 0.5$:



Propriété. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus $ARCH(p)$ stationnaire d'ordre 2. Alors X_k n'est pas indépendant de X_0 , mais $r(k) = \text{cov}(X_0, X_k) = 0$ pour tout $k \neq 0$. De plus la densité spectrale f de X est la même que celle d'un bruit blanc: $f(\lambda) = C$ pour tout $\lambda \in [-\pi, \pi]$, où $C > 0$.

On s'aperçoit donc qu'un processus $ARCH(p)$ ne pourra pas être identifié à partir de sa densité spectrale. Une conséquence de ceci est également qu'un $ARCH(p)$ ne peut pas être gaussien (car sinon ce serait un bruit blanc, ce qu'il n'est pas). On peut cependant montrer qu'en considérant $(X_k^2)_k$ au lieu de $(X_k)_k$ on se ramène à un processus ARMA:

Propriété. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus $ARCH(p)$ stationnaire d'ordre 2. Alors $Y = (X_k^2)_{k \in \mathbb{Z}}$ est un processus $ARMA(p)$.

Proof. On considère $v_t = X_t^2 - \sigma_t^2$, bruit blanc faible, et on montre que $X_t^2 = v_t + (a_0 + \sum a_i X_{t-i}^2) + \sum b_j (X_{t-j}^2 - v_{t-j})$, d'où (X_t^2) ARMA faible. \square

Corollaire. On en déduit qu'un $GARCH$ ne peut pas être gaussien.

2.3. Processus longue mémoire.

Definition (0). Let $X = (X_k)_{k \in \mathbb{Z}}$ be a second order stationary process

$$X \text{ is a Long Memory Process} \iff \sum_{k \in \mathbb{Z}} |r(k)| = \infty$$

Exemple: $X = (X_k)_{k \in \mathbb{Z}}$ with $X_k = X_0$ for all $k \in \mathbb{Z}$.

Consequences: If X LM process:

- The spectral density of X , if it exists, is not continuous;
- This definition is not really satisfying

Definition (1). $X = (X_k)_{k \in \mathbb{Z}}$ is a LM stationary second order process

$$r(k) = |k|^{-D} L(|k|), \text{ for } k \neq 0, \text{ with}$$

- $D \in]0, 1[$ LM parameter
- L slowly varying function in ∞ , i.e.: $\forall t > 0, \lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$

Counter-example: $X = (X_k)_{k \in \mathbb{Z}}$ with $X_k = X_0$ for all $k \in \mathbb{Z}$ is not LM.

Definition (2). $X = (X_k)_{k \in \mathbb{Z}}$ is a LM stationary second order process

$$f(\lambda) = |\lambda|^{D-1} M\left(\frac{1}{|\lambda|}\right), \text{ for } \lambda \rightarrow 0, \text{ with}$$

- $D \in]0, 1[$ LM parameter
- M a slowly varying function in ∞

Remarks:

- Definition (1) implies Definition (2) (Abélian Theorem);
- Definition (2) + decreasing of $r(\cdot)$ implies Definition (1) (Tauberian Theorem).

Other possible definitions [Samorodnitsky and Taqqu (1994)]

Definition. $X = (X_t)_{t \in \mathbb{R}}$ H -self similar process with stationary increments (H -SSSI)

$$\begin{cases} (X_{cs})_s \stackrel{\mathcal{L}}{\sim} c^H (X_s)_s \text{ for any } c > 0 \\ (X_{t+s} - X_t)_t \text{ stationary process for any } s \in \mathbb{R} \end{cases}$$

Definition. $X = (X_k)_{k \in \mathbb{Z}}$ is a LM stationary process when

$$X_k = Y_{k+1} - Y_k, \text{ for } k \in \mathbb{Z}, \text{ with } (Y_k) \text{ } H\text{-SSSI}$$

2.3.1. Two famous examples of LM processes. Fractional Gaussian Noise

$Y = \{Y_t, t \in \mathbb{R}\}$ is a Fractional Brownian Motion

Definition (Kolmogorov, 1940, Lévy, 1965).

\iff

Y is a centered Gaussian process with stationary increments such as $\mathbb{E}Y_t^2 = \sigma^2 |t|^{2H}$, $\sigma^2 > 0$, $H \in (0, 1]$

Consequences:

- (1) Y is the only Gaussian H -SSSI process
- (2) $X = (Y_{t+1} - Y_t)_{t \in \mathbb{Z}}$, Fractional Gaussian noise.

$$\implies X \text{ is LM if } H > 1/2: r(k) \sim \sigma^2 H(2H - 1) |k|^{2H-2} |k| \rightarrow \infty$$

Simulations of a FGN trajectory How to generate a FGN path (X_1, \dots, X_n) ?

- (1) Natural idea: Cholesky decomposition of $\Sigma = (r(|j - i|))_{1 \leq i, j \leq n} = RR'$
 $X = RZ$, with Z a sample of Gaussian i.i.d.r.v.

- (2) Best choice: plug Σ in a circulant matrix and use the spectral decomposition of a circulant matrix \implies spectral decomposition of $\Sigma^{1/2}$

FARIMA(p, d, q) processes

Definition (Granger et Joyeux, 1980). If $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$ white noise, $X = \{X_t, t \in \mathbb{R}\}$ FARIMA(p, d, q) process when

$$\begin{aligned} \iff (1 - B)^d P(B)(X) &= Q(B)(\varepsilon) \text{ avec } P \in \mathbb{R}_p[X], Q \in \mathbb{R}_q[X] \\ \iff X_k &= \sum_{j=0}^{\infty} \left(\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \right) \eta_{k-j} \text{ with } (\eta_t) \text{ ARMA}(p, q): \\ &\eta_k + \theta_1 \eta_{k-1} + \dots + \theta_p \eta_{k-p} = \varepsilon_k + \phi_1 \varepsilon_{k-1} + \dots + \phi_q \varepsilon_{k-q} \end{aligned}$$

Consequence: $f(\lambda) = \frac{\sigma^2}{\pi} \left| \frac{Q(e^{i\lambda})}{P(e^{i\lambda})} \right|^2 \frac{1}{|1 - e^{i\lambda}|^{2d}}$ for $\lambda \neq 0$

\implies Existence of X since $f \geq 0$ measurable function

$\implies X$ is LM if $0 < d < 1/2$: $f(\lambda) \sim C |\lambda|^{-2d}$ when $\lambda \rightarrow 0$

Simulations of a FARIMA trajectory. How to generate a FARIMA path (X_1, \dots, X_n) ?

(1) Best Gaussian idea: use also the circulant matrix method...

(2) Non Gaussian idea: truncation of $\sum_{j=0}^{\infty} \left(\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \right) \eta_{k-j}$

$$\implies X_k = \sum_{j=0}^M \left(\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \right) \eta_{k-j} \quad \text{with } M \text{ large number}$$

Limit theorem for the sum of Short Memory processes

Definition. Let $\varepsilon = (\varepsilon_k)_{k \in \mathbb{Z}}$ be a L^2 -white noise and $(a_k)_{k \geq 0} \in \ell^2(\mathbb{R})$. A one-sided linear process $X = (X_k)_{k \in \mathbb{Z}}$ is defined by

$$X_k = \sum_{j \geq 0} a_j \varepsilon_{k-j} \quad \text{for } k \in \mathbb{Z}$$

Consequence: If $|a_j| = j^{-\beta} L(j)$, $1/2 < \beta < 1$, X LM with $D = 2\beta - 1$.

Theorem (Ibragimov, 1962). If X is a one-sided linear process with $\sum_{k \in \mathbb{Z}} r(k) = \sigma^2 \neq 0$, then

$$\left(N^{-1/2} \sum_{j=1}^{[Nt]} X_t \right)_{t \geq 0} \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \sigma^2 (W_t)_{t \geq 0}.$$

A general limit theorem for the sum of LM processes

Theorem (Rosenblatt, 1961, Davydov 1970). If X is a Gaussian or one-sided linear process, $\mathbb{E}(X_0) = 0$, $\text{var}(X_0) = 1$, LM (Definition 1) with parameter $D \in]0, 1[$,

$$\left(\frac{1}{N^{1-D/2} L^{1/2}(N)} \sum_{j=1}^{[Nt]} X_t \right)_{t \geq 0} \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} (B_{1-D/2}(t))_{t \geq 0}$$

Limit theorem for functionals of Gaussian LM processes

H_n be the Hermite polynomial of degree n : $H_1(X) = X$, $H_2(X) = X^2 - 1, \dots$

Call Hermite rank m of $f \in L^2(\mathcal{N}(0, 1))$ if $\exists (c_j)_j$ such as $f = \sum_{j=m}^{\infty} c_j H_j$.

Theorem (Rosenblatt, 1961, Taqqu, 1975, Dobrushin and Major, 1979). If the Hermite rank of f is m , if X Gaussian process, $\mathbb{E}(X_0) = 0$, $\text{var}(X_0) = 1$, LM with parameter $D \in]0, 1/m[$,

$$\left(\frac{1}{N^{1-\frac{mD}{2}} L^{1/2}(N)} \sum_{j=1}^{[Nt]} [f(X_t) - \mathbb{E}(f(X_0))] \right)_t \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \frac{\mathbb{E}(f(X_0)H_m(X))}{m!} (Z_{m,D}(t))_t$$

Two particular cases

(1) If $m = 1$, $0 < D < 1$, the limit process is $Z_{1,D}(t) = B_{1-D/2}(t)$,

$$\text{ex: } \left(\frac{1}{N^{1-D/2} L^{1/2}(N)} \sum_{j=1}^{[Nt]} X_t \right)_{t \geq 0} \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} (B_{1-D/2}(t))_{t \geq 0}$$

(2) If $m = 2$, $0 < D < 1/2$, limit process $Z_{2,D}(t)$, Rosenblatt process,

$$\text{ex: } \left(\frac{1}{N^{1-D} L^{1/2}(N)} \sum_{j=1}^{[Nt]} [X_t^2 - 1] \right)_{t \geq 0} \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \frac{1}{2} (Z_{2,D}(t))_{t \geq 0}$$

Theorem (Surgailis, 1980, Giraitis, 1985, Ho and Hsin, 1997, Surgailis, 2000). *K function and $K_\infty^r = \left(\frac{\partial^r}{\partial x^r} \int K(x+y) d\mu_X(y) \right)_0$. If $K_\infty^r = 0$ for $r < k$ and $|K_\infty^k| \neq 0 < \infty$, if X LM one-sided linear process, with parameter $D \in]0, 1/k[$,*

$$\frac{1}{N^{1-\frac{kD}{2}} L^{2k}(N)} \sum_{j=1}^N [K(X_t) - \mathbb{E}(K(X_0))] \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} C(k, \mu) Z_{k,D}(1)$$

3. IDENTIFICATION D'UN PROCESSUS STATIONNAIRE À TEMPS DISCRET

On s'intéresse ici à l'identification de la partie bruit d'un processus à temps discret. Lorsque l'on considère une suite de v.a.i.i.d., on connaît un certain nombre de techniques (estimation, tests) et des résultats asymptotiques qui permettent d'en savoir plus sur cette suite (par exemple estimer sa variance, la probabilité d'une occurrence, tester si les variables sont gaussiennes,...). Pour un processus à temps discret, il faut compter avec la structure de dépendance du bruit, ce qui nous oblige à donner des extensions aux résultats asymptotiques classiques (loi des grands nombres, théorème de la limite centrale,...).

3.1. Comportement asymptotique des processus stationnaires à temps discret.

Proposition. *Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus du second ordre à temps discret stationnaire d'espérance m et de densité spectrale f continue en 0. Alors:*

$$\begin{aligned} \overline{X_n} &\xrightarrow[n \rightarrow +\infty]{p.s.} m \text{ et} \\ \lim_{n \rightarrow \infty} \mathbb{E}(\sqrt{n}(\overline{X_n} - m))^2 &= 2\pi f(0). \end{aligned}$$

Proof. 1/ La convergence presque-sûre nécessite des arguments trop longs à exposer ici (voir par exemple Azencott et Dacunha-Castelle, 1984, pour une preuve).

2/ On a facilement $\mathbb{E}(\sqrt{n}(\overline{X_n} - m))^2 = n \text{var}(\overline{X_n}) = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n r_X(|j-i|)$ où r_X est l'autocovariance de X . On montre alors que $n \text{var}(\overline{X_n}) = \frac{1}{n} \sum_{k=-n}^n (n-|k|) r_X(|k|)$. Mais comme f est continue en 0 on a $\sum_{k \in \mathbb{Z}} |r_X(k)| < \infty$. Par suite

$$\frac{1}{n} \sum_{k=-n}^n (n-|k|) r_X(|k|) = \sum_{k=-n}^n r_X(|k|) - \frac{1}{n} \sum_{k=-n}^n |k| r_X(|k|).$$

Mais

$$\begin{aligned} \frac{1}{n} \sum_{k=-n}^n |k| r_X(|k|) &= \frac{1}{n} \sum_{k=-[\sqrt{n}]}^{[\sqrt{n}]} |k| r_X(|k|) + \frac{2}{n} \sum_{k=[\sqrt{n}]+1}^n |k| r_X(|k|) \\ \implies \frac{1}{n} \sum_{k=-n}^n |k| |r_X(|k|)| &\leq \frac{1}{\sqrt{n}} \sum_{k=-[\sqrt{n}]}^{[\sqrt{n}]} |r_X(|k|)| + 2 \sum_{k=[\sqrt{n}]+1}^n |r_X(|k|)| \\ &\xrightarrow[n \rightarrow +\infty]{} 0 \text{ car } \sum_{k \in \mathbb{Z}} |r_X(k)| < \infty. \end{aligned}$$

En conséquence, $\frac{1}{n} \sum_{k=-n}^n (n-|k|) r_X(|k|) \xrightarrow[n \rightarrow +\infty]{} \sum_{k \in \mathbb{Z}} r_X(k) = 2\pi f(0)$. □

Proposition. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus gaussien stationnaire d'espérance m et de densité spectrale f . Alors:

$$(1) \text{ Si } \lim_{k \rightarrow \infty} r(k) = 0 \text{ alors } \overline{X}_n \xrightarrow[n \rightarrow +\infty]{p.s.} m.$$

$$(2) \text{ Si la densité spectrale } f \text{ existe et est continue en } 0, \text{ alors } \sqrt{n}(\overline{X}_n - m) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 2\pi f(0)).$$

Proof. La convergence presque-sûre est délicate, mais le théorème central limite se déduit du comportement asymptotique précédent (une suite de distributions gaussiennes de moyenne 0 et de variance tendant vers $2\pi f(0)$, lois de $\sqrt{n}(\overline{X}_n - m)$, converge quand $n \rightarrow \infty$ vers une distribution gaussienne de moyenne 0 et de variance $2\pi f(0)$). \square

Le cas gaussien permet donc d'avoir des théorèmes limite sous des conditions assez faibles. Pour le cas plus général, on peut commencer par donner un théorème central limite pour les processus M -dépendants qui se définissent ainsi:

Définition. Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire tel que $\text{var } X_0 = \sigma^2 < \infty$. On suppose que X est M -dépendant, où $M \in \mathbb{N}$, c'est-à-dire que pour tout t tel que $t \geq M + 1$, alors $(X_{-k})_{k \in \mathbb{N}}$ est indépendant de $(X_{t+k})_{k \in \mathbb{N}}$.

Théorème (Théorème central limite pour les processus M -dépendants). Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire tel que $\text{var } X_0 = \sigma^2 < \infty$, d'autocovariance r_X , et M -dépendant, avec $M \in \mathbb{N}$. Alors

$$\sqrt{n}(\overline{X}_n - \mathbb{E}(X_0)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \gamma^2) \quad \text{avec} \quad \gamma^2 = \sum_{k=-M}^M r_X(k) = 2\pi f(0).$$

Proof. La preuve se fait en découpant n en r morceaux, blocs, de taille p , avec r et p qui tendent vers l'infini. Ainsi on considère $Z_i = \frac{1}{p} \sum_{k=1}^{p-M} X_{(i-1)p+k}$ pour $i = 1, \dots, r$. Les variables aléatoires Z_i sont des v.a.i.i.d. du fait de la stationnarité de X et de la M -dépendance. A p fixé, on peut appliquer un théorème central limite classique aux variables Z_i et on obtient que $\sqrt{r} \overline{Z}_r \xrightarrow[r \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \gamma_p^2)$ où $\overline{Z}_r = \frac{1}{r} \sum_{i=1}^r Z_i$ et $\gamma_p^2 = \sigma^2 + 2 \sum_{i=1}^M (1 - \frac{i}{p}) r_X(i) \xrightarrow[p \rightarrow +\infty]{P} \gamma^2$.

Posons $U_i^{(p)} = \frac{1}{p} \sum_{k=0}^{M-1} X_{ip-k}$ pour $i = 1, \dots, r$. Ces variables aléatoires $U_i^{(p)}$ sont également des v.a.i.i.d. et $\text{pvar}(\overline{U}_r^{(p)}) \xrightarrow[n \rightarrow +\infty]{} 0$ avec $\overline{U}_r^{(p)} = \frac{1}{r} \sum_{i=1}^r U_i^{(p)}$. Il ne reste plus qu'à écrire que $\overline{X}_n = \frac{p}{n} \sum_{i=1}^r Z_i + \overline{U}_r^{(p)}$ pour montrer le théorème de la limite centrale voulu en faisant tendre p vers l'infini et en utilisant le Théorème de Slutsky. \square

Une fois ce théorème démontré, on peut l'utiliser même pour des processus qui ne sont pas M -dépendants, en faisant tendre M vers l'infini. Ainsi on obtient le théorème suivant:

Théorème (Théorème central limite pour les processus ARMA et ARCH). Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus ARMA(p, q) ou ARCH(p) stationnaire tel que $\text{var } X_0 = \sigma^2 < \infty$. Alors

$$\sqrt{n}(\overline{X}_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \gamma^2) \quad \text{avec} \quad \gamma^2 = 2\pi f(0).$$

Pour un processus linéaire centré, plus généralement on obtient le même théorème central limite sous la condition $\sum_{k \in \mathbb{N}} |k r_X(k)| < \infty$.

On peut maintenant également donner une loi des grands nombres et un théorème de la limite centrale vérifiés par les chaînes de Markov.

Théorème (Théorème ergodique). Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène irréductible sur $E = (e_1, \dots, e_p)$ ayant pour unique mesure de probabilité invariante μ . Alors pour toute fonction $h : E \rightarrow \mathbb{R}$ mesurable, et quelque soit la loi de X_0 ,

$$\frac{1}{n+1} \sum_{j=0}^n h(X_j) \xrightarrow[n \rightarrow +\infty]{p.s.} \int h(x) d\mu(x) = \sum_{i=1}^p h(e_i) \mu(e_i).$$

Théorème (Théorème de la limite centrale pour les chaînes de Markov à espace d'états). Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène irréductible sur $E = (e_1, \dots, e_p)$ ayant pour unique

mesure de probabilité invariante μ . Alors pour toute fonction $h : E \rightarrow \mathbb{R}$ mesurable, et quelque soit la loi de X_0 ,

$$\sqrt{n} \left(\frac{1}{n+1} \sum_{j=0}^n h(X_j) - \int h(x) d\mu(x) \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \int h^2(x) d\mu(x) - \left(\int h(x) d\mu(x) \right)^2 \right).$$

Voyons maintenant des utilisations de ces résultats. En premier lieu, on peut définir les moments empiriques d'ordre 2 d'un processus à temps discret:

Définition. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus à temps discret. On appelle covariance empirique de (X_1, \dots, X_n) pour $p = 0, 1, \dots, n-1$,

$$\hat{c}_n(p) = \frac{1}{n} \sum_{k=1}^{n-p} (X_{k+p} - \bar{X}_n)(X_k - \bar{X}_n) \text{ et } \hat{c}_n(-p) = \hat{c}_n(p).$$

Lorsque l'on sait que le processus est centré alors pour $p = 0, 1, \dots, n-1$,

$$\hat{c}_n(p) = \frac{1}{n} \sum_{k=1}^{n-p} X_{k+p} X_k \text{ et } c_n(-p) = c_n(p).$$

On appelle corrélation empirique de (X_1, \dots, X_n) pour $p = 0, 1, \dots, n-1$,

$$\hat{\rho}_n(p) = \frac{\hat{c}_n(p)}{\hat{c}_n(0)}.$$

Remarque.

Un certain nombre de logiciels propose de tracer le "correlogram" de la série (souvent la fonction ACF) soit le graphe de $k \mapsto \hat{\rho}_n(k)$ pour $k \in \{0, 1, \dots, m\}$, avec m souvent proche de \sqrt{N} . Il est bien clair que l'on a toujours $\hat{\rho}_n(0) = 1$ et nous allons voir que sous des hypothèses assez générales on peut espérer que ce correlogramme soit une bonne approximation du graphe de la fonction de corrélation. On définit maintenant l'estimateur "naturel" de la densité spectrale qui consiste à remplacer la covariance par la covariance empirique dans l'expression de la densité spectrale.

Définition. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus à temps discret centré. On appelle périodogramme de (X_1, \dots, X_n) la fonction $I_n(\lambda) : [-\pi, \pi[\rightarrow \mathbb{R}_+$ telle que:

$$I_n(\lambda) = \frac{1}{2\pi} \sum_{k=1}^{n-1} \hat{c}_n(k) e^{-ik\lambda} = \frac{1}{2\pi n} \left| \sum_{k=1}^n X_k e^{-ik\lambda} \right|^2.$$

On peut alors montrer la propriété suivante:

Propriété. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus à temps discret centré stationnaire ayant des moments d'ordre 2. Si les covariances $r(k)$ sont telles que $\sum |r(k)| < \infty$ alors

- (1) $\lim_{n \rightarrow \infty} \mathbb{E}(I_n(\lambda)) = f(\lambda)$ pour tout $\lambda \in [-\pi, \pi[$.
- (2) si X est un processus gaussien, $\lim_{n \rightarrow \infty} \text{var}(I_n(\lambda)) = \begin{cases} f^2(\lambda) & \text{si } \lambda \in]-\pi, 0[\cup]0, \pi[\\ 2f^2(\lambda) & \text{si } \lambda = -\pi, 0. \end{cases}$
- (3) si X est un processus gaussien, $\lim_{n \rightarrow \infty} \text{cov}(I_n(\lambda), I_n(\lambda')) = 0$ si $\lambda \neq \lambda'$.

Avec cette propriété on voit que I_n est un estimateur asymptotiquement non biaisé de f , mais un estimateur non convergent dans le cas gaussien. On pourrait généraliser cette propriété dans le cas plus général d'un processus ayant des moments d'ordre 4. Cependant cet estimateur sera intéressant lorsqu'on l'intègre par rapport à une fonction:

Définition. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus à temps discret du second ordre. Pour g une fonction continue sur $[-\pi, \pi[$, on définit le périodogramme intégré par:

$$J_n(g) = \int_{-\pi}^{\pi} g(\lambda) I_n(\lambda) d\lambda.$$

Proposition. Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus centré stationnaire ARMA(p, q) ou ARCH(p) tel que $\mathbb{E} X_0^4 < \infty$ (ce qui, dans le cas d'un processus ARCH(p) signifie que $(\mathbb{E} \varepsilon_0^4)^{1/2} \sum_{j=1}^p b_j < 1$), alors:

- (1) Pour g une fonction continue sur $[-\pi, \pi[$, $J_n(g) \xrightarrow[n \rightarrow +\infty]{p.s.} \int_{-\pi}^{\pi} f(\lambda)g(\lambda)d\lambda$.
- (2) Pour $(g_i)_{1 \leq i \leq m}$ une famille de fonctions continues sur $[-\pi, \pi[$ alors:

$$\sqrt{n} \left(J_n(g_i) - \int_{-\pi}^{\pi} f(\lambda)g(\lambda)d\lambda \right)_{1 \leq i \leq m} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, \Sigma)$$

$$\text{avec } \Sigma = (\sigma_{ij})_{1 \leq i, j \leq m} \text{ et } \sigma_{ij} = 4\pi \int_{-\pi}^{\pi} f^2(\lambda)g_i(\lambda)g_j(\lambda)d\lambda.$$

Conséquence.

- Rappelons que pour un ARCH(p) stationnaire d'ordre 2, f est une fonction constante.
- Ce double résultat s'applique aussi pour le carré d'un processus ARCH(p) puisque ce carré est un processus ARMA. La condition sera donc que X soit stationnaire et tel que $\mathbb{E} X_0^8 < \infty$, ce qui est possible dès que $(\mathbb{E} \varepsilon_0^8)^{1/4} \sum_{j=1}^p b_j < 1$.
- Ce double résultat s'étend également aux chaînes de Markov homogène à espace d'états fini irréductibles, et ceci quelque soit la loi de X_0 .
- Comme $\hat{c}_n(p) = \int_{-\pi}^{\pi} \cos(p\lambda)I_n(\lambda)d\lambda$, on peut appliquer le théorème de la limite centrale précédent. On obtient ainsi que:

$$\sqrt{n} \left(\hat{c}_n(p) - r(p) \right)_{0 \leq p \leq m} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_{m+1} \left(0, \left(4\pi \int_{-\pi}^{\pi} f^2(\lambda) \cos(\lambda i) \cos(\lambda j) d\lambda \right)_{0 \leq i, j \leq m} \right).$$

- En utilisant la Delta-méthode pour la fonction $G : (x_0, \dots, x_m) \in \mathbb{R}^* \times \mathbb{R}^m \mapsto \left(\frac{x_1}{x_0}, \frac{x_2}{x_0}, \dots, \frac{x_m}{x_0} \right) \in \mathbb{R}^m$, on peut déduire un théorème de la limite centrale pour $(\hat{\rho}_n(1), \dots, \hat{\rho}_n(m))$. Cela permet notamment de tester si X est un bruit blanc car sous une telle hypothèse, $f(\lambda) = \sigma^2/2\pi$ et on a alors

$$\sqrt{n} (\hat{\rho}_n(i))_{1 \leq i \leq m} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, I_m).$$

Comme dit plus haut ce résultat est également vrai pour un ARCH(p), ce qui limite la puissance d'un tel test.

- En utilisant un noyau comme fonction g que l'on fait se rapprocher d'une impulsion de Dirac en λ_0 , on obtient un estimateur de la densité $f(\lambda_0)$.

3.2. Estimation paramétrique et semi-paramétrique pour un processus stationnaire à

temps discret. Notation: On suppose que l'on a observé une trajectoire (X_1, \dots, X_N) d'un processus paramétrique de paramètre $\theta = (\theta_1, \dots, \theta_\ell)$ où $\ell \in \mathbb{N}^*$ tel que la "vraie" valeur prise par ce paramètre θ^* soit inconnue. La méthode qui suit pourrait être généralisée à d'autres exemples de processus, mais ici nous ne considérerons que les cas paramétriques suivants:

- Les paramètres $b_1, \dots, b_p, c_1, \dots, c_q$ d'un processus ARMA(p, q) stationnaire dont on supposera que le bruit blanc $(\varepsilon_k)_k$ admet une densité par rapport à la mesure de Lebesgue;
- Les paramètres b_1, \dots, b_p d'un processus ARCH(p) stationnaire dont on supposera que le bruit blanc $(\varepsilon_k)_k$ admet une densité par rapport à la mesure de Lebesgue;
- Les probabilités de transition $p_{ij} = \Pr(X_1 = e_j | X_0 = e_i)$ d'une chaîne de Markov homogène à espace d'états fini irréductibles. Comme la matrice est stochastique, il n'y a que $p(p-1)$ inconnues puisque par exemple $p_{ip} = 1 - \sum_{j=1}^{p-1} p_{ij}$.

On définit $L_N(\theta, \omega)$ une vraisemblance pour $\theta \in \Theta$ (ensemble localement compact de \mathbb{R}^ℓ). Soit

$$\hat{\theta}_N = \underset{\theta \in \Theta}{\text{Argmax}} L_N(\theta) = \underset{\theta \in \Theta}{\text{Argmax}} (\log L_N(\theta)).$$

Par “une” vraisemblance on veut dire: soit la densité du vecteur (X_1, \dots, X_N) (pour les processus ARMA(p, q)), soit la densité conditionnelle par rapport à X_0, X_{-1}, \dots (pour les processus ARCH(p) ou les chaînes de Markov à espace d'états fini irréductibles). Rappelons que dans ce dernier cas

$$L_N(\theta) = f_{((X_1, \dots, X_N) | X_0, X_{-1}, \dots)} = f_{(X_N | X_{N-1}, X_{N-2}, \dots)} f_{(X_{N-1} | X_{N-2}, X_{N-3}, \dots)} \times \dots \times f_{(X_1 | X_0, X_{-1}, \dots)}$$
 (attention à ne pas confondre cette densité f , qui peut d'ailleurs être une probabilité dans le cas d'une mesure discrète, avec la densité spectrale vue avant). Dans le cas d'une chaîne de Markov homogène irréductible sur (e_1, \dots, e_p) , le modèle est totalement posé et la densité revient à un produit de probabilités et le logarithme de L_N devient:

$$\log L_N(\theta) = \sum_{i=1}^p \sum_{j=1}^{p-1} \log(p_{ij}) \sum_{k=1}^{N-1} \mathbb{I}_{(X_k, X_{k+1})=(e_i, e_j)} + \sum_{i=1}^p \log(1-p_{i,1} - \dots - p_{i,p-1}) \sum_{k=1}^{N-1} \mathbb{I}_{(X_k, X_{k+1})=(e_i, e_p)}.$$

Notons que dans un tel cas la statistique constituée par le vecteur $(\sum_{k=1}^{N-1} \mathbb{I}_{(X_k, X_{k+1})=(e_i, e_j)})_{1 \leq i \leq p, 1 \leq j \leq p-1}$ est exhaustive (on ne peut pas prendre tous les i et j car la statistique serait liée).

Théorème. *Sous les hypothèses précédentes, si $\mathbb{E} X_0^4 < \infty$ et*

- (1) *si $\mathbb{E} [|\log f_{(X_1 | X_0, X_{-1}, \dots)}|] < \infty$ on a $\hat{\theta}_N \xrightarrow[N \rightarrow +\infty]{p.s.} \theta^*$.*
- (2) *si $\mathbb{E} [\log^2 f_{(X_1 | X_0, X_{-1}, \dots)}] < \infty$ on a $\sqrt{N}(\hat{\theta}_N - \theta^*) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, I(\theta^*)^{-1})$, où*

$$I_{ij}(\theta^*) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \log f_{\theta}(\lambda)}{\partial \theta_i} \right)_{\theta^*} \cdot \left(\frac{\partial \log f_{\theta}(\lambda)}{\partial \theta_j} \right)_{\theta^*} d\lambda.$$

Remarque.

On retrouve donc ici une vitesse en \sqrt{N} comme dans le cas de variables indépendantes. Cependant une des limitations de la méthode du maximum de vraisemblance réside dans le fait qu'il faille connaître exactement la loi du processus pour en déduire $L_N(\theta)$. L'utilisation de cet estimateur reste malgré tout intéressant pour les processus ARCH(p) (pour lesquels on peut utiliser la vraisemblance conditionnelle gaussienne même si l'innovation n'est pas gaussienne) et les chaînes de Markov. Mais dans le cas d'un processus ARMA(p, q), d'un processus gaussien ou linéaire et des processus longue mémoire en général, on préférera utiliser une approximation de l'estimateur $\hat{\theta}_N$ dite approximation de Whittle (voir ci-dessous) qui offre deux avantages: pas de nécessité de connaître a priori la loi du processus et un grand gain en terme de calcul tout en conservant une vitesse de convergence en \sqrt{N} .

Théorème. *Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus gaussien ou un processus linéaire, dont la densité spectrale s'écrit de manière paramétrique f_{θ} et vérifie certaines propriétés (dont l'identifiabilité: $f_{\theta_1} = f_{\theta_2} \implies \theta_1 = \theta_2$). On considère le contraste, dit encore contraste de Whittle:*

$$U_N(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log(f_{\theta}(\lambda)) + \frac{I_N(\lambda)}{f_{\theta}(\lambda)} d\lambda.$$

Soit $\widetilde{\theta}_N = \underset{\theta \in \Theta}{\text{Argmin}} U_N(\theta)$, appelé estimateur de θ par minimum de contraste (ou maximum de vraisemblance approché de Whittle). Alors, si X possède un moment d'ordre 4, avec une matrice $J(\theta^)$:*

$$\sqrt{N}(\widetilde{\theta}_N - \theta^*) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, J(\theta^*)^{-1}).$$

Remarque.

Ce résultat est valable pour les processus ARMA(p, q), FARIMA(p, d, q) et les bruits gaussiens fractionnaires. On peut remarquer cependant qu'un processus GARCH n'est pas identifié par sa densité spectrale. Cet estimateur peut cependant aussi être utilisé en l'appliquant aux carrés de ce processus.

Remarque.

Notons tout d'abord que la matrice $J = I$ lorsque le processus est gaussien. Sinon, elle fait intervenir les moments d'ordre 4 de l'innovation. Ensuite, concrètement, quand on travaille sur des données, les intégrales sont approchées par des sommes de Riemann calculées pour les $\lambda_k = \pi k/n$. On peut montrer que l'erreur d'approximation n'est pas préjudiciable au théorème de la limite centrale ci-dessus.

Conséquence.

On préférera nettement l'estimateur par contraste de Whittle pour les processus identifiés par leur densité spectrale à car il peut s'appliquer facilement à des échantillons de grande taille (alors que par exemple dans le cas gaussien, l'estimateur de maximum de vraisemblance classique requiert d'inverser la matrice de covariance ce qui peut être impossible numériquement dès que $n > 10^4$) et comme dit précédemment, il ne requiert pas la connaissance de la loi de X ou de ε . Pour les processus GARCH(p, q), mais aussi toute une famille générique de processus, les processus affines causaux, une autre méthode d'estimation semi-paramétrique peut être utilisée: la méthode d'estimation par quasi-vraisemblance gaussienne. Définissons d'abord la famille des processus affines causaux:

Définition. Si $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc, on dit que $(X_t)_{t \in \mathbb{Z}}$ est un processus affine causal si

$$X_t = M((X_{t-k})_{k \in \mathbb{N}^*}) \varepsilon_t + f((X_{t-k})_{k \in \mathbb{N}^*}) \quad \text{pour tout } t \in \mathbb{Z},$$

où M et f sont des fonctions à valeurs respectives dans \mathbb{R}_+ et \mathbb{R} et définies sur les suites numériques réelles.

Les processus ARMA ($M = 1$ et $f((x_k)_{k \in \mathbb{N}^*}) = \sum_{i=1}^{\infty} c_i$), et les processus GARCH(p, q) ($f = 0$ et $M((x_k)_{k \in \mathbb{N}^*}) = \sqrt{c_0 + \sum_{i=1}^{\infty} c_i x_i^2}$) sont des exemples de processus affines causaux. Pour la stationnarité, nous allons rajouter des hypothèses sur ces deux fonctions:

Propriété. On suppose que les fonctions M et f sont Lipschitziennes, ce qui signifie que pour tout $x = (x_i)_{i \in \mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}}$ et $y = (y_i)_{i \in \mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}}$, il existe des suites $(a_i(M))$ et $(a_i(f))$ de réels positifs telles que

$$|M(x) - M(y)| \leq \sum_{i=1}^{\infty} a_i(M) |x_i - y_i| \quad \text{et} \quad |f(x) - f(y)| \leq \sum_{i=1}^{\infty} a_i(f) |x_i - y_i|.$$

Pour $r \geq 1$, si $\mathbb{E}(|\varepsilon_0|^r) < \infty$, alors (X_t) est un processus stationnaire vérifiant $\mathbb{E}(|X_0|^r) < \infty$ lorsque:

$$\sum_{i=1}^{\infty} a_i(f) + (\mathbb{E}(|\varepsilon_0|^r))^{1/r} \sum_{i=1}^{\infty} a_i(M) < 1.$$

On suppose désormais que le modèle est semi-paramétrique c'est-à-dire que pour $\theta \in \mathbb{R}^d$, $M = M_\theta$ et $f = f_\theta$, que les fonctions $\theta \rightarrow M_\theta$ et $\theta \rightarrow f_\theta$ sont connues ainsi que l'échantillon (X_1, \dots, X_n) est connu mais θ est inconnu.

On utilise la méthode de quasi-vraisemblance gaussien pour estimer θ . Pour cela on suppose dans un premier temps que la loi de ε_0 est gaussienne. Notons $f_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \dots)$, $M_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \dots)$. Alors la log-densité conditionnelle de X_t sachant $(X_{t-1}, X_{t-2}, \dots)$ est:

$$q_t(\theta) = -\frac{1}{2} \left[\frac{(X_t - f_\theta^t)^2}{(M_\theta^t)^2} + \log((M_\theta^t)^2) \right].$$

Mais X_0, X_{-1}, \dots , sont inconnus, et on va les remplacer par des 0. Posons $\hat{f}_\theta^t = f_\theta(X_{t-1}, \dots, X_1, 0, \dots)$ et $\hat{M}_\theta^t = M_\theta(X_{t-1}, \dots, X_1, 0, \dots)$, alors on définit la quasi log-densité devient

$$\hat{q}_t(\theta) = -\frac{1}{2} \left[\frac{(X_t - \hat{f}_\theta^t)^2}{(\hat{M}_\theta^t)^2} + \log((\hat{M}_\theta^t)^2) \right].$$

On définit alors l'estimateur par quasi-vraisemblance gaussienne par: $\hat{\theta}_n = \arg \max_{\theta \in \Theta} \hat{L}_n(\theta)$ avec

$$\hat{L}_n(\theta) = \sum_{t=1}^n \hat{q}_t(\theta) \text{ la quasi-vraisemblance gaussienne.}$$

On peut évidemment appliquer cet estimateur même si le bruit n'est en réalité pas gaussien. On peut alors montrer le résultat suivant:

Théorème. *Sous les conditions de stationnarité, si $r \geq 4$, et sous d'autres conditions sur M_θ et f_θ (notamment l'identifiabilité et $M_\theta \geq \underline{M} > 0$), alors il existe des matrices $F(\theta_0)^{-1}$ et $G(\theta_0)$ telles que*

$$(3) \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_d(0, F(\theta_0)^{-1}G(\theta_0)F(\theta_0)^{-1}).$$

Remarque.

Ce théorème est notamment vérifié par les processus ARMA et GARCH.

3.3. Identification et test d'adéquation pour les processus ARMA et ARCH. Estimation de l'ordre d'un processus ARMA: Une autre question apparaît une fois l'estimation des paramètres obtenue: comment estimer l'ordre (p, q) de l'ARMA qui dans les deux techniques précédentes est supposé connu? Pour ce faire on utilisera un critère de sélection de modèle, par exemple le critère BIC vu précédemment. Dans le cas d'un processus ARMA celui-ci s'écrit:

$$BIC(p, q) = \log \left(\frac{1}{N} \sum_{j=1}^N \hat{\varepsilon}_j^2 \right) + \frac{\log N}{N} (p + q),$$

où $\hat{\varepsilon}_j = \frac{\hat{P}_N(B)}{\hat{Q}_N(B)} \cdot X_j$ (il faut donc que Q soit inversible causal, donc que les racines de Q soient en

dehors du disque trigonométrique). Les estimateurs $\hat{P}_N(B)$ et $\hat{Q}_N(B)$ sont calculés en remplaçant leurs coefficients a_i et b_j par les estimateurs obtenus par une des méthodes évoquées plus haut. On calcule ce critère pour tous $0 \leq p \leq p_{\max}$ et $0 \leq q \leq q_{\max}$ et on choisira

$$(\hat{p}_N, \hat{q}_N) = \underset{0 \leq p \leq p_{\max}, 0 \leq q \leq q_{\max}}{\text{Arg min}} BIC(p, q).$$

Estimation de l'ordre d'un processus ARCH: On peut soit reprendre la technique qui vient d'être présentée pour les ARMA en l'appliquant aux carrés du processus. On peut également reprendre la formule générale du critère BIC et l'appliquer à l'estimation de la log-vraisemblance conditionnelle. Auquel cas, on a:

$$BIC(p, q) = \log \left(\prod_{j=1}^N \hat{f}(X_j | X_{j-1}, X_{j-2}, \dots) \right) + \frac{\log N}{N} p,$$

où \hat{f} est obtenue en remplaçant les coefficients (b_0, \dots, b_p) par leurs estimations par maximum de vraisemblance.

Test d'adéquation à un processus ARMA: Pour tester l'adéquation à un processus ARMA, on peut utiliser un test dit de portemanteau (ce qui signifie "fourre-tout" en anglais). Après avoir calculé les résidus estimés $(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N)$ obtenus à partir (X_1, \dots, X_N) et des coefficients estimés par une méthode de type maximum de vraisemblance ou minimum de contraste: comme on a $\varepsilon = \frac{Q(B)}{P(B)}X$ on définit

$$\hat{T}_N(k) = N \sum_{j=1}^k (\hat{\rho}_\varepsilon(k))^2 \text{ où } \hat{\rho}_\varepsilon(k) = \frac{\frac{1}{N} \sum_{i=p}^{N-k} \hat{\varepsilon}_i \hat{\varepsilon}_{i+k} - \left(\frac{1}{N} \sum_{i=p}^N \hat{\varepsilon}_i \right)^2}{\frac{1}{N} \sum_{i=p}^N \hat{\varepsilon}_i^2 - \left(\frac{1}{N} \sum_{i=p}^N \hat{\varepsilon}_i \right)^2}.$$

Notons que $\widehat{\rho}_\varepsilon(k)$ est la corrélation empirique des résidus, qui doit tendre vers 0 si le modèle est bien un ARMA(p, q) dès que $k \neq 0$, suivant un Théorème de la Limite Centrale que nous avons vu un peu plus haut (d'où le N devant la statistique de test). Notons également que les $\widehat{\varepsilon}_i$ ne sont calculables que lorsque $i \geq p + 1$). On doit choisir k suffisamment grand pour donner plus de pertinence au test. Ainsi, sous l'hypothèse que le processus est bien un processus ARMA(p, q), dont le bruit admet un moment d'ordre 4 on peut alors montrer que:

$$\widehat{T}_N(k) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \chi^2(k - p - q).$$

Le fait que l'on doit retirer $p + q$ dans le nombre de degrés de liberté du χ^2 limite est assez classique et correspond au fait que cette statistique est obtenue en estimant $p + q$ paramètres.

Test d'adéquation à un processus ARCH: On peut reprendre la technique présentée pour les ARMA et l'appliquer aux carrés d'un processus ARCH. On peut aussi directement mettre en place un test du portemanteau pour les ARCH en "estimant" l'innovation $(\varepsilon_k)_k$ par

$$\widehat{\varepsilon}_k = X_k \left(\widehat{b}_0 + \sum_{j=1}^p \widehat{b}_j X_{k-j}^2 \right)^{-1/2}.$$

Il n'y a plus alors qu'à reprendre la statistique de test $\widehat{T}_N(k)$ décrite plus haut pour les ARMA.

4. PRÉDICTION POUR UN PROCESSUS À TEMPS DISCRET

Pouvoir prévoir est généralement une des possibilités offertes par les statistiques, peut-être même la plus importante.

4.1. Définitions et propriétés générales. On commence à faire quelques rappels sur l'espérance conditionnelle qui est essentielle pour donner des formulations théoriques à la prédiction.

Rappel. Si X et Y sont deux variables aléatoires continues de densité jointe $f_{(X,Y)}(x, y)$ alors

$$\mathbb{E}(X | Y = y) = \frac{\int x f_{(X,Y)}(x, y) dx}{\int f_{(X,Y)}(x, y) dx}.$$

Propriété. (1) $\mathbb{E}(Y | \mathcal{B})$ est une variable aléatoire de $\mathbb{L}^2(\Omega, \mathcal{B}, P)$; de plus, $\mathbb{E}(Y | X) = h(X)$, avec h une fonction borélienne.

(2) Pour Y_1 et Y_2 deux variables aléatoires sur (Ω, \mathcal{A}, P) , et $(a, b, c) \in \mathbb{R}^3$, alors

$$\mathbb{E}(aY_1 + bY_2 + c | \mathcal{B}) = a\mathbb{E}(Y_1 | \mathcal{B}) + b\mathbb{E}(Y_2 | \mathcal{B}) + c.$$

(3) Si $Y_1 \leq Y_2$, alors $\mathbb{E}(Y_1 | \mathcal{B}) \leq \mathbb{E}(Y_2 | \mathcal{B})$.

(4) Si $Y \in \mathbb{L}^2(\Omega, \mathcal{B}, P)$, alors $\mathbb{E}(Y | \mathcal{B}) = Y$; ainsi $\mathbb{E}(g(X) | X) = g(X)$ pour g une fonction mesurable réelle.

(5) On a $\mathbb{E}(\mathbb{E}(Y | \mathcal{B})) = \mathbb{E}Y$, donc $\mathbb{E}(\mathbb{E}(Y | X)) = \mathbb{E}Y$.

(6) Si $Y^{-1}(\mathcal{B}(\mathbb{R}))$ et \mathcal{B} sont indépendantes alors $\mathbb{E}(Y | \mathcal{B}) = \mathbb{E}Y$; ainsi, si X et Y sont indépendantes, $\mathbb{E}(Y | X) = \mathbb{E}Y$.

Proposition. Si (Y, X_1, \dots, X_n) est un vecteur gaussien, alors $\mathbb{E}(Y | (X_1, \dots, X_n)) = a_0 + a_1 X_1 + \dots + a_n X_n$, où les a_i sont des réels.

Définition. On suppose que (X_1, \dots, X_N) est connue. On appelle prédiction \widehat{X}_{N+p} à l'horizon p $\widehat{X}_{N+p} = f_p(X_1, \dots, X_N)$ où f_p est une fonction mesurable.

Définition. Soit X un processus à temps discret du second ordre. On appelle erreur moyenne quadratique (MSE) de la prédiction \widehat{X}_{N+p} le réel

$$\mathbb{E}(\widehat{X}_{N+p} - X_{N+p})^2.$$

Proposition. On suppose que X un processus à temps discret du second ordre. La prédiction \widehat{X}_{N+p} optimale au sens de la moyenne quadratique (ou des moindres carrés) est définie par:

$$\widehat{X}_{N+p} = \widehat{f}(X_1, \dots, X_N) \text{ avec } \widehat{f} = \underset{f \in \mathbb{L}^2}{\text{Arg min}} \left(\mathbb{E} (X_{N+p} - f(X_1, \dots, X_N))^2 \right).$$

Alors,

$$\widehat{X}_{N+p} = \mathbb{E} (X_{N+p} \mid (X_1, \dots, X_N)).$$

\widehat{X}_{N+p} est donc une variable aléatoire dépendant de (X_1, \dots, X_N) telle que

$$\mathbb{E} (X_{N+p} \mid (X_1, \dots, X_N)) = \frac{\int x_{N+p} f_{(X_1, \dots, X_N, X_{N+p})}(X_1, \dots, X_N, x_{N+p}) dx_{N+p}}{\int f_{(X_1, \dots, X_N, X_{N+p})}(X_1, \dots, X_N, x_{N+p}) dx_{N+p}}.$$

Propriété (Cas des processus ARMA). Soit $(X_k)_{k \in \mathbb{Z}}$ un processus ARMA(p, q) avec $P(B) \cdot X = Q(B) \cdot \varepsilon$ où les racines de P et de Q sont à l'extérieur du disque trigonométrique (donc X est stationnaire, causal et inversible). On suppose également que (X_1, \dots, X_N) est connu. Soit \widehat{X}_{N+1}

le prédicteur optimal au sens des moindres carrés. Alors $\widehat{X}_{N+1} = - \sum_{k=0}^{N-1} c_{k+1} X_{N-k}$ où les (c_k) sont définis par le développement en série entière de

$$\frac{P}{Q}(z) = \sum_{k \geq 0} c_k z^k.$$

Propriété (Cas des processus ARCH(p)). Soit $(X_k)_{k \in \mathbb{Z}}$ un processus ARCH(p) défini par

$$X_k = \xi_k \sqrt{a_0 + a_1 X_{k-1}^2 + \dots + a_p X_{k-p}^2} \quad \text{pour tout } k \in \mathbb{Z},$$

où X est stationnaire d'ordre 2 (donc $(a_1 + \dots + a_p) \mathbb{E} \xi_0^2 < 1$). On suppose également que (X_1, \dots, X_N) est connu. Soit \widehat{X}_{N+1} le prédicteur optimal au sens des moindres carrés. Alors $\widehat{X}_{N+1} = 0$.

Propriété (Cas des chaînes de Markov à espace d'états fini). Soit $(X_k)_{k \in \mathbb{N}}$ une chaîne de Markov homogène à espace d'états fini $(e_1, \dots, e_p) \in \mathbb{R}^p$, de matrice de transition $P = (p_{ij})_{1 \leq i, j \leq p}$. Soit

$$\widehat{X}_{N+1} \text{ le prédicteur optimal au sens des moindres carrés. Alors } \widehat{X}_{N+1} = \mathbb{E} [X_{N+1} \mid X_N] = \sum_{j=1}^p p_{X_N, j} e_j.$$

Proposition. On suppose que X est un processus à temps discret du second ordre. Alors la prédiction **linéaire** \widehat{X}_{N+p} optimale au sens de la moyenne quadratique (ou des moindres carrés) est définie par:

$$\widehat{X}_{N+p} = \widehat{a}_1 X_1 + \dots + \widehat{a}_N X_N + \widehat{b} \text{ avec } (\widehat{a}_i)_{1 \leq i \leq N} = \underset{(a_i)_{1 \leq i \leq N} \in \mathbb{R}^N}{\text{Arg min}} \left\{ \mathbb{E} (X_{N+p} - (a_1 X_1 + \dots + a_N X_N + b))^2 \right\}.$$

Les coefficients estimés sont obtenus par régression (théorique).

Proposition. Si X est un processus gaussien, alors la prédiction linéaire par moindres carrés est aussi la prédiction par moindres carrés (donc l'espérance conditionnelle).

4.2. Prédiction par filtrage exponentiel. L'étude théorique précédente présuppose que l'on ait choisi un modèle de bruit pour prédire. On peut cependant prédire sans avoir besoin de connaître le modèle de bruit. On suppose cependant que la tendance et la saisonnalité sont composées de certaines fonctions. Le lissage exponentiel est une démarche possible pour prédire X_{N+p} lorsque (X_1, \dots, X_N) est connu.

On suppose donc que la tendance et la saisonnalité s'écrivent sous une forme connue a priori (les f_i , fonctions non constantes, et r sont connues), soit:

$$a(t) = \sum_{i=1}^k a_i f_i(t) \text{ et } S(t) = \sum_{i=1}^r s_i g_i(t) \text{ pour } t \in T,$$

avec $g_i(t) = \mathbb{I}_{\{t=i, [r]\}}$ sont r -périodiques (on ne suppose pas ici que $\sum_{i=1}^r S(i) = 0$).

Notation. • $X = (X_1, \dots, X_N)$.

- $F_i = (f_i(1), \dots, f_i(N))$ pour $i = 1, \dots, k$ et $G_i = (g_i(1), \dots, g_i(N))$ pour $i = 1, \dots, r$.
- $U = (\varepsilon(1), \dots, \varepsilon(N))$.

Le modèle s'écrit alors vectoriellement:

$$X = \sum_{i=1}^k a_i f_i + \sum_{i=1}^r s_i g_i + U.$$

Proposition. On peut estimer les coefficients (a_i) et (s_i) par une régression linéaire par moindres carrés pondérés qui accorde un poids exponentiellement plus important à X_N qu'à X_M si $M < N$. Pour cela, on minimise la distance dans \mathbb{R}^N dépendant d'un paramètre β :

$$\| X - (a_1 F_1 + \dots + s_r G_r) \|_{\beta}^2 = {}^t (X - (a_1 F_1 + \dots + s_r G_r)) \Omega^{-1}(\beta) (X - (a_1 F_1 + \dots + s_r G_r)),$$

$$\text{en notant } \Omega(\beta) = \begin{pmatrix} \beta^N & 0 & 0 & \dots & 0 \\ 0 & \beta^{N-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \text{ et avec } Z \text{ la matrice } Z = (F_1 \dots F_k G_1 \dots G_r),$$

$$\text{Alors } \widehat{X_{N+p}} = (f_1(N+p), \dots, g_r(N+p)) ({}^t Z \Omega(\beta) Z)^{-1} {}^t Z \Omega(\beta) X.$$

Remarque.

Il faut avoir en tête que si β est proche de 0 alors on prend seulement en compte les toutes dernières valeurs de la série, alors que β proche de 1 fait que l'on considère toutes les valeurs. En pratique, on peut arbitrairement choisir β entre 0.5 et 0.9. Mais le mieux est d'estimer une valeur de β adaptée au jeu de données. Pour ce faire, on utilise les données précédentes (X_{10}, \dots, X_{N-p}) (le 10 est pris ici de manière arbitraire; on pourrait aussi bien choisir $X_{\sqrt{N}}$ ou $X_{N/2}$). Pour chacune de ces valeurs on connaît la valeur réellement obtenue à l'horizon p ; il suffit donc pour une grille de valeur de β dans $[0, 1]$, de calculer la somme des carrés des écarts entre ces valeurs obtenues et les prédictions faites avec le filtre exponentielle de paramètre β . On choisira $\widehat{\beta}_N$ qui minimise cette somme de distance au carré. Cette technique peut être étendue à des prédictions linéaires: c'est que l'on appelle une prédiction par le filtre de