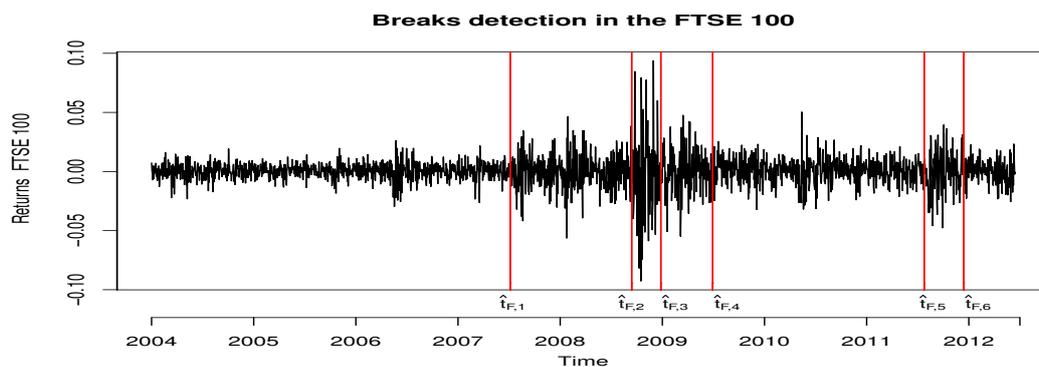


M2 TIDE 2019

Plan détaillé du cours “Econométrie des séries temporelles”

JEAN-MARC BARDET (UNIVERSITÉ PARIS 1, SAMM)



Outline

1. Times series: definitions, stationarity and properties
2. Parametric and non-parametric estimation of additive trends and seasonalities
3. Common examples of stationary time series: ARMA, GARCH, long memory processes, ...
4. Semi-parametric estimation, model selection and tests for stationary time series
5. 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] Boutahar M. et Royer-Carenzi, M. (2019). *Méthodes en séries temporelles et applications avec R*, Ellipses.
- [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*. 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 STAFV: <http://www.stafv.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://smai.emath.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 climatiques et financières,...

Objectifs. *Les objectifs de ce cours sont:*

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

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 $\forall (u_1, \dots, u_n) \in \mathbb{R}^n, u_1 X_1 + \dots + u_n X_n$ est une variable gaussienne.

Propriété. • Si X et Y sont des vecteurs issus d'un vecteur gaussien, alors (Indépendance de X et $Y \iff \text{cov}(X, Y) = 0$).

- 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 Processus stationnaires et moments

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 stationnaire.

1. (Les covariances $r(k)$ vérifient $\sum |r(k)|^2 < \infty$) $\iff (f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} r(k) e^{-ik\lambda}$ pour presque tout $\lambda \in [-\pi, \pi[$ et est de carré intégrable sur $[-\pi, \pi[$).
2. (Les covariances $r(k)$ sont telles que $\sum |r(k)| < \infty$) $\implies (f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} r(k) e^{-ik\lambda}$ pour tout $\lambda \in [-\pi, \pi[$ et f continue sur $[-\pi, \pi[$).

Ces résultats découlent de résultats classiques en analyse de Fourier (les autocovariances jouant le rôle des coefficients de Fourier, à (2π) près).

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 $(X_k)_{k \in \mathbb{Z}}$.

La densité spectrale apparaît dans les résultats suivants:

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

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

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

2. Si la densité spectrale f existe et est continue en 0 (notamment si $\sum |r(k)| < \infty$), alors $\sqrt{n}(\bar{X}_n - m) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 2\pi f(0))$.

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 $\sqrt{n}(\bar{X}_n - m)$, 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 1 (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

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

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 $\sqrt{r p} \bar{Z}_r \xrightarrow[r \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \gamma_p^2)$ où $\bar{Z}_r = \frac{1}{r} \sum_{i=1}^r Z_i$ et $\gamma_p^2 = \sigma^2 + 2 \sum_{i=1}^M (1 - \frac{i}{p}) r_X(i) \xrightarrow[p \rightarrow \infty]{\mathcal{P}} \gamma^2$. 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 $p \text{var}(\bar{U}_r^{(p)}) \xrightarrow[n \rightarrow +\infty]{} 0$ avec $\bar{U}_r^{(p)} = \frac{1}{r} \sum_{i=1}^r U_i^{(p)}$. Il ne reste plus qu'à écrire que $\bar{X}_n = \frac{p}{n} \sum_{i=1}^r Z_i + \bar{U}_r^{(p)}$ 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 pour un théorème très utile, à la source de tests classiques d'indépendance pour les processus stationnaires, les tests portemanteaux. Commençons par définir les estimateurs "naturels" des autocovariances et des autocorrélations pour des processus stationnaires:

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

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

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

$$\hat{\rho}_X(p) = \frac{\hat{r}_X(p)}{\hat{r}_X(0)}.$$

Théorème 2 (Théorème central limite pour les autocorrélations de suite de v.a.i.i.d.). *Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire constitué par des v.a.i.i.d. de variance $0 < \sigma_X^2 < \infty$. Alors pour $K_{\max} \in \mathbb{N}$ fixé,*

$$\sqrt{n} (\widehat{r}_X(k))_{1 \leq k \leq K_{\max}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_{K_{\max}}(0, \sigma_X^2).$$

Proof. On commence par supposer que (X_t) est centrée et on définit $\widetilde{r}_X(p) = \frac{1}{n} \sum_{k=1}^{n-p} X_{k+p} X_k$. Il est clair que les (X_t) étant indépendantes, alors les $Y_t = X_t X_{t+p}$ forment une suite stationnaire p -dépendante. De plus $\mathbb{E}(Y_t) = 0$ et $\text{var}(Y_t) = \sigma_X^4$ pour $|p| \geq 1$. On peut donc appliquer le TLC pour les processus p -dépendants et on le fait plus généralement à $Z_t = u_1 X_t X_{t+1} + \dots + u_{K_{\max}} X_t X_{t+K_{\max}}$, où $(u_1, \dots, u_{K_{\max}})$ est un vecteur quelconque de $\mathbb{R}^{K_{\max}}$. Les (Z_t) forment une suite stationnaire p -dépendante avec $\mathbb{E}(Z_t) = 0$ et $\text{var}(Z_t) = \sigma_X^4(u_1^2 + \dots + u_{K_{\max}}^2)$ car $\text{cov}(X_t X_{t+i}, X_t X_{t+j}) = 0$ sauf pour $i = j$ où l'on a σ_X^4 . On applique le TLC pour les processus K_{\max} -dépendants et on a:

$$\sqrt{n-p} \left(\frac{1}{n-K_{\max}} \sum_{i=1}^{n-K_{\max}} Z_t - 0 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_X^4(u_1^2 + \dots + u_{K_{\max}}^2)).$$

Comme p ne dépend pas de n on peut en déduire le même TLC en remplaçant $n-p$ par n par le Lemme de Slutsky. Il est clair que ce résultat s'étend à une série stationnaire d'espérance m en considérant dans ce cas $\widetilde{r}_X(p) = \frac{1}{n} \sum_{k=1}^{n-p} (X_{k+p} - m)(X_k - m)$. Il nous reste à considérer pour $1 \leq p \leq K_{\max}$,

$$\begin{aligned} \widetilde{r}_X(p) - \widehat{r}_X(p) &= \frac{1}{n} \sum_{k=1}^{n-p} (X_{k+p} - m)(X_k - m) - (X_{k+p} - \overline{X}_n)(X_k - \overline{X}_n) \\ &= -\frac{1}{n} (m - \overline{X}_n) \sum_{k=1}^{n-p} (X_{k+p} - m) + (X_k - m) - (m - \overline{X}_n)^2 \\ \mathbb{E} [|\widetilde{r}_X(p) - \widehat{r}_X(p)|] &\leq \frac{n-p}{n} \mathbb{E} [(m - \overline{X}_n)^2] + 2 \sqrt{\mathbb{E} [(m - \overline{X}_n)^2]} \sqrt{\frac{1}{n^2} \sum_{k=1}^{n-p} \sigma_X^2}, \end{aligned}$$

en utilisant l'inégalité de Cauchy-Schwarz. Or $\mathbb{E} [(m - \overline{X}_n)^2] = \sigma_X^2/n$ et après calculs on en déduit donc que $\sqrt{n} \mathbb{E} [|\widetilde{r}_X(p) - \widehat{r}_X(p)|] \xrightarrow[n \rightarrow +\infty]{} 0$ et par Markov

$$\sqrt{n} \mathbb{E} [|\widetilde{r}_X(p) - \widehat{r}_X(p)|] \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} 0.$$

En utilisant tout ce qui précède, on en déduit que:

$$\sqrt{n} (u_1 \widehat{r}_X(p) + \dots + u_{K_{\max}} \widehat{r}_X(K_{\max})) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_X^4(u_1^2 + \dots + u_{K_{\max}}^2)).$$

De ceci on en déduit le résultat final vectoriel. □

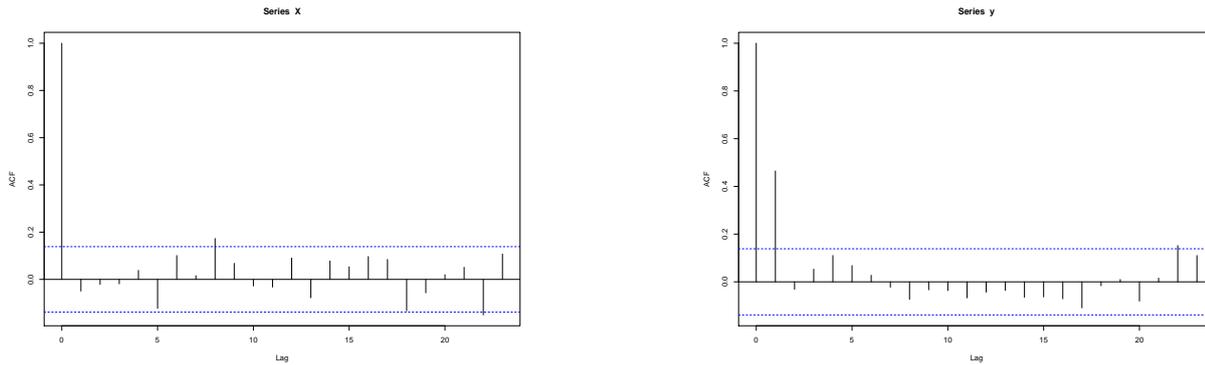
Une conséquence intéressante s'obtient de ce résultat:

Conséquence. *Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire constitué par des v.a.i.i.d. de variance $0 < \sigma_X^2 < \infty$. Alors pour $K_{\max} \in \mathbb{N}$ fixé et avec $I_{K_{\max}}$ la matrice identité de taille K_{\max} ,*

$$\sqrt{n} (\widehat{\rho}_X(k))_{1 \leq k \leq K_{\max}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_{K_{\max}}(0, I_{K_{\max}}).$$

Proof. Sous ces hypothèses, il est bien connu que $\widehat{r}_X(0) \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} \sigma_X^2$. Par conséquent, en utilisant le Lemme de Slutsky et le théorème centrale limite pour les autocovariances on en déduit celui pour les autocorrélations. □

De ce résultat, on en déduit les graphiques appelés "correlogrammes" où sont représentés les autocorrélations empirique en fonction du "lag", et dont voici deux exemples sur la figure suivante (à gauche le correlogramme d'un bruit blanc, à droite celui d'un processus MA[1]):



Sur ces graphes, est représenté l'intervalle de confiance asymptotique à 95%. Ceci donne une première idée de la structure de dépendance du processus (lorsque celui-ci est supposé stationnaire, sinon aucune conclusion ne peut être retirée). Mais, comment aller plus loin qu'une vérification graphique pour attester de la dépendance ou l'indépendance de la série? C'est ce que proposent les deux tests suivants:

Théorème 3 (Comportement asymptotique des tests portemanteaux d'indépendance). *Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire ayant des moments d'ordre 2. On désire tester:*

$$\begin{cases} H_0 : (X_t) \text{ forme une suite de variables indépendantes} \\ H_1 : (X_t) \text{ ne forme pas une suite de variables indépendantes} \end{cases} .$$

On définit deux statistique de test:

- La statistique de Box-Pierce: $\hat{T}_{BP} = n \sum_{k=1}^{K_{\max}} \hat{\rho}_X^2(k);$
- La statistique de Ljung-Box: $\hat{T}_{LB} = n(n+2) \sum_{k=1}^{K_{\max}} \frac{\hat{\rho}_X^2(k)}{n-k}.$

Dans les deux cas, on a

$$\begin{matrix} \hat{T}_{BP} \\ \hat{T}_{LB} \end{matrix} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_{K_{\max}}^2.$$

Proof. Ceci s'obtient directement dans le cas \hat{T}_{BP} en écrivant cette statistique comme la norme au carré du vecteur $\sqrt{n}(\hat{\rho}_X(k))_{1 \leq k \leq K_{\max}}$ et en utilisant le TLC précédent et le fait que l'application norme est continue. Pour \hat{T}_{LB} cela s'obtient avec Slutsky. \square

Ces tests s'appellent "tests du portemanteau", mot anglais signifiant "fourre-tout", sans doute parce que l'hypothèse alternative peut revêtir un nombre considérable de formes...

1.4 Combinaison linéaire de processus stationnaires

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 \mathbb{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 \mathbb{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 démontrer 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 $c \in \mathbb{Z}$, $(X_{t_1-m}, X_{t_1-m+1}, \dots, X_{t_1+m}, X_{t_2-m}, \dots, X_{t_n+m})$ a la 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 $g(X_{t_1-m}, X_{t_1-m+1}, \dots, X_{t_1+m}, X_{t_2-m}, \dots, X_{t_n+m}) = (\sum_{i=-m}^m a_i X_{t_1-i}, \dots, \sum_{i=-m}^m a_i X_{t_n-i})$, la fonction 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. Concernant les moments d'ordre 2, 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 \mathbb{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 \mathbb{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 \mathbb{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 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 stationnarité il est possible d'obtenir un résultat théorique très puissant, qui n'a cependant que peu d'intérêt en pratique:

Théorème 4 (Décomposition de Cramèr-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 \mathbb{L}^2

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

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 Cramèr-Wald n'apporte pas beaucoup de renseignements (voir le cas des processus GARCH).

2 Estimation des tendance et composante saisonnière

2.1 Tendances et saisonnalité

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. si $t \mapsto S(t)$ est une fonction périodique non nulle de période $r > 0$ 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, $(X_t - \widehat{a}(t))_t$, où $\widehat{a}(t)$ 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, $(X_t - \widehat{S}(t))_t$, où $\widehat{S}(t)$ 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.

2.2 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,...).

2.2.1 Estimation semi-paramétrique par régression

On voudrait connaître la tendance et la saisonnalité du processus en supposant connue une trajectoire (X_1, \dots, X_N) .

On suppose que la tendance et la saisonnalité s'écrivent sous une forme connue a priori (les f_i , fonctions quelconques, et r sont supposés connus), soit:

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

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

Notation. • $X = {}^t(X_1, \dots, X_N)$.

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

Le modèle s'écrit alors vectoriellement:

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

Proposition. *On peut estimer les coefficients (a_i) et (s_i) par une régression par moindres carrés en minimisant une distance dans \mathbb{R}^n :*

$$\| X - (a_1 F_1 + \dots + s_{r-1} G_{r-1}) \|^2, \quad \text{et on notant } Z \text{ la matrice } Z = (F_1 \dots F_k G_1 \dots G_{r-1}),$$

1. *si U est un bruit dont on ne connaît pas la variance, on utilise une estimation par moindres carrés ordinaires et:*

$${}^t(\widehat{a}_1, \dots, \widehat{a}_k, \widehat{s}_1, \dots, \widehat{s}_{r-1}) = ({}^t Z Z)^{-1} {}^t Z X,$$

2. *si U est un bruit tel que $\mathbb{E} U^t U = \Sigma$ est une matrice connue, on utilise une estimation par moindres carrés généralisés, et:*

$${}^t(\widehat{a}_1, \dots, \widehat{a}_k, \widehat{s}_1, \dots, \widehat{s}_{r-1}) = ({}^t Z \Sigma^{-1} Z)^{-1} {}^t Z \Sigma^{-1} X.$$

On déduit aisément de ces expressions un premier résultat de convergence pour les estimateurs de la tendance et de la saisonnalité:

Propriété. *Dans le cadre de régression précédent, si (ε_n) est un bruit blanc de variance finie, si la matrice $({}^t Z Z)^{-1}$ tend vers 0 en norme, alors les estimateurs des paramètres sont non biaisés et convergents en probabilité quand $N \rightarrow \infty$.*

Sous les mêmes hypothèses, une condition nécessaire et suffisante de convergence presque sûre est $\max_{1 \leq i \leq k+r-1} |(Z({}^t Z Z)^{-1} Z)_{ii}| \xrightarrow[N \rightarrow \infty]{} 0$. La normalité asymptotique (théorème de la limite centrale vérifié par les estimateurs des paramètres) est également impliquée par cette condition. Cependant, quand on n'est plus dans le cas d'un bruit blanc, les résultats de convergence peuvent être plus complexes, le comportement avec N de la matrice Σ ayant un rôle important...

Exemple.

Soit le cas où la tendance est une constante, où il n'y a pas de saisonnalité et le bruit est stationnaire et admet une matrice de covariance Σ . A-t-on toujours convergence? Et si $\lim_{n \rightarrow \infty} \mathbb{E}(u_0 u_n) = 0$?

Remarque.

- On peut noter que ces estimateurs sont des combinaisons linéaires des X_i .
- Une telle régression nécessite la connaissance a priori des fonctions f_i . Pour ce faire on peut faire différentes hypothèses: on peut considérer que $a(\cdot)$ est un polynôme (typiquement $f_i = t^i$), ou bien que $a(\cdot)$ est un polynôme trigonométrique (typiquement $f_i(t) = \sin(it)$ ou bien $f_i(t) = \cos(it)$). Concrètement, on utilisera plutôt une régression polynomiale lorsque les données semblent prendre une certaine direction, alors que la modélisation par un polynôme trigonométrique permet d'avoir des prédictions qui restent dans le même ordre de grandeur que les données connues. Pour aller un peu plus loin,

on peut même essayer de décomposer la tendance $a(t)$ dans une certaine base (ce qui peut être fait en Fourier ou avec une base d'ondelettes par exemple).

La question qui se pose ensuite est celle de savoir comment choisir le nombre k de fonctions f_i considérées. Cela revient dans le cas polynomial à choisir le degré maximal du polynôme que l'on utilise dans la régression. Il est clair que ce nombre k doit être inférieur à $N - r$, sinon les différents paramètres a_i ne pourront pas être estimés. Une technique efficace pour obtenir "mathématiquement" un choix "optimal" de k est d'utiliser un critère de sélection de modèle. Pour ses propriétés de convergence même dans des cas non gaussiens, le **critère BIC** (Bayesian Information Criterium) est un choix souvent intéressant. Pour k fixé, on calcule

$$BIC(k) = -2\log(\text{Vraisemblance maximisée du modèle à } (k+r-1) \text{ paramètres}) + \log N (k+r-1)$$

et on cherche k qui minimise ce critère. Cela revient, après des approximations, à chercher

$$\hat{k} = \text{Argmax}_{k=0,1,\dots,k_{\max}} \left(\log(\hat{\sigma}_{k+r-1}^2) + \log N k \right), \quad \text{où } \hat{\sigma}_{k+r-1}^2 = \frac{1}{n} \|X - (\hat{a}_1 F_1 + \dots + \hat{s}_{r-1} G_{r-1})\|^2.$$

où k_{\max} est un entier suffisamment grand mais plus petit que $N - r$.

Cas particulier de l'estimation du saisonnier

L'estimation par moindres carrés ordinaires permet d'estimer le saisonnier d'une série chronologique détendancialisée. On suppose connue (X_1, \dots, X_N) , où r est la période (connue) du saisonnier. On écrit $X_t = S(t) + Y_t$, où $S(t) = \sum_{i=1}^{r-1} s_i (g_i(t) - g_r(t))$ pour $t \in T$, avec $g_i(t) = \mathbb{I}_{\{t=i, [r]\}}$ (voir plus haut). Dans ce cas, on a:

$$Z = \begin{pmatrix} 1 & 0 & 0 & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & -1 & \cdot & -1 \\ 1 & 0 & 0 & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & -1 & \cdot & -1 \\ 1 & 0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & -1 & \cdot & -1 \end{pmatrix}, \quad {}^t Z \cdot Z = N \begin{pmatrix} 2 & 1 & \cdot & 1 \\ 1 & 2 & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & 2 \end{pmatrix},$$

et

$$({}^t Z \cdot Z)^{-1} = \frac{1}{rN} \begin{pmatrix} r-1 & -1 & \cdot & -1 \\ -1 & r-1 & \cdot & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & \cdot & r-1 \end{pmatrix}.$$

On montre alors que dans le cadre d'une estimation par moindres carrés ordinaires:

$$\hat{s}_i = \frac{1}{N} \sum_{k=1}^N X_{i+r(k-1)} - \frac{1}{rN} \sum_{k=1}^{rN} X_k, \quad \text{pour } i = 1, \dots, r-1.$$

Exercice.

Montrer que si $\lim_{n \rightarrow \infty} \mathbb{E} u_0 u_n = 0$, où $u = (u_k)_k$ est un bruit gaussien stationnaire, il y a convergence en probabilité de cet estimateur.

2.2.2 Estimation non-paramétrique de la tendance

On supposera donc ici que la série n'admet pas de composante saisonnière, et que l'on a $X_t = a(t) + u_t$ pour $t \in \{t_1, \dots, t_N\}$. Plutôt que de poser a priori un modèle pour la fonction a comme cela a été fait avec la régression, on peut estimer directement cette fonction avec une méthode non-paramétrique. On présente ici deux types de méthodes:

- Méthode d'estimation par noyau.

Définition. On appelle noyau $K : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable telle que $\int_{\mathbb{R}} K(t)dt = 1$ et $K(0) > 0$.

L'idée de la méthode par noyau est que pour $x_0 \in R$ et $h > 0$ dit taille de fenêtre, $\frac{1}{h} K\left(\frac{x - x_0}{h}\right)$ converge vers une masse de Dirac lorsque $h \rightarrow 0$, ce qui permet sous certaines conditions d'avoir $\int_{\mathbb{R}} \frac{1}{h} K\left(\frac{x - x_0}{h}\right) f(x) dx \rightarrow f(x_0)$ quand $h \rightarrow 0$, pour toute fonction f . Aussi peut-on définir un estimateur de la tendance a par:

$$\hat{a}_{N,h}(t) = \frac{\frac{1}{Nh} \sum_{j=1}^N X_j K\left(\frac{t-t_j}{h}\right)}{\frac{1}{Nh} \sum_{j=1}^N K\left(\frac{t-t_j}{h}\right)}.$$

Sous certaines hypothèses sur la manière dont h converge vers 0 en fonction de N , on peut montrer que $\hat{a}_{N,h}(t) \rightarrow a(t)$ pour tout t lorsque a est suffisamment régulière.

On peut également estimer de façon automatique une taille de fenêtre \hat{h}_N adaptée aux données et optimale en un certain sens: pour ce faire on utilise par exemple le principe de la validation croisée, c'est-à-dire que pour $h > 0$ fixé et pour chaque $i = 1, \dots, N$, l'on calcule $\hat{a}_{N,h}^{(i)}(i)$ à partir de l'échantillon $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_N)$ et l'on compare $\hat{a}_{N,h}^{(i)}(i)$ à X_i qui est la valeur obtenue en $t = i$. Ainsi on pourra choisir

$$\hat{h}_N = \operatorname{Argmin}_{h>0} \sum_{i=1}^N (\hat{a}_{N,h}^{(i)}(i) - X_i)^2 \implies \hat{a}_{N,\hat{h}_N}(t) \text{ estimateur de } a.$$

- Régressions localisées

Pour estimer la tendance, il est aussi possible d'utiliser des **régressions localisées** de type Spline, Loess ou Lowess. Celles-ci s'obtiennent en fixant une taille de fenêtre et on fait une régression polynomiale (pour les Splines, de degré 3) ou linéaire mais pondérées (pour Loess ou Lowess) de X dans la fenêtre que l'on fait glisser. En fait ce sont également des estimations non-paramétriques, au sens où elles ne supposent pas connues les fonctions pouvant composer la tendance (comme dans le cas de la régression), mais juste une certaine régularité de la fonction tendance. Plus précisément, considérons l'approche par moyenne mobile: en X_t , une idée pour approcher $a(t)$ serait de supposer que a ne change pas trop autour de t et que l'on peut moyenniser sur les valeurs de temps proche de t .

Par exemple, on pourra considérer l'exemple important de la moyenne mobile élémentaire (dite uniforme):

$$\hat{a}_\ell(t) = \frac{1}{2\ell + 1} \sum_{i=-\ell}^{\ell} X_{t+i} \quad \text{pour } t \in \{1, \dots, N\}.$$

L'intérêt réside dans le fait de faire ainsi baisser la variance du bruit u_t en moyennant, ce qui se montre facilement dans le cas où le bruit u est un bruit blanc. Mais il faut choisir ℓ : trop grand, on lissera trop jusqu'à n'obtenir qu'une constante pour tout t , trop petit, on ne fera guère mieux que d'estimer $a(t)$ par X_t , et la fonction a sera donc très irrégulière.

Une idée pour dépasser cela sera de considérer une moyenne mobile pondérée dans laquelle on n'accorde pas le même poids à tous les X_i autour de X_t , mais plutôt un poids décroissant en fonction de la distance entre i et t . Aussi considérera-t-on différentes formes de pondérations dépendant d'un paramètre que nous noterons β :

$$\hat{a}_\beta(t) = \frac{\sum_{i=-t+1}^{N-t} W_\beta(|t-i|) X_{t+i}}{\sum_{i=-t+1}^{N-t} W_\beta(|t-i|)},$$

les W_β étant des "poids" (à valeurs donc dans $[0, \infty[$ et décroissants en fonction de $|t-i|$) pouvant prendre plusieurs formes, par exemple:

- Des poids exponentiels: $W_\beta(x) = \beta^x$ et $\beta \in [0, 1]$ (on notera que pour $\beta = 1$ on trouve $\hat{a}_\beta(t) = \bar{X}_N$ pour tout t ;
- Des poids tricubes: $W_\beta(x) = \left(1 - \left(\frac{x}{\beta N}\right)^3\right)^3 \mathbb{I}_{x \leq \beta N}$ et $\beta \in [0, 1]$, βN mesurant une taille de fenêtre.

On appellera un tel estimateur un estimateur LOWESS (LOcally WEighted Scatterplot Smoothing). Le choix de β peut aussi se faire par validation croisée, comme cela est présenté ci-dessus pour obtenir $\hat{\beta}_N = \text{Argmin}_\beta \sum_{i=1}^N (\hat{a}_\beta^{(i)}(i) - X_i)^2$.

Notons que d'une manière générale, la généralité des estimateurs non-paramétriques se traduit souvent pas une moins bonne vitesse de convergence de \hat{a}_N vers a que les estimateurs paramétriques dans le cas où le vrai modèle est paramétrique. En général, ils ne permettent pas des prédictions, mais peuvent inspirer des méthodes de prédictions (voir plus loin). Pour détendancionaliser une série et ainsi permettre une identification de la partie aléatoire de la série, ils demeurent des outils précieux. Enfin, pour estimer la partie tendance quand une saisonnalité est présente, il est d'abord possible d'appliquer une moyenne mobile uniforme dont la longueur totale est un multiple de T , période de la saisonnalité. Cela permet une première estimation de la tendance. Sur la série détendancionalisée, on pourra alors estimer la saisonnalité. Puis, en désaisonnalisant, on pourra à nouveau estimer la tendance par une des techniques précédentes. Cette suite de procédures peut s'enchaîner pour arriver à une forme de stabilité. C'est ce que propose notamment la procédure `st1` en R.

3 Les processus ARMA, GARCH et longue mémoire

3.1 Processus ARMA

Les processus AR (Auto-Regressive), MA (Moving Average) et ARMA ont été introduits par Slutsky (1927) et Yule (1921) dans le cadre de modélisation de cycles en économie.

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 $(a_i)_{0 \leq p} \in \mathbb{R}^{p+1}$, $P(B) \cdot X = (\sum_{j=0}^p a_j X_{n-j})_{n \in \mathbb{Z}}$.
- 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:

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

- 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:

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

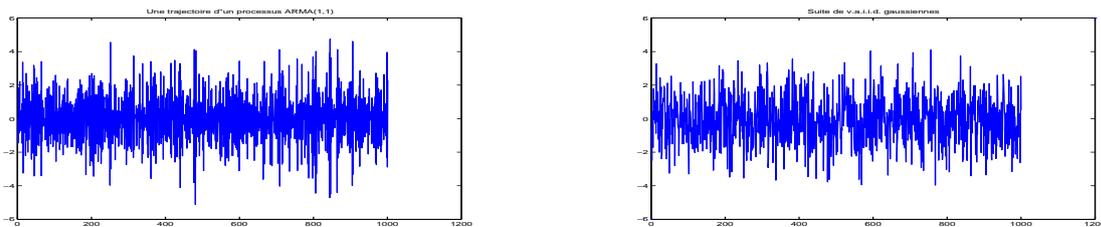
- Un processus $X = (X_n)_{n \in \mathbb{Z}}$ tel que $P(B) \cdot X = Q(B) \cdot \varepsilon$ est appelé un processus ARMA(p, q). Ceci revient à écrire qu'il existe $(b_1, \dots, b_p) \in \mathbb{R}^p$ et $(c_1, \dots, c_q) \in \mathbb{R}^q$ tels que:

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

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 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

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

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:

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

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) est un processus linéaire causal ayant pour densité spectrale:

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

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

3.2 Les processus GARCH

Les processus ARCH(p) ont été introduits par Engle (1982) et généralisés en GARCH(p, q) par Bollerssev (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 qu’il existe une solution stationnaire à l’équation

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

On peut se demander quelles sont les conditions qui pourraient permettre d’obtenir des processus stationnaires vérifiant 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 $\mathbb{E} [\log(a_1 \varepsilon_0^2 + b_1)] < 0$ et il vérifie alors (3).

Proof. On note la fonction $\alpha(x) = a_1 x^2 + b_1$. Alors $\sigma_t^2 = a_0 + \alpha(\varepsilon_{t-1}) \sigma_{t-1}^2$ pour tout $t \in \mathbb{Z}$. En itérant, pour $m \geq 1$, on obtient:

$$\sigma_t^2 = a_0 \left(1 + \sum_{j=1}^{m-1} \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-j}) \right) + \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-m}) \sigma_{t-m}^2.$$

Supposons que $\mathbb{E} [\log(a_1 \varepsilon_0^2 + b_1)] = \mathbb{E} [\log(\alpha(\varepsilon_0^2))] < 0$. Par conséquent, d’après la loi forte des grands nombres:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \log(\alpha(\varepsilon_{t-i}^2)) &\xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E} [\log(\alpha(\varepsilon_0^2))] \\ \implies \left(\alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-n}) \right)^{1/n} &\xrightarrow[n \rightarrow +\infty]{p.s.} \exp \left(\mathbb{E} [\log(\alpha(\varepsilon_0^2))] \right) < 1 \end{aligned}$$

On peut alors utiliser le critère de Cauchy de convergence des séries numériques à termes positifs: si $u_n^{1/n} \xrightarrow[n \rightarrow +\infty]{} \ell < 1$ alors $\sum u_n$ converge. Par conséquent:

$$a_0 \left(1 + \sum_{j=1}^n \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-j}) \right) \xrightarrow[n \rightarrow +\infty]{p.s.} s^2(t) = a_0 \left(1 + \sum_{j=1}^{\infty} \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-j}) \right),$$

limite existant presque sûrement, ce qui implique également que $\alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-n}) \xrightarrow[n \rightarrow +\infty]{p.s.} 0$. Ainsi, pour obtenir une solution stationnaire (X_t) à l’équation de récurrence, ce qui induit que (σ_t^2) est également stationnaire, alors $\alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-n}) \sigma_{t-n}^2 \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ et

$$X_t = \sqrt{a_0} \varepsilon_t \left(1 + \sum_{j=1}^{\infty} \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-j}) \right)^{1/2} \quad \text{presque sûrement.} \quad (3)$$

On peut montrer l’unicité de cette solution: si $X_t = \varepsilon_t \sigma'(t)$ avec $\sigma_t'^2 = a_0 + \alpha(\varepsilon_{t-1}) \sigma_{t-1}'^2$ stationnaire, en utilisant l’itération précédente, on a $\sigma_t^2 - \sigma_t'^2 = \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-n}) (\sigma_{t-n}^2 - \sigma_{t-n}'^2) \xrightarrow[n \rightarrow +\infty]{p.s.} 0$.

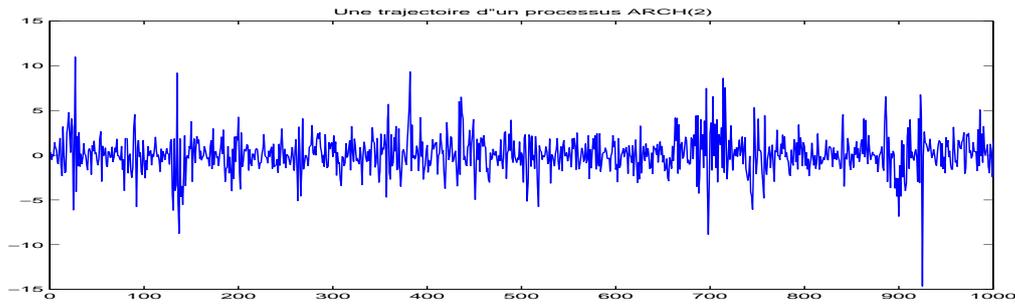
Si $\mathbb{E} [\log(a_1 \varepsilon_0^2 + b_1)] = \mathbb{E} [\log(\alpha(\varepsilon_0^2))] > 0$, avec $\sigma_t^2 \geq a_0 \left(1 + \sum_{j=1}^n \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-j}) \right)$ pour tout $n \in \mathbb{N}$, en utilisant également la règle de Cauchy on montre que $\sum_{j=1}^n \alpha(\varepsilon_{t-1}) \alpha(\varepsilon_{t-2}) \times \dots \times \alpha(\varepsilon_{t-j}) \xrightarrow[n \rightarrow +\infty]{} +\infty$ donc il n’existe pas de solution stationnaire.

Si $\mathbb{E}[\log(a_1\varepsilon_0^2 + b_1)] = \mathbb{E}[\log(\alpha(\varepsilon_0^2))] = 0$, on a $\sigma_t^2 \geq a_0 \left(1 + \sum_{j=1}^n \alpha(\varepsilon_{t-1})\alpha(\varepsilon_{t-2}) \times \cdots \times \alpha(\varepsilon_{t-j})\right)$ pour tout $n \in \mathbb{N}$. Pour qu'une solution stationnaire existe, il faut que la série converge p.s., donc que $\alpha(\varepsilon_{t-1})\alpha(\varepsilon_{t-2}) \times \cdots \times \alpha(\varepsilon_{t-j}) \xrightarrow[n \rightarrow +\infty]{} 0$, soit $\sum_{i=1}^n \log(\alpha(\varepsilon_{t-i})) \xrightarrow[n \rightarrow +\infty]{p.s.} -\infty$. Mais $(\log(\alpha(\varepsilon_{t-i})))_{i \geq 1}$ est une suite de variables aléatoires centrées, donc $\sum_{i=1}^n \log(\alpha(\varepsilon_{t-i}))$ est une marche aléatoire centrée qui ne peut converger p.s. vers $-\infty$ (par exemple grâce au TLC on obtient $\mathbb{P}(\sum_{i=1}^n \log(\alpha(\varepsilon_{t-i})) \geq 0) \xrightarrow[n \rightarrow +\infty]{} 1/2$). Donc il n'existe pas de solution stationnaire dans ce cas. \square

Conséquence. *Plusieurs conséquences découlent de ce résultat:*

1. *Tout d'abord, si $a_0 = 0$, l'unique solution stationnaire est $X_t = 0$ pour tout $t \in \mathbb{Z}$. On verra également que la variance de X_t est proportionnelle à a_0 .*
2. *Si le bruit blanc fort $(\varepsilon_t)_t$ est centré et de variance 1, alors par l'Inégalité de Jensen, $\mathbb{E}[\log(a_1\varepsilon_0^2 + b_1)] \leq \log(\mathbb{E}[a_1\varepsilon_0^2 + b_1]) \leq \log(a_1 + b_1)$. On en déduit donc que si $a_1 + b_1 < 1$ alors la solution stationnaire existe toujours.*

Voici un exemple d'une trajectoire du processus ARCH(2) avec $a_0 = 1$, $a_1 = 0.3$ et $a_2 = 0.5$:



Propriété. *Un processus GARCH(1, 1) vérifiant l'équation (2) est stationnaire d'ordre 2 si et seulement si $a_1 + b_1 < 1$.*

Proof. \implies Si X a un moment d'ordre 2 et est stationnaire d'ordre 2 alors $\mathbb{E}[X_t^2] = \mathbb{E}[\sigma_t^2] = a_0 + a_1 \mathbb{E}[X_{t-1}^2] + b_1 \mathbb{E}[\sigma_{t-1}^2]$, d'où l'équation $\mathbb{E}[X_0^2] = a_0 + a_1 \mathbb{E}[X_0^2] + b_1 \mathbb{E}[X_0^2]$, soit:

$$\mathbb{E}[X_0^2] = \frac{a_0}{1 - a_1 - b_1} \implies a_1 + b_1 < 1.$$

\Leftarrow Si $a_1 + b_1 < 1$ et $\mathbb{E}(\varepsilon_0^2) = 1$ alors $\mathbb{E}[\log(a_1\varepsilon_0^2 + b_1)] < 0$ et X est donc stationnaire strict. On a donc X qui s'écrit sous la forme (3). Il sera stationnaire d'ordre 2 si $\mathbb{E}[X_0^2] < \infty$. Or d'après (3) et en utilisant l'indépendance des (ε_t) ,

$$\begin{aligned} \mathbb{E}[X_t^2] &= a_0 \mathbb{E}[\varepsilon_t^2] \mathbb{E}\left[1 + \sum_{j=1}^{\infty} \alpha(\varepsilon_{t-1})\alpha(\varepsilon_{t-2}) \times \cdots \times \alpha(\varepsilon_{t-j})\right] \\ &= a_0 \left(1 + \sum_{j=1}^{\infty} \mathbb{E}[\alpha(\varepsilon_{t-1})] \mathbb{E}[\alpha(\varepsilon_{t-2})] \times \cdots \times \mathbb{E}[\alpha(\varepsilon_{t-j})]\right) \\ &= a_0 \left(1 + \sum_{j=1}^{\infty} (a_1 + b_1)^j\right) = \frac{a_0}{1 - a_1 - b_1}. \end{aligned}$$

\square

Ce dernier résultat se généralise au cas des processus GARCH(p, q):

Propriété. *Un processus GARCH(p, q) vérifiant l'équation (2) est stationnaire d'ordre 2 si et seulement si $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1$.*

La preuve de ce résultat est très récente en ce qui concerne la condition nécessaire (elle n'a été montrée qu'en 2000...). Pour la condition suffisante on a:

Proof. \implies Si X a un moment d'ordre 2 et est stationnaire d'ordre 2 alors $\mathbb{E}[X_t^2] = \mathbb{E}[\sigma_t^2] = a_0 + a_1 \mathbb{E}[X_{t-1}^2] + \dots + a_p \mathbb{E}[X_{t-p}^2] + b_1 \mathbb{E}[\sigma_{t-1}^2] + \dots + b_q \mathbb{E}[\sigma_{t-q}^2]$, d'où l'équation $\mathbb{E}[X_0^2] = a_0 + a_1 \mathbb{E}[X_0^2] + \dots + a_p \mathbb{E}[X_0^2] + b_1 \mathbb{E}[X_0^2] + \dots + b_q \mathbb{E}[X_0^2]$, soit:

$$\mathbb{E}[X_0^2] = \frac{a_0}{1 - a_1 - \dots - a_p - b_1 - \dots - b_q} \implies \sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1.$$

□

De ceci, on peut déduire les propriétés suivantes:

Propriété. *Soit $X = (X_k)_{k \in \mathbb{Z}}$ un processus GARCH(p, q) stationnaire d'ordre 2. Alors $r(k) = \text{cov}(X_0, X_k) = 0$ 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 GARCH(p, q) ne pourra pas être identifié à partir de ses covariances ou de sa densité spectrale. 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 GARCH(p, q) stationnaire d'ordre 2 tel que $\mathbb{E}(\varepsilon^4) < \infty$. Alors $Y = (X_k^2)_{k \in \mathbb{Z}}$ est un processus ARMA($\max(p, q), p$) faible non centré.*

Proof. On considère $v_t = X_t^2 - \sigma_t^2 = (\varepsilon_t^2 - 1)\sigma_t^2$. La série (v_t) est un bruit blanc faible car (v_t) est stationnaire, $\mathbb{E}(v_t) = 0$ et pour $t > 0$, $\text{cov}(v_0, v_t) = \text{cov}((\varepsilon_0^2 - 1)\sigma_0^2, (\varepsilon_t^2 - 1)\sigma_t^2) = 0$ car $(\varepsilon_t^2 - 1)$ est indépendant des 3 autres termes et d'espérance nulle. De plus, on montre que $X_t^2 = v_t + (a_0 + \sum_{i=1}^p a_i X_{t-i}^2) + \sum_{j=1}^q b_j (X_{t-j}^2 - v_{t-j}) = a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j (X_{t-j}^2 - v_{t-j}) + v_t - \sum_{j=1}^q b_j v_{t-j}$, d'où (X_t^2) ARMA faible non centré. □

Conséquence. *On en déduit qu'un processus GARCH(p, q) n'est pas un bruit blanc fort et donc qu'il ne peut pas être un processus gaussien.*

3.3 Processus affine causaux à mémoire infinie

Des deux types de processus étudiés précédemment, deux extensions "naturelles" peuvent être considérées:

- Les processus AR(∞): $X_t = \varepsilon_t + \sum_{i=1}^{\infty} a_i X_{t-i}$ pour $t \in \mathbb{Z}$;
- Les processus ARCH(∞): $X_t = \varepsilon_t \sigma_t$ avec $\sigma_t^2 = a_0 + \sum_{i=1}^{\infty} a_i X_{t-i}^2$ pour tout $t \in \mathbb{Z}$.

Il est possible de généraliser encore de la manière suivante. Tout d'abord, dans la suite on note \mathbb{R}^{∞} l'ensemble des suites $(x_n)_{n \in \mathbb{N}^*}$ telles que (x_n) n'a qu'un nombre fini de termes non nuls. Ceci amène à la définition suivante:

Definition (0). *Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc tel que $\mathbb{E}[|\varepsilon_0|^r] < \infty$, où $r \geq 1$. On dit que $(X_t)_{t \in \mathbb{R}}$ est un processus affine causal s'il existe $F : \mathbb{R}^{\infty} \rightarrow \mathbb{R}$ et $M : \mathbb{R}^{\infty} \rightarrow \mathbb{R}$ deux fonctions telles que*

$$X_t = F((X_{t-k})_{k \geq 1}) + \varepsilon_t M((X_{t-k})_{k \geq 1}) \quad \text{pour tout } t \in \mathbb{Z}. \quad (4)$$

Naturellement on se pose la question de connaître les conditions pour que (X_t) existe, soit stationnaire, ait des moments... Pour cela, on utilisera un résultat encore plus général sur les séries à mémoire infinie démontré par Doukhan et Wintenberger (2007):

Proposition. *On suppose que (X_t) est définie par (4). On suppose que F et G ont toutes leurs dérivées partielles qui existent presque partout sur \mathbb{R}^∞ et on note*

$$\alpha_i(F) = \sup_{x \in \mathbb{R}^\infty} \left| \frac{\partial}{\partial x_i} F(x) \right| \quad \text{et} \quad \alpha_i(M) = \sup_{x \in \mathbb{R}^\infty} \left| \frac{\partial}{\partial x_i} M(x) \right|.$$

Alors (X_t) est une série chronologique stationnaire stricte, avec un moment d'ordre r (donc $\mathbb{E}[|X_0|^r] < \infty$) qui s'écrit sous une forme causale $X_t = H((\varepsilon_{t-k})_{k \geq 0})$ avec H une unique fonction telle que $H : \mathbb{R}^\infty \rightarrow \mathbb{R}$ dès que:

$$\sum_{i=1}^{\infty} \alpha_i(F) + (\mathbb{E}[|\varepsilon_0|^r])^{1/r} \sum_{i=1}^{\infty} \alpha_i(M) < 1. \quad (5)$$

Exemple.

- Pour un processus AR(∞), cela revient à écrire: $\sum_{i=1}^{\infty} |a_i| < 1$.
- Pour un processus ARCH(∞), cela revient à écrire: $(\mathbb{E}[|\varepsilon_0|^r])^{1/r} \sum_{i=1}^{\infty} \sqrt{a_i} < 1$.

3.4 Processus longue mémoire

Definition (0). Soit $X = (X_k)_{k \in \mathbb{Z}}$ une série chronologique stationnaire d'ordre 2. Alors on dit que X est un processus longue mémoire si $\sum_{k \in \mathbb{Z}} |r_X(k)| = \infty$

Exemple: $X = (X_k)_{k \in \mathbb{Z}}$ avec $X_k = X_0$ pour tout $k \in \mathbb{Z}$.

Consequences: Si X est un processus longue mémoire:

- La densité spectrale de X , si elle existe, n'est pas continue sur $[-\pi, \pi]$;
- Cette définition n'est pas vraiment satisfaisante car elle demande nécessairement d'avoir un moment d'ordre 2 pour mesurer une mémoire infinie. De plus, une densité spectrale avec une discontinuité en une fréquence mais des limites à droite et à gauche, ne caractérise pas le même type de "longue" mémoire qu'un pôle dans la densité spectrale.

On va donc un peu plus "quantifier" la première définition:

Definition (1). $X = (X_k)_{k \in \mathbb{Z}}$ est un processus stationnaire d'ordre 2 à longue mémoire lorsque

$$r_X(k) = |k|^{-D} L(|k|), \quad \text{pour } k \neq 0, \quad \text{avec}$$

- $D \in]0, 1[$ le paramètre de longue-mémoire
- L une fonction à variations lentes en ∞ , i.e.: $\forall t > 0, \lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$

Contre-exemple: $X = (X_k)_{k \in \mathbb{Z}}$ avec $X_k = X_0$ pour tout $k \in \mathbb{Z}$ n'est plus un processus longue mémoire suivant cette définition.

Une autre définition est possible à partir de la densité spectrale:

Definition (2). $X = (X_k)_{k \in \mathbb{Z}}$ est un processus stationnaire d'ordre 2 à longue mémoire lorsque

$$f(\lambda) = |\lambda|^{D-1} M\left(\frac{1}{|\lambda|}\right) \quad , \quad \text{pour } \lambda \in [-\pi, 0[\cup]0, \pi], \quad \text{avec}$$

- $D \in]0, 1[$ le paramètre de longue-mémoire;
- M une fonction à variations lentes en ∞ .

Remarque. • La définition (1) implique la définition (2) (Théorème abélien);

- La définition (2) associée à l'hypothèse d'une décroissance de la fonction $r(\cdot)$ implique la définition (1) (Théorème taubérien).

Pour permettre une définition de la longue mémoire qui ne soit pas liée aux moments d'ordre 2, voici une autre possible définition [Samorodnitsky and Taqqu (1994)]. On définit d'abord l'auto-similarité:

Definition. On dit que $X = (X_t)_{t \in \mathbb{R}}$, un processus H -auto-similaire à accroissements stationnaires lorsque:

$$\begin{cases} (X_{cs})_s \stackrel{\mathcal{L}}{\sim} c^H (X_s)_s \quad \text{pour tout } c > 0 \\ (X_{t+s} - X_t)_t \quad \text{est un processus stationnaire pour tout } s \in \mathbb{R} \end{cases}$$

On peut alors énoncer une nouvelle définition de la longue mémoire qui ne soit plus relative aux moments d'ordre 2:

Definition (3). $X = (X_k)_{k \in \mathbb{Z}}$ est un processus stationnaire à longue mémoire lorsque $X_k = Y_{k+1} - Y_k$, pour $k \in \mathbb{Z}$, avec (Y_k) un processus H -auto-similaire à accroissements stationnaires et $H \in]1/2, 1[$.

Cette définition est cependant beaucoup moins générale que les précédentes puisqu'elle demande l'autosimilarité de Y ce qui est une condition forte. On remarque après quelques calculs que si Y admet des moments d'ordre 2, alors l'autocovariance de X satisfait la condition $r_X(k) \sim C |k|^{2H-2}$ pour $k \rightarrow \infty$ et ainsi la définition (2) est satisfaite avec $D = 2 - 2H$.

3.4.1 Deux exemples célèbres de processus longue mémoire

Bruits gaussiens fractionnaires

Definition (Kolmogorov, 1940, Lévy, 1965). On dit que $Y = (Y_t)_{t \in \mathbb{R}}$ est un mouvement brownien fractionnaire de paramètres $\sigma^2 > 0$ et $H \in (0, 1]$ lorsque Y est un processus gaussien centré à accroissements stationnaires tel que pour tout $t \in \mathbb{R}$, $\mathbb{E}[Y_t^2] = \sigma^2 |t|^{2H}$.

Remarque. La famille des mouvements browniens fractionnaires correspond à celle des processus gaussiens autosimilaires à accroissements stationnaires. Notons que pour $H = 1/2$, on retombe sur le classique mouvement brownien de Wiener.

Cela nous permet de définir une famille de processus long-mémoire (voir définition (3))

Definition. Soit Y un mouvement brownien fractionnaire de paramètres $\sigma^2 > 0$ et $H \in (0, 1]$. Alors $X = (Y_{t+1} - Y_t)_{t \in \mathbb{Z}}$ est appelé un bruit gaussien fractionnaire de paramètres $\sigma^2 > 0$ et $H \in (0, 1]$.

On a alors la propriété suivante:

Propriété. Si X est un bruit gaussien fractionnaire de paramètres $\sigma^2 > 0$ et $H \in]1/2, 1]$ alors $r_X(k) \sim \sigma^2 H(2H - 1) |k|^{2H-2}$ pour $|k| \rightarrow \infty$: X est un processus longue mémoire. Lorsque $H \in]0, 1/2]$, X est un processus courte mémoire, c'est-à-dire que $\sum |r_X(k)| = \infty$.

Simulations d'une trajectoire (X_1, \dots, X_n) d'un bruit gaussien fractionnaire:

1. Une idée naturelle: utiliser la décomposition de Cholesky de $\Sigma = (r(|j - i|))_{1 \leq i, j \leq n}$, matrice de Toeplitz, qui s'écrira sous la forme $\Sigma = R R'$. Il suffira alors de calculer $X = R Z$ où Z est un échantillon de v.a. i.i.d. gaussiennes standardes.
2. Un meilleur choix: plonger Σ dans une matrice circulant et utiliser la décomposition spectrale (en terme de valeurs propres et matrice de passage) d'une matrice circulante \implies décomposition spectrale d'une "racine" de Σ notée $\Sigma^{1/2}$ comme matrice extraite de taille n d'une racine de la matrice circulante, et calcul de $X = \Sigma^{1/2} Z$ où Z est un échantillon de v.a. i.i.d. gaussiennes standardes.

Processus FARIMA(p, d, q)

Definition (Granger et Joyeux, 1980). Si $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc, on dira que $X = (X_t)_{t \in \mathbb{Z}}$ est un processus FARIMA(p, d, q), où $d \in]-1/2, 1/2[$ et $(p, q) \in \mathbb{N}^2$ lorsque

$$\iff (1 - B)^d P(B)(X) = Q(B)(\varepsilon) \quad \text{avec } P \in \mathbb{R}_p[X], Q \in \mathbb{R}_q[X], \text{ deux polynômes premiers entre eux et } P(z) \neq 0 \text{ pour tout } |z| \leq 1;$$

$$\iff X_k = \sum_{j=0}^{\infty} \left(\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \right) \eta_{k-j} \quad \text{avec } (\eta_t) \text{ un processus ARMA}(p, q) \text{ stationnaire, soit:}$$

$$\eta_k + \theta_1 \eta_{k-1} + \dots + \theta_p \eta_{k-p} = \varepsilon_k + \phi_1 \varepsilon_{k-1} + \dots + \phi_q \varepsilon_{k-q}.$$

Propriété. De la seconde écriture, comme $\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \sim \frac{1}{\Gamma(d)} j^{d-1}$ quand $j \rightarrow \infty$ (pour cela on utilise le développement asymptotique $\Gamma(x) \sim \sqrt{2\pi} e^{1-x} (x-1)^{x-1/2}$ pour $x \rightarrow \infty$), on peut montrer que $\sum_{j=0}^{\infty} \left| \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \right|^2 < \infty$ pour tout $d < 0.5$, on obtient que X est un bien un processus linéaire du second ordre centré stationnaire.

Conséquence: Grâce aux résultats précédents obtenus sur la densité spectrale on obtient que celle-ci existe sur $[-\pi, 0[\cup]0, \pi]$ et est telle que:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{Q(e^{i\lambda})}{P(e^{i\lambda})} \right|^2 \frac{1}{|1 - e^{i\lambda}|^{2d}} \quad \text{pour } \lambda \in [-\pi, 0[\cup]0, \pi].$$

Ceci permet de montrer facilement en utilisant la définition (2) qu'un processus FARIMA est bien un processus longue mémoire si $d \in]0, 1/2[$ car:

Propriété. Pour X un processus FARIMA(p, d, q), X est un processus stationnaire longue mémoire lorsque $0 < d < 1/2$ car alors $f(\lambda) \sim C |\lambda|^{-2d}$ quand $\lambda \rightarrow 0$. Si $d \leq 0$, X est un processus à mémoire courte.

Simulations d'une trajectoire (X_1, \dots, X_n) d'un processus FARIMA:

1. Meilleure idée dans le cas gaussien (un FARIMA n'est pas forcément un processus gaussien): utiliser la méthode la matrice circulante...
2. Dans le cas non gaussien, utiliser une troncature de $\sum_{j=0}^{\infty} \left(\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \right) \eta_{k-j}$, soit $X_k = \sum_{j=0}^M \left(\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \right) \eta_{k-j}$ avec M grand, par exemple $M = 10000$.

3.4.2 Des processus affines causaux longue mémoire

Soit Z un processus FARIMA(0, d , 0). De sa définition, on peut aussi écrire que:

$$(1 - B)^d Z = \varepsilon \implies Z_t = \varepsilon_t - \sum_{j=1}^{\infty} \alpha_j(d) Z_{t-j} \quad \text{pour tout } t \in \mathbb{Z},$$

où $\alpha_j(d) = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}$, avec $\alpha_j(d) \sim \frac{1}{\Gamma(-d)} j^{-d-1}$ quand $j \rightarrow \infty$. Un FARIMA(0, d , 0) est donc un cas particulier de processus AR(∞) à longue mémoire, et donc un processus affine causal stationnaire à longue mémoire.

Plus généralement, on peut poser

$$X_t = M((X_{t-k})_{k \geq 1}) \varepsilon_t + \sum_{j=1}^{\infty} a_j X_{t-j} \quad \text{pour tout } t \in \mathbb{Z}, \quad (6)$$

avec $M : \mathbb{R}^{\infty} \rightarrow \mathbb{R}$ une fonction différentiable sur \mathbb{R}^{∞} et la suite (a_j) telle que:

$$a_j = L(j) j^{-\kappa} \quad \text{avec } \kappa \in]1, 3/2[\text{ et } L(\cdot) \text{ une fonction à variations lentes.}$$

Si on suppose de plus que pour $r \geq 2$,

$$\sum_{j=1}^{\infty} a_j + (\mathbb{E}[|\varepsilon_0|^r])^{1/r} \sum_{j=1}^{\infty} \alpha_j(M) < 1,$$

avec $\alpha_i(M) = \sup_{x \in \mathbb{R}^{\infty}} \left| \frac{\partial}{\partial x_i} M(x) \right|$, alors il existe une unique solution (X_t) stationnaire de (6), avec des moments d'ordre r , et longue mémoire de paramètre.

4 Identification d'un processus causal affine stationnaire

On s'intéresse ici à l'identification de la partie bruit d'un processus causal affine stationnaire. 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 une série chronologique, 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,...).

4.1 Comportement asymptotique des processus causaux affines

On considère donc ici des processus affines causaux qui pourront être aussi bien des processus ARMA, GARCH, AR(∞), ARCH(∞),...

Théorème 5 (Théorème central limite pour les processus causaux affines stationnaires).

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus affine causal stationnaire défini par (4). On suppose également qu'il existe $C > 0$ et $\rho \in [0, 1[$ tels que

$$\max \{ \alpha_i(F), \alpha_i(M) \} \leq C \rho^i \quad \text{pour tout } i \in \mathbb{N}. \quad (7)$$

Pour $k \in \mathbb{N}^*$, on note $p = (p_1, \dots, p_k) \in \mathbb{N}^k$ et $|p| = p_1 + \dots + p_k$. On suppose de plus que $\mathbb{E}[|X_0|^{2|p|}] < \infty$ et on définit

$$\begin{aligned} m^{(p)} &= \mathbb{E}[X_1^{p_1} X_2^{p_2} \dots X_k^{p_k}], \quad \text{le moment d'ordre } p \\ \widehat{m}_N^{(p)} &= \frac{1}{N} \sum_{i=1}^{N-k+1} X_i^{p_1} X_{i+1}^{p_2} \dots X_{i+k-1}^{p_k}, \quad \text{le moment empirique d'ordre } p. \end{aligned}$$

$$1. \text{ Avec } \sigma^2(p) = \sum_{\ell \in \mathbb{Z}} \left(\mathbb{E}(X_1^{p_1} X_2^{p_2} \dots X_k^{p_k} X_{\ell+1}^{p_1} \dots X_{\ell+k}^{p_k}) - (m^{(p)})^2 \right),$$

$$\sqrt{n} (\widehat{m}_N^{(p)} - m^{(p)}) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(p)),$$

2. Plus généralement, soit $I \in \mathbb{N}^*$, $(k_1, \dots, k_I) \in (\mathbb{N}^*)^I$, et pour $i = 1, \dots, I$, $p^{(i)} = (p_1^{(i)}, \dots, p_{k_i}^{(i)})$. Sous l'hypothèse $\mathbb{E}[|X_0|^{2 \max(|p_1|, \dots, |p_I|)}] < \infty$, alors

$$\sqrt{N} (\widehat{m}_N^{(p_i)} - m^{(p_i)})_{1 \leq i \leq I} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_I(0, \Sigma(p_1, \dots, p_I)), \quad (8)$$

$$\text{où } \Sigma(p_1, \dots, p_I) = \left(\sum_{\ell \in \mathbb{Z}} \left(\mathbb{E}(X_1^{p_1^{(i)}} X_2^{p_2^{(i)}} \dots X_k^{p_k^{(i)}} X_{\ell+1}^{p_1^{(j)}} \dots X_{\ell+k}^{p_k^{(j)}}) - m^{(p_i)} m^{(p_j)} \right) \right)_{1 \leq i, j \leq I}.$$

Nous pouvons donner deux conséquences directes de ce résultat pour les moments empiriques d'ordre 1 et 2 d'un processus affine causal stationnaire:

Corollaire. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus affine causal défini par (4) vérifiant l'hypothèse (7).

1. Si $\mathbb{E}[X_0^2] < \infty$, alors

$$\sqrt{N} \bar{X}_N \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \gamma^2) \quad \text{avec } \gamma^2 = 2\pi f(0) = \sum_{\ell \in \mathbb{Z}} r_X(\ell).$$

2. Si $\mathbb{E}[X_0^4] < \infty$, alors avec $\Sigma(m) = \left(\sum_{k \in \mathbb{Z}} (\mathbb{E}[X_0 X_i X_k X_{k+j}] - r_X(i) r_X(j)) \right)_{0 \leq i, j \leq m}$,

$$\sqrt{n} (\widehat{r}_N(i) - r_X(i))_{0 \leq i \leq m} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_{m+1}(0, \Sigma(m)).$$

Par conséquent, on en déduit également un théorème de la limite centrale pour les auto-corrélations empiriques, soit:

$$\sqrt{N} (\hat{\rho}_N(i) - \rho(i))_{1 \leq i \leq m} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_m(0, \Gamma(m)).$$

avec $\Gamma_{ij}(m) = \frac{1}{r_X^2(0)} \sum_{k \in \mathbb{Z}} (\mathbb{E} [X_0 X_i X_k X_{k+j}] - r_X(i) r_X(j))$. Ce qui nous permet de voir que le corrélogramme peut être utilisé pour évaluer la structure de dépendance du processus.

Remarque. *Le résultat précédent montre en revanche que les théorèmes portemanteaux ne peuvent pas être appliqués tels quels à des processus affines causaux, même lorsqu'ils sont des bruits blancs faibles. Un exemple est celui des GARCH(p,q). Dans ce cas, il est facile de voir que la matrice $\Gamma(m)$ précédente est bien diagonale, mais les termes diagonaux s'écrivent $\frac{1}{r_X^2(0)} \mathbb{E} [X_0^2 X_i^2]$, et ne sont donc pas tous égaux à 1 comme dans le cas d'un bruit blanc fort. Nous verrons qu'en travaillant de manière adéquate sur les résidus ils peuvent l'être.*

On remarque également que ces résultats ne sont pas valables pour des processus à longue mémoire: ils sont même faux, la vitesse de convergence des autocorrélations n'est plus en \sqrt{N} dans ce cas.

4.2 Estimation semi-paramétrique pour un processus affine causal stationnaire

On suppose que l'on a observé une trajectoire (X_1, \dots, X_N) d'un processus affine causal dont la dynamique est gouvernée par un paramètre $\theta = (\theta_1, \dots, \theta_d)$ où $d \in \mathbb{N}^*$ tel que la "vraie" valeur prise par ce paramètre θ^* soit inconnue. Plus précidément, on suppose que:

$$X_t = F_{\theta^*}((X_{t-k})_{k \geq 1}) + \varepsilon_t M_{\theta^*}((X_{t-k})_{k \geq 1}) \quad \text{pour tout } t \in \mathbb{Z}, \quad (9)$$

avec:

- (X_1, \dots, X_N) est une trajectoire observée;
- Pour $\theta \in \Theta \subset \mathbb{R}^d$, les fonctions $F_\theta : \mathbb{R}^\infty \rightarrow \mathbb{R}$ et $M_\theta : \mathbb{R}^\infty \rightarrow \mathbb{R}$ sont connues;
- $\theta^* \in \Theta \subset \mathbb{R}^d$ est inconnu;
- $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc dont la loi est inconnue, mais dont on sait qu'il possède un moment d'ordre r .

Notre but ici est l'estimation de θ^* , ce qui nous place dans un cadre semi-paramétrique (puisque la loi du bruit blanc n'est pas supposée être connue).

Exemples: Voici plusieurs exemples d'estimation que nous allons traiter:

- 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 $(\varepsilon_k)_k$ est tel que $\mathbb{E} [\varepsilon_0^2] = \sigma_\varepsilon^2 < \infty$, que l'on estime également. Ici $\theta = (b_1, \dots, b_p, c_1, \dots, c_q, \sigma_\varepsilon^2)$.
- Les paramètres $a_0, a_1, \dots, a_p, b_1, \dots, b_q$ d'un processus GARCH(p) stationnaire dont on supposera que le bruit blanc $(\varepsilon_k)_k$ est tel que $\mathbb{E} [\varepsilon_0^2] = 1$ (nécessité de fixer cette valeur pour éviter la surparamétrisation, a_0 jouant également le rôle d'une variance). Ici $\theta = (a_0, a_1, \dots, a_p, b_1, \dots, b_q)$.

- Les paramètres a_0, μ et p d'un processus ARCH(∞) défini par $X_t = \varepsilon_t \left(a_0 + \mu \sum_{j=1}^{\infty} j^{-p} X_{t-j}^2 \right)^{1/2}$ pour tout $t \in \mathbb{Z}$. Ici, la stationnarité et l'obtention d'un moment d'ordre 2, demande que $\mu \sum_{j=1}^{\infty} j^{-p} < 1$, donc $p > 1$, et on a $\theta = (a_0, \mu, p)$.
- Les paramètres d et σ_ε^2 d'un processus FARIMA(0, d , 0) avec $d \in]-1/2, 1/2[$. Pour cet exemple on a $\theta = (\sigma_\varepsilon^2, d)$.

Dans la suite on propose une méthode dont le principe est unique pour estimer θ^* dans des cadres aussi divers que ceux des exemples évoqués ci-dessous, et plus généralement pour des processus affines causaux stationnaires: celle de l'estimateur par maximum de quasi-vraisemblance gaussienne.

Pour cela on suppose dans un premier temps afin de construire l'estimateur 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 $\widehat{f}_\theta^t = f_\theta(X_{t-1}, \dots, X_1, 0, \dots)$ et $\widehat{M}_\theta^t = M_\theta(X_{t-1}, \dots, X_1, 0, \dots)$, alors on définit la quasi log-densité conditionnelle par

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

Pour les densités conditionnelles, on connaît la propriété suivante:

$$f_{((X_1, \dots, X_N) | X_0, X_{-1}, \dots)} = f_{(X_N | X_{N-1}, X_{N-2}, \dots)} f_{(X_{N-1} | X_{N-2}, X_{N-3}, \dots)} \times \dots \times f_{(X_1 | X_0, X_{-1}, \dots)}.$$

En passant au logarithme, le produit devient une somme. Ceci nous permet de définir \widehat{L}_n la quasi-log-vraisemblance conditionnelle gaussienne et l'estimateur par maximum de quasi-vraisemblance (QMLE) gaussienne par:

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\text{Argmax}} \widehat{L}_n(\theta) \quad \text{avec} \quad \widehat{L}_n(\theta) = \sum_{t=1}^n \widehat{q}_t(\theta). \quad (10)$$

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

Théorème 6. *On suppose que $r \geq 2$ et on définit*

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^d, \sum_{i=1}^{\infty} \alpha_i(F_\theta) + (\mathbb{E}[|\varepsilon_0|^r])^{1/r} \sum_{i=1}^{\infty} \alpha_i(M)_\theta < 1 \right\}. \quad (11)$$

On suppose que Θ est un compact de $\Theta(r)$, que $\theta^ \in \Theta$ et:*

1. *Il existe $\underline{M} > 0$, tel que pour tout $\theta \in \Theta(r)$ et tout $x \in \mathbb{R}^\infty$, $M_\theta(x) \geq \underline{M} > 0$.*
2. *Si pour θ_1 et θ_2 dans Θ , $(M_{\theta_1}(\cdot) = M_{\theta_2}(\cdot))$ ou $(F_{\theta_1}(\cdot) = F_{\theta_2}(\cdot))$ alors $(\theta_1 = \theta_2)$.*
3. *Il existe $\ell > \max(1, 3/r)$ tel que $\sup_{\theta \in \Theta} \alpha_j(F_\theta) + \sup_{\theta \in \Theta} \alpha_j(M_\theta) = O(j^{-\ell})$.*

Alors on a: $\widehat{\theta}_N \xrightarrow[N \rightarrow \infty]{p.s.} \theta^*$.

Exemples: Cette convergence presque-sûre de l'estimateur est valable:

- Pour estimer tous les paramètres d'un processus ARMA ou d'un processus GARCH.
- Pour estimer tous les paramètres de modèles type ARMA-GARCH, TARCH ou APARCH.
- Pour estimer le vecteur de paramètre θ d'un modèle AR(∞) ou ARCH(∞) dont les coefficients (qui sont infinis) s'écrivent $a_i(\theta)$ et vérifient les conditions de décroissance précédentes.
- Pour $r \geq 3$, ce théorème est donc également valable pour les processus affine causal à longue-mémoire de type $F_\theta(x) = \sum_{j \geq 1} a_j(\theta) x_j$ avec $a_j(\theta) = L_\theta(j) j^{-1-d}$ et $\alpha_j(M_\theta) = O(j^{-\ell})$, où $d \in]0, 1/2[$ (typiquement un processus FARIMA(p, d, q)).

Sous quelques hypothèses supplémentaires, on peut également montrer la normalité asymptotique de $\widehat{\theta}_N$:

Théorème 7. *On se place sous les conditions du Théorème 6 avec $r \geq 4$. On rajoute également les hypothèses suivantes:*

1. *On suppose que pour tout $\theta \in \Theta$, F_θ et G_θ sont 3 fois différentiables sur \mathbb{R}^∞ ;*
2. *Il existe $\ell' > 3/2$ tel que*

$$\sup_{\theta \in \Theta} \sup_{x \in \mathbb{R}^\infty} \left(\left\| \frac{\partial^2}{\partial \theta \partial x_j} F_\theta(x) \right\| + \left\| \frac{\partial^2}{\partial \theta \partial x_j} M_\theta(x) \right\| + \left\| \frac{\partial^3}{\partial \theta^2 \partial x_j} F_\theta(x) \right\| + \left\| \frac{\partial^3}{\partial \theta^2 \partial x_j} M_\theta(x) \right\| \right) = O(j^{-\ell'}).$$

Alors

$$\sqrt{N} (\widehat{\theta}_N - \theta^*) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_d(0, F(\theta^*)^{-1} G(\theta^*) F(\theta^*)^{-1}). \quad (12)$$

avec $(F(\theta^*))_{i,j} = \mathbb{E} \left[\frac{\partial^2 q_0(\theta^*)}{\partial \theta_i \partial \theta_j} \right]$ et $(G(\theta^*))_{i,j} = \mathbb{E} \left[\frac{\partial q_0(\theta^*)}{\partial \theta_i} \frac{\partial q_0(\theta^*)}{\partial \theta_j} \right]$ pour $i, j \in \{1, \dots, d\}$.

Il est clair qu'en utilisant le Lemme de Slutsky, on peut remplacer $F(\theta^*)$ et $G(\theta^*)$ par $F(\widehat{\theta}_N)$ et $G(\widehat{\theta}_N)$ dans une renormalisation du TLC précédent, ce qui peut permettre d'obtenir des régions de confiance et des tests de Wald sur les paramètres du processus.

Exemples: Cette normalité asymptotique est valable pour les mêmes exemples que précédemment, à savoir, sans aucune condition supplémentaire si ce n'est $r \geq 4$, pour estimer tous les paramètres d'un processus ARMA, GARCH, ARMA-GARCH, TARCH ou APARCH. Et avec des conditions de différentiabilité en θ , pour estimer les paramètres de modèles AR(∞) ou ARCH(∞). En revanche, pour des processus de type FARIMA, même si ce résultat est valable, il nécessite d'être précisé pour passer de $\ell > 3/2$ à $\ell > 1$ dans ce cas précis.

4.3 Sélection de modèle et test d'adéquation pour les processus affines causaux

Nous venons de voir qu'une fois un modèle posé, on peut estimer ses paramètres avec une méthode performante. Les tests de Wald, comme les tests de Student ou de Fisher en régression linéaire par moindres carrés, pourraient être mis en place pour tester la nullité des coefficients des modèles. Mais, comme désormais on préfère le faire en régression, on va utiliser des critères de sélection de modèles AIC et BIC pour identifier un modèle.

On commence d'abord par la définition suivante:

Definition. Pour $X = (X_t)_t$ un modèle causal affine stationnaire semi-paramétrique défini par (9), dont on a observé une trajectoire (X_1, \dots, X_N) et pour lequel le paramètre $\theta \in \Theta \subset \mathbb{R}^d$ est inconnu. On définit alors:

- le critère AIC pour ce modèle: $\widehat{AIC} = -2 \widehat{L}_N(\widehat{\theta}_N) + 2d$;
- le critère BIC pour ce modèle: $\widehat{BIC} = -2 \widehat{L}_N(\widehat{\theta}_N) + \log(N)d$.

Notons que d pourrait aussi être remplacé par "nombre de paramètres estimés". Ces critères sont les acronymes de Akaike Information Criterion (Akaike, 1973) et Bayesian Information Criterion (Shwartz, 1978). La valeur de ces deux critères importe peu, ce sont plutôt les comparaisons entre ces valeurs pour différents modèles qui vont être essentielles. Ainsi va-t-on plutôt définir ces critères pour des familles de modèles:

Definition. Soit \mathcal{M} une famille de modèles affines causaux stationnaires semi-paramétriques, telle que si $m \in \mathcal{M}$ il existe $\theta(m)$ tel que m soit défini par (9). On observe une trajectoire (X_1, \dots, X_N) d'un processus stationnaire. Si l'on note $|m|$ la dimension de l'espace dans lequel $\theta(m)$ varie, on définit alors:

- le critère AIC pour un modèle $m \in \mathcal{M}$ par: $\widehat{AIC}(m) = -2 \widehat{L}_N(\widehat{\theta}_N(m)) + 2|m|$;
- le critère BIC pour un modèle $m \in \mathcal{M}$ par: $\widehat{BIC}(m) = -2 \widehat{L}_N(\widehat{\theta}_N(m)) + \log(N)|m|$.

La famille de modèle peut-être hiérarchique (par exemple si elle contient tous les modèles AR(p) avec $p \in \{0, 1, \dots, p_{\max}\}$) ou non (par exemple si elle contient tous les modèles ARMA(p, q) avec $p \in \{0, 1, \dots, p_{\max}\}$ et $q \in \{0, 1, \dots, q_{\max}\}$). Elle peut également être composée de différentes familles de modèles (par exemple l'ensemble des ARMA(p, q) avec $p \in \{0, 1, \dots, p_{\max}\}$ et $q \in \{0, 1, \dots, q_{\max}\}$ et des GARCH(p', q') avec $p' \in \{0, 1, \dots, p'_{\max}\}$ et $q' \in \{0, 1, \dots, q'_{\max}\}$).

Le critère mesure une forme de distance (dissemblance de Kullback) entre le modèle m et un modèle optimal, le second est relatif à l'inverse d'une probabilité d'avoir le modèle m sachant que (X_1, \dots, X_N) a été observé. Parmi tous les modèles de la famille \mathcal{M} , on choisira le(s) modèles qui minimisent ces critères:

$$\widehat{m}_{AIC} = \underset{m \in \mathcal{M}}{\text{Argmin}} \widehat{AIC}(m) \quad \text{et} \quad \widehat{m}_{BIC} = \underset{m \in \mathcal{M}}{\text{Argmin}} \widehat{BIC}(m).$$

On peut alors montrer le résultat suivant:

Théorème 8. *Si \mathcal{M} est une famille de taille finie de modèles affines causaux stationnaires semi-paramétriques, et s'il existe un "vrai" modèle $m^* \in \mathcal{M}$ tel que (X_1, \dots, X_N) soit une trajectoire observée issue de la dynamique (9) où $\theta = \theta(m^*)$. On suppose les conditions du Théorème 7 pour tous les modèles $m \in \mathcal{M}$ et qu'il existe $\ell'' > 5/2$ tel que*

$$\sup_{\theta(m) \in \Theta} \sup_{x \in \mathbb{R}^\infty} \left(\left\| \frac{\partial}{\partial x_j} F_{\theta(m)}(x) \right\| + \left\| \frac{\partial}{\partial x_j} M_{\theta(m)}(x) \right\| \left\| \frac{\partial^2}{\partial \theta \partial x_j} F_{\theta(m)}(x) \right\| + \left\| \frac{\partial^2}{\partial \theta \partial x_j} M_{\theta(m)}(x) \right\| + \left\| \frac{\partial^3}{\partial \theta^2 \partial x_j} F_{\theta(m)}(x) \right\| + \left\| \frac{\partial^3}{\partial \theta^2 \partial x_j} M_{\theta(m)}(x) \right\| \right) = O(j^{-\ell''}).$$

Alors:

$$\mathbb{P}(\widehat{m}_{BIC} = m^*) \xrightarrow{N \rightarrow \infty} 1 \quad \text{et} \\ \sqrt{N} (\widehat{\theta}_N(\widehat{m}_{BIC}) - \theta(m^*)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_{|m^*|}(0, F(\theta(m^*))^{-1} G(\theta(m^*)) F(\theta(m^*))^{-1}). \quad (13)$$

Ainsi, sous des hypothèses pas trop lourdes, le critère BIC est consistant, c'est-à-dire qu'il est capable de retrouver asymptotiquement le vrai modèle lorsque celui-ci appartient à la famille de modèles. Par exemple, ceci peut être appliqué pour retrouver le vrai ordre (p^*, q^*) d'un processus ARMA lorsque celui-ci est inconnu et que les données sont bien issues d'un processus ARMA. Le critère BIC permettra également de distinguer asymptotiquement entre un ARMA et un GARCH pour modéliser, avec l'obtention de l'ordre. De plus, la seconde propriété montre que l'on peut directement utiliser le modèle choisi par le BIC pour obtenir un TLC optimal sur le paramètre, et l'on ne perd rien par rapport à connaître le vrai modèle à l'avance. On notera cependant que la consistance n'est pas montrée pour les processus à longue mémoire.

Concernant le critère AIC, le résultat de consistance que l'on peut obtenir sous les mêmes hypothèses est le suivant:

$$\mathbb{P}(\widehat{m}_{AIC} \supset m^*) \xrightarrow{N \rightarrow \infty} 1,$$

ce qui signifie que le critère AIC a asymptotiquement tendance à sur-paramétriser. Faut-il malgré tout conserver le critère AIC pour sélectionner un modèle de processus? Oui, pour deux raisons:

1. Le critère AIC a des propriétés d'efficience plutôt que de consistance, c'est-à-dire qu'il aura tendance à choisir un modèle qui minimisera le risque (c'est-à-dire les erreurs de prédiction: voir les propriétés oracles;
2. Surtout, et même si la famille de modèles considérée est grande, il n'y a souvent aucune raison pour qu'un vrai modèle existe (on rentre donc dans une problématique d'apprentissage statistique plutôt que de statistique inférentielle). Dans un tel cadre, le critère AIC conserve des propriétés oracle, c'est-à-dire qu'il va sélectionner le modèle optimal en terme d'erreur de minimisation du risque dans la famille de modèles.

Vient alors à nouveau une question de statistique inférentielle: une fois le modèle choisi, ses paramètres estimés, comment être sûr que le modèle est acceptable, en vue par exemple de l'utiliser pour des prédictions? C'est ce que permet de faire le test d'adéquation suivant, de type portemanteau.

Le problème de test considéré est le suivant:

$$\begin{cases} H_0 : \exists m^* \in \mathcal{M}, \text{ tel que } (X_1, \dots, X_n) \text{ est une trajectoire de } X \text{ satisfaisant (9) avec } \theta^* = \theta(m^*) \\ H_1 : \nexists m^* \in \mathcal{M}, \text{ tel que } (X_1, \dots, X_n) \text{ est une trajectoire de } X \text{ satisfaisant (9) avec } \theta^* = \theta(m^*) \end{cases}$$

Pour ce faire, on commence par définir les résidus pour un modèle $m \in \mathcal{M}$ quelconque:

$$\widehat{e}_t(m) := (\widehat{M}_{\widehat{\theta}(m)}^t)^{-1} (X_t - \widehat{f}_{\widehat{\theta}(m)}^t),$$

calculés à partir de la trajectoire observée. Pour $K \in \mathbb{N}^*$ fixé, on définit également le vecteur d'autocorrélation des carrés des résidus, soit:

$$\widehat{\rho}(m) := (\widehat{\rho}_1(m), \dots, \widehat{\rho}_K(m))',$$

$$\text{avec pour } k = 1, \dots, K, \widehat{\rho}_k(m) := \frac{\widehat{\gamma}_k(m)}{\widehat{\gamma}_0(m)} \text{ et } \widehat{\gamma}_k(m) := \frac{1}{n} \sum_{t=k+1}^n (\widehat{e}_t^2(m) - 1)(\widehat{e}_{t-k}^2(m) - 1).$$

Théorème 9. *Sous les hypothèses du Théorème 8, en notant $\mu_4 = \mathbb{E}[\varepsilon^4]$ et si $\mathbb{E}[\varepsilon_0^3] = 0$ (vrai notamment si la distribution du bruit est symétrique), alors*

$$\begin{aligned} 1. \text{ Avec } V(\theta^*) &= I_K + \frac{J_K(m^*) F(\theta^*)^{-1}}{(\mu_4 - 1)^2} \left(G(\theta^*) F(\theta^*)^{-1} + (\mu_4 - 1) I_K \right) J_K'(m^*), \text{ où l'on pose} \\ J_K(m^*) &= -2 \left(\mathbb{E} \left[(\varepsilon_0^2 - 1) \frac{\partial}{\partial \theta_j} \log(M_{\theta^*}^i) \right] \right)_{1 \leq i \leq K, j \in m^*}, \text{ on a} \\ \sqrt{N} \widehat{\rho}(m^*) &\xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_K(0, V(\theta^*)). \end{aligned} \quad (14)$$

$$2. \text{ Si } \widehat{Q}_K(m^*) := N \widehat{\rho}(m^*)' (V(\widehat{\theta}(m^*)))^{-1} \widehat{\rho}(m^*), \text{ alors}$$

$$\widehat{Q}_K(m^*) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi^2(K). \quad (15)$$

$$3. \text{ Les points 1. et 2. sont également vrai si } m^* \text{ est remplacé par } \widehat{m}_{BIC}.$$

Il est facile de voir que pour un processus ARMA, la matrice J_K est nulle et ainsi $V(\theta^*) = I_K$: le test revient à un test classique de type portemanteau de Box-Pierce, mais sur les carrés des résidus. Dans les autres cas classiques, cette matrice J_K peut être naturellement estimée

par $\widehat{J}_K = \frac{1}{N} \sum_{t=1}^{N-i} (\widehat{e}_t^2(\widehat{m}_{BIC}) - 1) \left(\frac{\partial \log(M_{\widehat{\theta}(\widehat{m}_{BIC})}^{t+i})}{\partial \theta_j} \right)_{i,j}$ et μ^4 par $\widehat{\mu}_4 = \frac{1}{N} \sum_{t=1}^N \widehat{e}_t^4(\widehat{m}_{BIC})$, le

Lemme de Slutsky permettant une nouvelle fois d'utiliser la statistique de test $\widehat{Q}_K(\widehat{m}_{BIC})$ en remplaçant $V(\widehat{\theta}(\widehat{m}_{BIC}))$ par son estimateur $\widehat{V}(\widehat{\theta}(\widehat{m}_{BIC}))$ ainsi obtenue.

5 Prédiction pour un processus affine causal stationnaire

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

5.1 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 $\mathbb{E}(Y | (X_1, \dots, X_n)) = a_0 + a_1X_1 + \dots + a_nX_n$, 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:

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

Alors,

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

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

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

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 $\widehat{X}_{N+1} = 0$.

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:

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

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).

5.2 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:

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

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 :

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

en notant $\Omega(\beta) = \begin{pmatrix} \beta^N & 0 & 0 & \dots & 0 \\ 0 & \beta^{N-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$ et avec Z 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