

Master 2 Ingénierie Statistique

SÉRIES TEMPORELLES PAUL ROCHET

Contents

1	Inti	Introduction		
	1.1	Processus stochastiques	2	
	1.2	Auto-covariance et auto-corrélation	9	
	1.3	Innovation	Ę	
	1.4	Auto-corrélation partielle	7	
2	Tendance et saisonnalité			
	2.1	Modélisation paramétrique	10	
	2.2	Filtrage	12	
3	Pro	ocessus linéaires standards	17	
	3.1	Moyennes mobiles	19	
	3.2	Processus auto-régressifs	21	
	3.3	Processus ARMA	23	
4	Extensions des processus ARMA			
	4.1	Processus ARMA saisonniers	26	
	4.2	ARIMA et SARIMA	27	
	4.3	Modélisation avec variables exogènes	27	
	4.4	Modèles GARCH	28	
5	Infé	érence	29	
	5.1	Estimation des paramètres	29	
	5.2	Tests de stationnarité	31	
	5.3	Tests de blancheur		
	5.4	Ajustement d'un modèle SARIMA	33	
6	Prévision			
	6.1	Meilleur prédicteur linéaire	34	
	6.2	Lissage exponentiel	36	

1 Introduction

1.1 Processus stochastiques

Un processus stochastique est une famille de variables aléatoires $(X_t)_{t\in T}$ indexée par un espace T et définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans un même espace E. Dans ce cours, on se placera dans l'espace $\mathbb{L}^2 = \mathbb{L}^2(\Omega)$ des variables aléatoires de carré intégrable à valeurs dans un espace Euclidien E, voire même le plus souvent $E = \mathbb{R}$.

Si le processus est indexé par le temps, on parle de série temporelle. On peut avoir un processus à temps discret $(T = \mathbb{N} \text{ ou } \mathbb{Z})$ ou à temps continu $(T = \mathbb{R}_+ \text{ ou } \mathbb{R})$. Il existe des processus stochastiques indexés par des espaces de dimension supérieure comme en imagerie avec $T = \mathbb{Z}^2$ par exemple.

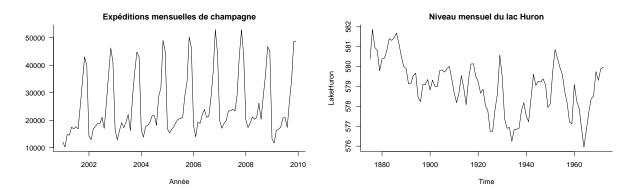


Figure 1: Exemples de séries temporelles mensuelles relevées sur plusieurs années.

On s'intéressera quasi exclusivement à des séries temporelles à temps discret avec $T = \mathbb{N}$ ou \mathbb{Z} . En pratique, on dispose d'un échantillon $x_1, ..., x_n$ que l'on suppose issu d'une réalisation du processus X_t observé à n temps successifs.

L'objectif de ce cours est de modéliser l'évolution d'un processus X_t dont on observe une réalisation $X = (X_1, ..., X_n)$ dans le temps, généralement à des fins de prédiction.

La loi d'un processus $(X_t)_{t \in T}$ est caractérisée par ses marginales finies dimensionnelles, c'est-à-dire par la loi des vecteurs $(X_{t_1}, ..., X_{t_k})$ pour tout $t_1 < ... < t_k \in T$, $k \in \mathbb{N}$. La loi d'un processus est donc potentiellement très compliquée. En pratique, on caractérisera souvent la loi d'un processus en se ramenant à des processus dits *stationnaires*.

Définition 1.1. Une série temporelle $(X_t)_{t\in\mathbb{Z}}$ est stationnaire (ou faiblement stationnaire ou stationnaire au second ordre) si elle est d'espérance constante au cours du temps et si pour tout $h \in \mathbb{N}, t \in \mathbb{Z}$, $cov(X_t, X_{t+h}) = cov(X_0, X_h) \in \mathbb{R}$ (la covariance entre X_t et X_{t+h} ne dépend pas de t).

En particulier, la stationnarité implique l'existence d'un moment d'ordre 2 qui est constant dans le temps. Grossièrement, une série temporelle est stationnaire si ses moments joints jusqu'à l'ordre 2 sont invariants

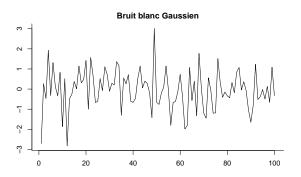
par translation dans le temps. On dira qu'une série temporelle est fortement stationnaire si sa loi est invariante par translation:

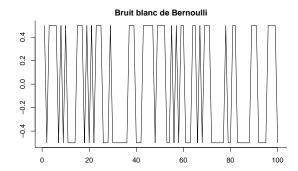
$$(X_{t_1},...,X_{t_k}) \stackrel{loi}{=} (X_{t_1+h},...,X_{t_k+h}), \forall h \in \mathbb{N}, \forall t_1,...,t_k.$$

Il existe des tests sur la stationnarité, les plus utilisés étant le test de Dickey-Fuller augmenté (ADF) dont l'hypothèse nulle implique la non-stationnarité du processus (on conclut donc la stationnarité en rejettant H_0) et le test de Kwiatkowski, Phillips, Schmidt et Shin (KPSS) dont l'hypothèse nulle sous-entend la stationnarité. Sous R, ces tests sont codés par les fonctions adf.test et kpss.test de la librairie tseries.

Définition 1.2. Un bruit blanc est un processus stochastique $(\epsilon_t)_{t\in T}$ indexé par un espace discret tel que ϵ_t est centré de variance $\sigma^2 < \infty$ pour tout t et pour tout s, t différents, $\operatorname{cov}(\epsilon_t, \epsilon_s) = 0$.

On dira qu'un bruit blanc $(\epsilon_t)_{t\in T}$ est fort si les ϵ_t sont iid. Il n'existe pas de bruit blanc indexé par un espace continu. Un bruit blanc Gaussien (dont toutes les marginales finies dimensionnelles sont Gaussiennes) est un bruit blanc fort.





Exercice. Montrer qu'un bruit blanc (resp. bruit blanc fort) est stationnaire (resp. fortement stationnaire).

Les bruits blancs ne sont bien sûr pas les seuls processus stationnaires mais la plupart des processus stationnaires classiques sont définis à partir de transformations linéaires de bruits blancs. Par exemple, une moyenne mobile (MA pour mobile average) d'ordre 1 est un processus stationnaire de la forme

$$X_t = \epsilon_t + a\epsilon_{t-1} , t \in \mathbb{Z}$$

où $a \in \mathbb{R}$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc. On vérifie facilement que ce processus est stationnaire (exercice).

1.2 Auto-covariance et auto-corrélation

Pour une série temporelle $(X_t)_{t\in\mathbb{Z}}$ stationnaire, on peut s'intéresser à sa fonction d'auto-covariance, définie par

$$\gamma(h) := \operatorname{cov}(X_t, X_{t+h}), h \in \mathbb{N}.$$

C'est bien une fonction du lag h uniquement (elle ne dépend pas de t) précisément parce que $(X_t)_{t\in\mathbb{Z}}$ est stationnaire. Soit $\sigma^2 = \gamma(0)$ la variance de X_t , on définit de la même manière la fonction d'auto-corrélation ou ACF (auto-corrélation function) par

$$\rho(h) := \operatorname{cor}(X_t, X_{t+h}) = \frac{\operatorname{cov}(X_t, X_{t+h})}{\sigma^2} , h \in \mathbb{N}.$$

On peut étendre ces fonctions à $h \in \mathbb{Z}$ par symétrie en posant $\gamma(h) = \gamma(-h)$.

Proposition 1.3. Une fonction $\gamma: \mathbb{Z} \to \mathbb{R}$ est la fonction d'auto-covariance d'une série temporelle stationnaire si et seulement elle est symétrique et de type positif, c'est-à-dire si pour tout $N \in \mathbb{N}$ et pour toute suite réelle $a = (a_k)_{k \in \mathbb{Z}}$ à support fini

$$\sum_{s,t\in\mathbb{Z}} a_s a_t \gamma(s-t) \ge 0.$$

Preuve. Soit γ la fonction d'auto-covariance d'un processus $(X_t)_{t\in\mathbb{Z}}$ stationnaire (qu'on suppose centré sans perte de généralité). Pour tout $a=(a_k)_{k\in\mathbb{Z}}$ à support fini,

$$0 \le \operatorname{var}\left(\sum_{t \in \mathbb{Z}} a_t X_t\right) = \sum_{s,t \in \mathbb{Z}} a_s a_t \gamma(s-t).$$

Réciproquement, soit $\gamma: \mathbb{Z} \to \mathbb{R}$ symétrique de type positif. Pour $N \in \mathbb{N}$, on considère la matrice de Toeplitz $\Sigma_N = (\gamma(s-t))_{s,t=-N,\dots,N}$ qui est de taille $(2N+1) \times (2N+1)$. Cette matrice est semi-définie positive par hypothèse, il existe donc un vecteur Gaussien $Y^{(N)} = (Y_t^{(N)})_{t=-N,\dots,N}$ centré de matrice de covariance Σ_N . On définit le processus $X^{(N)}$ indexé par \mathbb{Z} :

$$X_t^{(N)} = \left\{ \begin{array}{ll} Y_t^{(N)} & \text{si } -N \leq t \leq N, \\ 0 & \text{sinon.} \end{array} \right.$$

On peut alors montrer que $X^{(N)}$ converge en loi vers un processus stationnaire de fonction d'auto-covariance γ (on vérifie facilement la convergence des marginales finies dimensionnelles mais la convergence en loi de processus nécessite un critère supplémentaire de tension que l'on admettra).

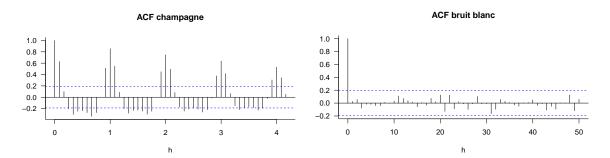
Les fonctions d'auto-covariance et auto-corrélation sont bien sûr inconnues en pratique mais elle peuvent être estimées empiriquement. Soient $X_1, ..., X_n$ les données issues de $(X_t)_{t \in \mathbb{Z}}$, on note $\widehat{m} = \frac{1}{n} \sum_{i=1}^n X_i$. Les fonctions d'auto-covariance et auto-corrélation empirique sont alors définies par

$$\widehat{\gamma}(h) := \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \widehat{m})(X_{i+h} - \widehat{m}) \text{ et } \widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}, h \in 0, ..., n-1.$$

et $\widehat{\gamma}(h) = \widehat{\rho}(h) = 0$ pour $h \ge n$. On peut bien entendu là aussi étendre ces fonctions à $h \in \mathbb{Z}$ en posant $\widehat{\gamma}(-h) = \widehat{\gamma}(h)$. On remarque que l'auto-covariance empirique $\widehat{\gamma}(h)$ est calculée avec la renormalisation en 1/n au lieu de l'habituelle 1/(n-h). La raison principale est d'obtenir une fonction de type positif, ce qui ne serait pas nécessairement le cas avec des renormalisation différentes pour chaque h.

Exercice. Montrer que la fonction d'auto-covariance empirique $\hat{\gamma}$ est de type positif.

L'ACF (et son estimation parfois appelée auto-corrélogramme) nous donne de l'information sur la dépendance du processus. Il est toujours utile de le représenter grahiquement car il met en évidence ces dépendances. Par exemple, un bruit blanc est caractérisé par un acf $\rho(h)$ nul pour $h \ge 1$. Par abus de notation, on parle parfois d'ACF pour désigner l'estimation.



1.3 Innovation

Pour un processus $(X_t)_{t\in\mathbb{Z}}$, l'innovation désigne la part du processus qui est imprévisible (linéairement) à partir des observations passées et donc qui est non corrélée au passé.

L'enveloppe linéaire H d'un processus $(X_t)_{t\in\mathbb{Z}}$ de carré intégrable est l'adhérence du sous-espace vectoriel de \mathbb{L}^2 engendré par les variables ..., X_t, X_{t+1}, \dots (en incluant la constante) :

$$H = \overline{\text{vect}\{1, X_t, t \in \mathbb{Z}\}}.$$

L'enveloppe linéaire d'un processus (pas nécessairement stationnaire) contient donc l'ensemble des variables aléatoires de \mathbb{L}^2 que l'on peut obtenir comme transformation linéaire du processus X_t , c'est-à-dire qui s'écrivent sous la forme

$$Y := \sum_{t \in \mathbb{Z}} a_t X_t \ , \ t \in \mathbb{Z},$$

et tel que $\text{var}(Y) < \infty$. De la même manière, on définit le passé (linéaire) de X_t comme l'adhérence de l'espace vectoriel engendré par la série jusqu'au temps t

$$H_t = \overline{\operatorname{vect}\{1, X_s, s \le t\}}, \ t \in \mathbb{Z}.$$

Clairement, les ensembles H_t sont croissants. On définit le passé lointain comme la limite en moins l'infini:

$$H_{-\infty} = \lim_{t \to -\infty} H_t = \bigcap_{t \in \mathbb{Z}} H_t.$$

Dans les deux cas extrêmes, un processus est dit purement déterminisite si $H_t = H_{t-1}$ pour tout $t \in \mathbb{Z}$ (et donc si $H_{-\infty} = H$) et purement innovant si $H_{-\infty} \subseteq \text{vect}\{1\}$. Par exemple, un bruit blanc est purement innovant tandis qu'un processus constant est purement déterministe.

Définition 1.4. Soit $(X_t)_{t\in\mathbb{Z}}$ un processus de carré intégrable, son processus des innovations est donné par

$$\epsilon_t = X_t - \Pi_{H_{t-1}}(X_t) , t \in \mathbb{Z}$$

où $\Pi_{H_{t-1}}$ est l'opérateur de projection orthogonale sur H_{t-1} dans \mathbb{L}^2 .

L'innovation d'un processus $(X_t)_{t\in\mathbb{Z}}$ est donc la partie imprévisible (linéairement) à partir des données passées. On remarque que H_t est la somme directe

$$H_t = H_{t-1} \oplus \operatorname{vect}\{\epsilon_t\}$$

et que les espaces H_{t-1} et vect $\{\epsilon_t\}$ sont orthogonaux. Si X_t est de variance finie, la variance de l'innovation est strictement inférieure à celle de X_t par la relation de Pythagore

$$\operatorname{var}(\epsilon_t) = \operatorname{var}(X_t) - \operatorname{var}(\Pi_{H_{t-1}}(X_t)) \le \operatorname{var}(X_t).$$

Un processus est purement déterminisite si et seulement si son innovation est presque sûrement nulle pour tout t.

Proposition 1.5. Si $(X_t)_{t\in\mathbb{Z}}$ est un processus stationnaire, alors son processus des innovations est un bruit blanc.

Preuve. On suppose le processus $(X_t)_{t\in\mathbb{Z}}$ centré sans perte de généralité. Son processus des innovations $(\epsilon_t)_{t\in\mathbb{Z}}$ est entièrement caractérisé par son auto-covariance γ . En effet, par définition de H_{t-1} , on peut écrire la projection orthogonale comme $\Pi_{H_{t-1}}(X_t) = \lim_{N\to\infty} \sum_{h=1}^N a_h^N X_{t-h}$ où $(a_1^{(N)}, ..., a_N^{(N)})$ minimise

$$(a_1,...,a_N) \mapsto \operatorname{var}\left(X_t - \sum_{h=1}^N a_h X_{t-h}\right) = \gamma(0) - 2\sum_{h=1}^N a_h \gamma(h) - 2\sum_{h,h'=1}^N a_h a_{h'} \gamma(h-h').$$

Le processus des innovations est donc stationnaire. Par construction, $\epsilon_t \in H_t \cap H_{t-1}^{\perp}$, ϵ_t est donc orthogonal à H_{t-1} et à H_{t-h} pour $h \geq 2$ par inclusion. D'où $\operatorname{cov}(\epsilon_t, \epsilon_{t-h}) = 0$ pour tout $h \geq 1$.

Exercice. Montrer que la réciproque de la proposition précédente n'est pas vraie.

Théorème 1.6. Soit $(X_t)_{t\in\mathbb{Z}}$ une série temporelle stationnaire purement innovante d'espérance m, il existe une suite réelle $(\psi_k)_{k\geq 1}$ de carré sommable tels que

$$X_t = m + \sum_{k>1} \psi_k \epsilon_{t-k} + \epsilon_t \ , \ t \in \mathbb{Z},$$

avec $(\epsilon_t)_{t\in\mathbb{Z}}$ le bruit blanc des innovations de $(X_t)_{t\in\mathbb{Z}}$.

Preuve. Soit ϵ_t l'innovation au temps t, on sait que $H_t = H_{t-1} \oplus \text{vect}\{\epsilon_t\}$. Par récurrence, on obtient pour tout h > 1.

$$H_t = H_{t-h} \oplus E_t^h$$

où $E_t^h = \text{vect}\{\epsilon_t, ..., \epsilon_{t-h}\}$ est orthogonal à H_{t-h} . De plus, comme les ensembles $E_t^h, h \ge 1$ sont croissants, on a

$$X_{t} = \Pi_{H_{t-h}}(X_{t}) + \sum_{k=1}^{h} \psi_{k} \epsilon_{t-k} + \epsilon_{t} , t \in \mathbb{Z},$$

où les coefficients $\psi_k, k \leq h$ ne dépendent pas de h (car les ϵ_t sont orthogonaux). On sait par hypothèse que la variance de X_t est finie avec

$$\forall h \ge 1 , \infty > \operatorname{var}(X_t) \ge \sigma^2 \left(1 + \sum_{k=1}^h \psi_k^2\right).$$

On déduit que $\sum_{k\geq 1}\psi_k^2<\infty$. En particulier, $\sum_{k=1}^h\psi_k\epsilon_{t-k}$ converge dans \mathbb{L}^2 quand $h\to\infty$ (on peut montrer par exemple que c'est une suite de Cauchy), la limite $\lim_{h\to\infty}\Pi_{H_{t-h}}(X_t)$ existe donc aussi dans \mathbb{L}^2 . Par construction, la limite appartient à $H_{-\infty}=\mathrm{vect}\{1\}$ car le processus est purement innovant. \square

La réciproque est vraie, ce qui se montre facilement. Cette représentation est la décomposition de Wold d'un processus stationnaire purement innovant.

Exercice. Calculer la fonction d'auto-covariance d'un processus centré stationnaire purement innovant en fonction des coefficients ψ_k de sa décomposition de Wold.

Exercice. Déterminer la décomposition de Wold du processus $X_t = \epsilon_t + \epsilon_{t-1}$.

1.4 Auto-corrélation partielle

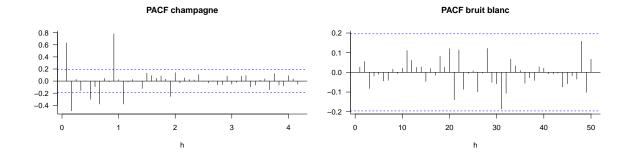
Un outil similaire indispensable pour l'analyse de séries temporelles est la fonction d'auto-corrélation partielle ou PACF. Pour $t \in \mathbb{Z}$, soit

$$H_t^0 = \text{vect}\{1\} \text{ et } H_t^h = \text{vect}\{1, X_t, ..., X_{t-h+1}\} , h = 1, 2, ...$$

Autrement dit, H_t^h est le sous-espace vectoriel de \mathbb{L}^2 de dimension au plus h engendré par le passé $X_t, ..., X_{t-h+1}$. On note $\Pi_{H_t^h}$ l'opérateur de projection orthogonale sur cet espace. La fonction d'auto-corrélation partielle ou PACF du processus $(X_t)_{t\in\mathbb{Z}}$ est donnée par

$$\alpha(h) := \operatorname{cor}\left(X_t - \Pi_{H_{t-1}^{h-1}}(X_t), X_{t-h} - \Pi_{H_{t-1}^{h-1}}(X_{t-h})\right), \ h \ge 1, \ t \in \mathbb{Z}.$$

Pour h=1, la projection $\Pi_{H_{t-1}^0}(.)$ est identiquement nulle, on a donc $\alpha(1)=\rho(1)$. De façon (peut être) plus parlante, la corrélation partielle à l'ordre h et la corrélation résiduelle entre X_t et X_{t-h} lorsqu'on a éliminé les dépendances linéaires avec la série aux h-1 temps intermédiaires t-1,...,t-h+1.



Pour tout lag $h \ge 1$, le coefficient $\alpha(h)$ peut s'obtenir par régression linéaire de X_t en fonction du passé jusqu'à l'ordre h: $X_{t-1},...,X_{t-h}$. Quitte à considérer la série recentrée $(X_t-m)_{t\in\mathbb{Z}}$ où $m=\mathbb{E}(X_t)$, on suppose sans perte de généralité que $\mathbb{E}(X_t)=0$. Soit $\beta_1^{(h)},...,\beta_h^{(h)}$ les coefficients de cette régression linéaire, solution de

$$X_{t} = \beta_{1}^{(h)} X_{t-1} + \dots + \beta_{h}^{(h)} X_{t-h} + \eta_{t}^{(h)}, \tag{1}$$

où $\eta_t^{(h)}$ est orthogonal (non corrélé) à $X_{t-1},...,X_{t-h}$. Ces coefficients ne dépendent pas du temps par stationnarité de $(X_t)_{t\in\mathbb{Z}}$. On remarque par ailleurs que

$$\beta_1^{(h)} X_{t-1} + \dots + \beta_h^{(h)} X_{t-h} = \Pi_{H_{t-1}^h}(X_t) \text{ et } \eta_t^{(h)} = X_t - \Pi_{H_{t-1}^h}(X_t).$$

On supposera implicitement que les coefficients $\beta_j^{(h)}$ sont uniques et donc que les variables $X_{t-1}, ..., X_{t-h}$ ne sont pas liées linéairement (l'espace H_{t-1}^{h-1} est donc de dimension h-1). On a alors la propriété suivante.

Proposition 1.7. Pour tout $h \ge 1$, $\alpha(h) = \beta_h^{(h)}$.

Avant de donner la preuve, remarquons la propriété fondamentale suivante: la corrélation partielle est entièrement déterminée par la fonction d'auto-covariance (elle n'apporte donc en théorie pas plus d'information même si elle s'interprète différemment). De plus, la fonction d'auto-covariance $\gamma(.)$ étant symétrique, la série renversée dans le temps

$$\overset{\leftarrow}{X_t} := X_{-t} \ , \ t \in \mathbb{Z}$$

est donc également stationnaire avec le même ACF et donc le même PACF. Une conséquence immédiate est que les coefficients de la régression linéaire de X_t par le passé $X_{t-1},...,X_{t-h}$ sont les mêmes (en ordre inversé) que ceux de la régression de X_t par le futur $X_{t+1},...,X_{t+h}$, autrement dit

$$X_{t} = \beta_{1}^{(h)} X_{t+1} + \dots + \beta_{h}^{(h)} X_{t+h} + \xi_{t}^{(h)}$$

où $\xi_t^{(h)}$ est orthogonal à $X_{t+1},...,X_{t+h}$ et de même variance que $\eta_t^{(h)}$.

Preuve de la proposition. On suppose toujours sans perte de généralité que $(X_t)_{t\in\mathbb{Z}}$ est centrée. Pour h=1, la régression linéaire $X_t=\beta_1^{(1)}X_{t-1}+\eta_t^{(1)}$ a pour solution

$$\beta_1^{(1)} = \frac{\operatorname{cov}(X_t, X_{t-1})}{\operatorname{var}(X_{t-1})} = \operatorname{cor}(X_t, X_{t-1}),$$

par stationnarité (et donc variance constante) de $(X_t)_{t\in\mathbb{Z}}$. Pour $h\geq 2$, on a d'après l'équation (1)

$$\Pi_{H_{t-1}^{h-1}}(X_t) = \left(\beta_1^{(h)} X_{t-1} + \dots + \beta_{h-1}^{(h)} X_{t-h+1}\right) + \beta_h^{(h)} \Pi_{H_{t-1}^{h-1}}(X_{t-h}).$$

Comme $\eta_t^{(h)}$ est orthogonal à H_{t-1}^h (et donc aussi à H_{t-1}^{h-1}), on déduit

$$X_{t} - \Pi_{H_{t-1}^{h-1}}(X_{t}) = \beta_{h}^{(h)} \left(X_{t-h} - \Pi_{H_{t-1}^{h-1}}(X_{t-h}) \right) + \eta_{t}^{(h)}, \tag{2}$$

et

$$\operatorname{cov}\left(X_{t} - \Pi_{H_{t-1}^{h-1}}(X_{t}), X_{t-h} - \Pi_{H_{t-1}^{h-1}}(X_{t-h})\right) = \beta_{h}^{(h)} \operatorname{var}\left(X_{t-h} - \Pi_{H_{t-1}^{h-1}}(X_{t-h})\right).$$

Par symétrie dans le temps (les séries X_t et $\overset{\leftarrow}{X}_t = X_{-t}$ ont les mêmes fonctions d'auto-covariance), on sait que var $\left(X_{t-h} - \Pi_{H_{t-1}^{h-1}}(X_{t-h})\right) = \operatorname{var}\left(X_t - \Pi_{H_{t-1}^{h-1}}(X_t)\right)$.

En pratique, on peut donc estimer l'auto-corrélation partielle $\alpha(h)$ par le coefficient $\widehat{\alpha}(h) = \widehat{\beta}_h^{(h)}$ de la régression linéaire des observations X_t en fonction du passé jusqu'à l'ordre h. On peut aussi utiliser une méthode itérative comme l'algorithme de Durbin-Levinson.

Proposition 1.8 (Algorithme de Durbin-Levinson). Pour tout $h \geq 1$, soit $\tau_h = \text{var}(\eta_t^{(h)})$. Alors, $\tau_1 = \gamma(0)(1 - \rho(1)^2)$, $\beta_1^{(1)} = \rho(1)$ et pour $h \geq 2$,

•
$$\alpha(h) = \beta_h^{(h)} = \frac{1}{\tau_{h-1}} \left(\gamma(h) - \sum_{j=1}^{h-1} \beta_j^{(h-1)} \gamma(h-j) \right)$$

- $\tau_h = \tau_{h-1} (1 \alpha(h)^2)$
- $\beta_j^{(h)} = \beta_j^{(h-1)} \alpha(h)\beta_{h-j}^{(h-1)}, j = 1, ..., h-1.$

Preuve. D'après la preuve précédente, $X_t = \beta_1^{(1)} X_{t-1} + \eta_t^{(1)}$ avec $\beta_1^{(1)} = \rho(1)$. On a bien

$$\tau_1 = \operatorname{var}(\eta_t^{(1)}) = \operatorname{var}(X_t) - \beta_1^{(1)2} \operatorname{var}(X_{t-1}) = \gamma(0) (1 - \rho(1)^2).$$

En remarquant que $X_t - \Pi_{H_{t-1}^{h-1}}(X_t) = \eta_t^{(h-1)}$, on trouve toujours d'après la preuve précédente

$$\beta_h^{(h)} = \frac{\operatorname{cov}\left(X_t - \Pi_{H_{t-1}^{h-1}}(X_t), X_{t-h} - \Pi_{H_{t-1}^{h-1}}(X_{t-h})\right)}{\operatorname{var}\left(\eta_t^{(h-1)}\right)} = \frac{\operatorname{cov}\left(X_t - \Pi_{H_{t-1,h-1}}(X_t), X_{t-h}\right)}{\tau_{h-1}}.$$

Or $\Pi_{H_{t-1}^{h-1}}(X_t) = \beta_1^{(h-1)} X_{t-1} + \ldots + \beta_{h-1}^{(h-1)} X_{t-h+1}$, d'où

$$\beta_h^{(h)} = \frac{1}{\tau_{h-1}} \left(\gamma(h) - \sum_{i=1}^{h-1} \beta_j^{(h-1)} \gamma(h-j) \right).$$

Pour montrer le deuxième point, on applique la variance dans l'équation (2), ce qui donne

$$\tau_{h-1} = \alpha(h)^2 \tau_{h-1} + \tau_h.$$

Enfin, comme $H_{t-1}^{h-1} \subseteq H_{t-1}^h$, $\Pi_{H_{t-1}^{h-1}}(X_t) = \Pi_{H_{t-1}^{h-1}}(\Pi_{H_{t-1}^h}(X_t))$, ou encore

$$\beta_1^{(h-1)} X_{t-1} + \dots + \beta_{h-1}^{(h-1)} X_{t-h+1} = \beta_1^{(h)} X_{t-1} + \dots + \beta_h^{(h)} X_{t-h+1} + \beta_h^{(h)} \Pi_{H^{h-1}} (X_{t-h}).$$

Par symétrie par rapport au temps, $\Pi_{H_{t-1}^{h-1}}(X_{t-h}) = \beta_1^{(h-1)} X_{t-h+1} + ... + \beta_{h-1}^{(h-1)} X_{t-1}$. On conclut par identification des coefficients.

2 Tendance et saisonnalité

Une série temporelle qui comporte une composante déterministe non constante (croissance linéaire, phénomène saisonnier, etc...) n'est pas stationnaire (pourquoi?). Dans ces cas-là, on va chercher à modéliser la série en la décomposant en une partie déterministe (non-aléatoire) et une partie résiduelle stochastique. Par exemple:

$$X_t = M_t + S_t + Z_t, t \in \mathbb{N}$$

οù

- M_t est la tendance, soit une fonction "simple" (par exemple linéaire, quadratique, linéaire par morceaux etc...) du temps qui dicte le comportement global du processus.
- S_t est la composante saisonnière, qui est une fonction périodique en t c'est-à-dire telle que S_t = S_{t+s} pour un s > 0, pour tout t ∈ N. Le plus petit s > 0 pour lequel cette propriété est vérifiée est appelé la période. Par convention, on imposera que S_t soit centrée: ∑_{t=1}^s S_t = 0 sans perte de généralité puisque une valeur moyenne non-nulle peut être reportée sur la tendance.
- Z_t est la composante stochastique qui modélise les aléas ou fluctuations autour de la partie déterministe. Ce processus n'est en général pas un bruit blanc mais on pourra souvent chercher à le modéliser par un processus stationnaire (ou dérivé d'un processus stationnaire).

Ce modèle simple permet d'isoler la ou les composantes déterministes, pour ensuite les éliminer et modéliser la composante résiduelle Z_t . Détecter des phénomènes de tendances et ou saisonalité se fait à partir du contexte (données mensuelles implique potentiellement une période de 12, hebdomadaire de 7 etc...) et des représentations graphiques de la série. Isoler la composante stochastique permet en général de détecter des phénomènes de dépendances masqués par les composantes déterministes.

2.1 Modélisation paramétrique

Une approche naturelle consiste à poser un modèle paramétrique sur la tendance et/ou la composante saisonnière pour les estimer.

Estimation de la tendance. On ajuste la tendance par un modèle linéaire en fonction du temps, typiquement par une fonction affine $S_t = a_0 + a_1 t$, polynomiale $S_t = a_0 + a_1 t + ... + a_k t^k$ ou autre (exponentielle, puissance, etc...). Les paramètres a_i sont estimés par la méthode des moindres carrés ordianires.

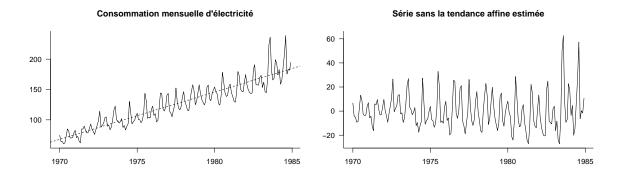


Figure 2: Estimation par MCO et élimination de la tendance affine.

Estimation de la composante saisonnière. On travaille sur la série $X_t - \widehat{M}_t$ où \widehat{M}_t est l'estimation de la tendance. La période p de la composante saisonnière (s'il y en a) est déterminée par le contexte ou "à la main" au vu des données (représentation graphique de la série, ACF et PACF). On peut alors proposer une modélisation paramétrique périodique sur la composante saisonnière S_t , par exemple $S_t = a \sin(2\pi t/p) + b \cos(2\pi t/p)$ ou alors l'estimer de façon "non-paramétrique" par la moyenne sur les temps espacés de s. Il y a donc s-1 valeurs à estimer puisqu'on a la condition d'identifiabilité $\sum_{t=1}^s S_t = 0$.

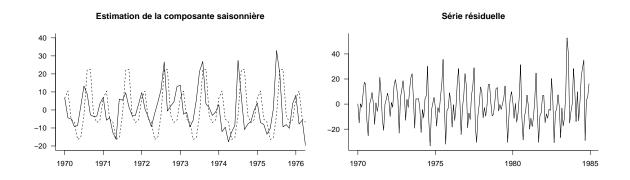
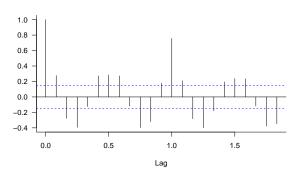


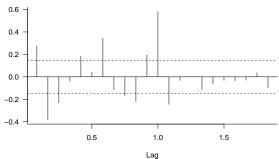
Figure 3: Estimation non-paramétrique de la composante saisonnière et série résiduelle.

On cherche maintenant à modéliser la série résiduelle $\widehat{Z}_t = X_t - \widehat{M}_t - \widehat{S}_t$ comme un processus stochastique, idéalement stationnaire. On peut étudier ses ACF et PACF estimés. Même s'il est toujours utile de les représenter graphiquement, ceux-ci n'ont de sens que si la série est stationnaire.

ACF de la série résiduelle

PACF de la série résiduelle





2.2 Filtrage

Définition 2.1. L'opérateur retard est l'opérateur qui à une série temporelle $(X_t)_{t\in\mathbb{Z}}$ associe la même série retardée d'une unité de temps. On le note généralement B (pour backward) ou L (pour lag).

Par abus de notation, on appliquera directement l'opérateur B aux valeurs de la série en écrivant

$$BX_t = X_{t-1}.$$

Son inverse B^{-1} décale la série d'une unité de temps dans le futur,

$$B^{-1}X_t = X_{t+1}$$

On a bien $B \circ B^{-1} = B^{-1} \circ B = I$, où I est l'identité. On note B^k et B^{-k} les opérateurs B et B^{-1} composés j fois,

$$B^k X_t = X_{t-k}$$
 et $B^{-k} X_t = X_{t+k}$.

Définition 2.2. Un filtre linéaire est un opérateur de la forme

$$F = \sum_{k \in \mathbb{Z}} \psi_k B^k,$$

où $\psi = (\psi_k)_{k \in \mathbb{Z}}$ est une suite complexe sommable: $\sum_{k \in \mathbb{Z}} |\psi_k| < \infty$. On dit que le filtre est causal si les coefficients $\psi_k, k < 0$ sont nuls.

Un filtre peut être vu comme une fonction $\Psi: x \mapsto \sum_{k \in \mathbb{Z}} \psi_k x^k$ appliquée à l'opérateur retard B. On note alors $F = \Psi(B)$ et

$$FX_t = \Phi(B)X_t = \sum_{k \in \mathbb{Z}} \psi_k B^k X_t = \sum_{k \in \mathbb{Z}} \psi_k X_{t-k} \ , \ t \in \mathbb{Z},$$

où on utilise ici la convention $B^0=I$. Si le filtre $F=\Psi(B)$ est causal, alors Ψ est analytique de rayon de convergence supérieur ou égal à 1.

Lemme 2.3. L'image d'un processus stationnaire par un filtre linéaire est stationnaire.

Preuve. Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire de fonction d'auto-covariance γ et $F=\sum_{k\in\mathbb{Z}}\psi_kB^k$ un filtre linéaire. Pour $h\geq 1$,

$$\sum_{j,k\in\mathbb{Z}} |\psi_j \psi_k \gamma(h+j-k)| \leq \gamma(0) \bigg(\sum_{k\in\mathbb{Z}} |\psi_k|\bigg)^2 < \infty.$$

On vérifie alors facilement que l'espérance de FX_t est constante et pour $h \ge 0$, la covariance à l'ordre h

$$cov(FX_t, FX_{t-h}) = cov\left(\sum_{k \in \mathbb{Z}} \psi_k X_{t-k}, \sum_{k \in \mathbb{Z}} \psi_k X_{t-h-k}\right) = \sum_{j,k \in \mathbb{Z}} \psi_j \psi_k \gamma(h+j-k)$$

ne dépend pas de t.

Exercice. Montrer que la somme F+G et la composée $F\circ G$ (que l'on note souvent FG par abus de notation) de deux filtres linéaires F,G sont des filtres linéaires. Vérifier que ces opérations sont commutatives et qu'elles préservent la causalité.

Définition 2.4. L'opérateur de différenciation discret est défini par $\Delta := I - B$.

Le filtre de différenciation Δ renvoie les variations d'une série temporelle

$$\Delta X_t = X_t - X_{t-1}.$$

ce qui s'apparente à une dérivée discrète. De même, la composée k-fois est l'opérateur de différenciation à l'ordre k,

$$\Delta^k X_t = (I - B)^k X_t = \sum_{j=0}^k (-1)^j \frac{k!}{j!(k-j)!} B^j.$$

Proposition 2.5. L'opérateur de différenciation à l'ordre $k \geq 1$ rend constant tout polynôme de degré inférieur ou égal à k du temps.

Preuve. Soit $M_t = a_0 + a_1 t + ... + a_k t^k$, on a par linéarité

$$\Delta^k M_t = a_0 \Delta^k 1 + a_1 \Delta^k t + \dots + a_k \Delta^k t^k.$$

Il suffit de montrer le résultat pour les monômes t^{j} . Pour k=1,

$$\Delta t = (I - B)t = t - (t - 1) = 1.$$

On procède par récurrence, supposons que le résultat est vrai pour k-1,

$$\Delta^k t^k = \Delta^{k-1} \Delta t^k = \Delta^{k-1} (t^k - (t-1)^k).$$

Or $t^k - (t-1)^k$ est un polynôme de degré k-1 en t, le résultat est donc vrai à l'ordre k.

L'opérateur de différenciation permet d'éliminer une tendance polynomiale dans une série temporelle de la forme $X_t = M_t + S_t + Z_t, t \in \mathbb{Z}$. Si M_t est un polynôme de degré k du temps alors, la série $\Delta_k X_t$ ne comporte plus de tendance polynômiale (seulement une constante) mais conserve une composante saisonnière, différente de l'originale.

Exercice. Montrer que si $(S_t)_{t\in\mathbb{Z}}$ est une série périodique de période $p\geq 1$, alors sa série différenciée à l'ordre $k\geq 1$ l'est également. Montrer que cette dernière est de somme nulle sur sa période même si $(S_t)_{t\in\mathbb{Z}}$ ne l'est pas.

Proposition 2.6. Soit $F = \sum_{j \in \mathbb{Z}} a_j B^j$ un filtre linéaire tel que pour $p \ge 1$,

$$\sum_{j\in\mathbb{Z}}a_j=1 \quad et \quad \sum_{j\in\mathbb{Z}}j^ka_j=0 \quad pour \ k=1,...,p.$$

Alors, F laisse invariant tout polynôme de degré inférieur ou égal à p.

Preuve. Exercice. \Box

Définition 2.7. Une moyenne mobile est un filtre F associé à une suite $(a_j)_{j\in\mathbb{Z}}$ non nulle à support fini. Autrement dit, un filtre F est une moyenne mobile s'il existe $k_1, k_2 \in \mathbb{N}$ et des coefficients $a_{-k_1}, ..., a_{k_2} \in \mathbb{R}$ avec $a_{-k_1} \neq 0$, $a_{k_2} \neq 0$ tels que

$$F = \sum_{j=-k_1}^{k_2} a_j B^j.$$

La moyenne mobile arithmétique est le filtre linéaire uniforme centré en 0 qui à une série X_t associe la moyenne des valeurs autour de X_t . Formellement, pour $k \in \mathbb{N}^*$, la moyenne mobile arithmétique d'ordre 2k+1 est le filtre

$$M_{2k+1} = \frac{1}{2k+1} \sum_{j=-k}^{k} B^k.$$

Dans le cas pair, on définit la moyenne mobile arithmétique d'ordre 2k par

$$M_{2k} = \frac{1}{4k}B^{-k} + \frac{1}{2k}\sum_{i=-k+1}^{k-1} B^k + \frac{1}{4k}B^k.$$

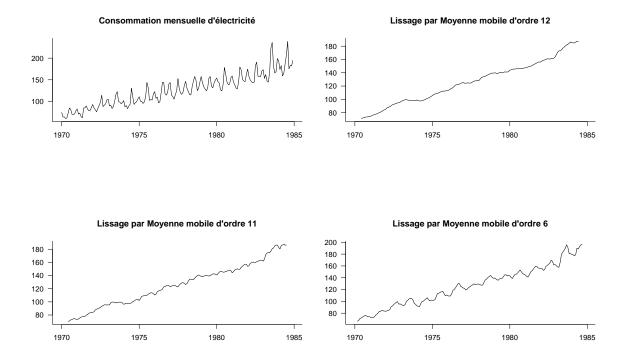
Proposition 2.8. La moyenne mobile arithmétique d'ordre p annule toute série périodique de période p de somme nulle sur sa période.

Preuve.

Exercice. Montrer que les moyennes arithmétiques laissent invariantes les fonctions affines du temps.

Les propriétés des moyennes mobiles arithmétiques permettent d'estimer la tendance d'une série temporelle de manière non paramétrique. Soit $X_t = M_t + S_t + Z_t$ une série temporelle avec M_t la tendance globale 'lisse' (pas nécessairement linéaire), S_t la composante saisonnière de période $p \geq 2$ et Z_t la partie résiduelle. Appliquer la moyenne mobile arithmétique M_p permet d'éliminier la composante saisonnière

tout en modifiant peu la tendance. Cela ne fonctionne que si l'ordre de la moyenne mobile est égal à (ou multiple de) la période. On appelle ce procédé lissage par moyenne mobile.



Les opérateurs de différenciation et moyenne mobile ont en quelque sorte des rôles complémentaires pour la décomposition en tendance et composante saisonnière d'une série temporelle.

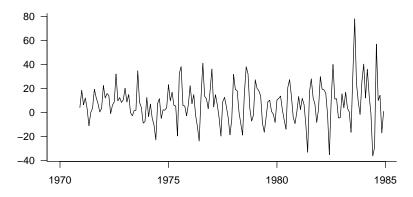
Proposition 2.9. Le filtre $I - B^p$ annule les séries périodiques de période p et rend constantes les fonctions affines du temps.

$$Preuve.$$
 Immédiat.

Le filtre $I-B^p$ permet donc d'éliminer d'un coup tendance affine et composante saisonnière d'une série temporelle de la forme $X_t=(a+bt)+S_t+Z_t, t\in\mathbb{Z}$ pour se ramener à la série "purement" stochastique

$$(I - B^p)(X_t) = bp + Z_t - Z_{t-p} , t \in \mathbb{Z}.$$

Consommation électrique avec filtre I-B¹²



Définition 2.10. Le pouvoir de réduction de variance d'un filtre F est la quantité

$$RV(F) = \frac{\operatorname{var}(F\epsilon_t)}{\operatorname{var}(\epsilon_t)}$$

 $o\dot{u}$ $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc.

Exercice. Calculer le pouvoir de réduction de variance des opérateurs de différenciation Δ^k et des moyennes mobiles arithmétiques M_p . Commenter.

Comme on peut s'y attendre, lisser un processus en applicant un filtre moyenne mobile réduit sa variance. En contrepartie, cela a tendance à créer/amplifier les corrélations temporelles.

Un filtre linéaire F est inversible s'il existe un filtre linéaire noté F^{-1} tel que $FF^{-1} = F^{-1}F = I$.

Lemme 2.11. Soit le filtre linéaire $F_{\lambda} = (I - \lambda B)$ avec $\lambda \in \mathbb{C}$. Alors,

- $Si |\lambda| < 1$, F_{λ} est inversible et $F_{\lambda}^{-1} = \sum_{k \geq 0} \lambda^k B^k$ est causal.
- $Si |\lambda| > 1$, F_{λ} est inversible et $F_{\lambda}^{-1} = -\sum_{k>1} \lambda^{-k} B^{-k}$ n'est pas causal.
- $Si |\lambda| = 1$, F_{λ} n'est pas inversible.

Preuve. Soit $(\psi_k)_{k\in\mathbb{Z}}$ telle que

$$I = \sum_{k \in \mathbb{Z}} \psi_k B^k (I - \lambda B) = \sum_{k \in \mathbb{Z}} (\psi_k - \lambda \psi_{k-1}) B^k.$$

Par identification, $\psi_0 = 1 + \lambda \psi_{-1}$ et $\psi_k = \lambda \psi_{k-1}$ pour tout $k \neq 0$, ce qui donne en raisonnant par récurrence $\psi_k = \lambda^k \psi_0$ pour $k \geq 1$ et $\psi_k = \lambda^k (\psi_0 - 1)$ pour $k \leq -1$. On sait de plus que $(\psi_k)_{k \in \mathbb{Z}}$ doit vérifier

$$\sum_{k \in \mathbb{Z}} |\psi_k| = |\psi_0| \sum_{k \ge 0} |\lambda^k| + |\psi_0 - 1| \sum_{k \ge 1} |\lambda^{-k}| < \infty.$$

Cette condition est vérifiée pour $\psi_0 = 1$ si $|\lambda| < 1$, $\psi_0 = 0$ si $|\lambda| > 1$ et n'est jamais vérifiée si $|\lambda| = 1$.

Ce résultat important permet d'identifier les conditions sous lesquelles un polynôme de l'opérateur retard est inversible, d'inverse causal ou non inversible. Ces conditions ne dépendent que du module des racines du polynôme.

Exercice. Soient $\Psi: x \mapsto \sum_{k \geq 0} \psi_k x^k$ et $\Phi: x \mapsto \sum_{k \geq 0} \phi_k x^k$ deux fonctions analytiques de rayon de convergence supérieur ou égal à 1.

- 1. Justifier que $\Psi(B)$ et $\Phi(B)$ sont des filtres linéaires causaux.
- 2. Montrer que $\Psi(B)\Phi(B) = (\Psi \times \Phi)(B)$.
- 3. Déduire que si Ψ n'a pas de racine dans la boule unité $\mathcal{B} = \{z \in \mathbb{C} : |z| \leq 1\}$, alors $\Psi(B)^{-1} = (1/\Psi)(B)$ est un filtre linéaire causal.

3 Processus linéaires standards

Définition 3.1. Un processus $(X_t)_{t\in\mathbb{Z}}$ est dit linéaire s'il peut s'écrire

$$X_t = m + \sum_{k \in \mathbb{Z}} \psi_k \epsilon_{t-k} \ , \ t \in \mathbb{Z}$$

 $o\grave{u}\ m\in\mathbb{R},\ (\epsilon_t)_{t\in\mathbb{Z}}\ est\ un\ bruit\ blanc\ et\ (\psi_k)_{k\in\mathbb{Z}}\ est\ une\ suite\ r\'eelle\ telle\ que\ \psi_0=1\ et\ \sum_{k\in\mathbb{Z}}|\psi_k|<\infty.$

La condition $\psi_0 = 1$ est simplement une condition d'identification. Sous réserve que la suite $(\psi_k)_{k \in \mathbb{Z}}$ n'est pas identiquement nulle, on peut s'y ramener en modifiant le bruit blanc.

Un processus linéaire (recentré) est l'image d'un bruit blanc par un filtre linéaire $\Psi(B) := \sum_{k \geq \in \mathbb{Z}} \psi_k B^k$,

$$X_t - m = \sum_{k \in \mathbb{Z}} \psi_k \epsilon_{t-k} = \Psi(B) \epsilon_t , \ t \in \mathbb{Z}.$$

On dira que cette représentation est

• causale si le filtre $\Psi(B)$ est causal et donc que les ψ_k sont nuls pour k < 0:

$$X_t - m = \sum_{k>0} \psi_k \epsilon_{t-k} , \ t \in \mathbb{Z}.$$

• inversible si le filtre $\Psi(B)$ est inversible et que son inverse $\Psi(B)^{-1} = (1/\Psi)(B)$ est causal, c'est-à-dire, s'il existe des coefficients $(\zeta_k)_{k\geq 0}$ avec $\sum_{k>0} |\zeta_k| < \infty$ tels que

$$\epsilon_t = \sum_{k>0} \zeta_k (X_{t-k} - m) , t \in \mathbb{Z}.$$

En notant $\Psi: z \mapsto \sum_{k \in \mathbb{Z}} \psi_k z^k$, la représentation $X_t = m + \Psi(B)\epsilon_t, t \in \mathbb{Z}$ est causale si Ψ est analytique sur la boule unité $\mathcal{B} = \{z \in \mathbb{C} : |z| \le 1\}$ et inversible si la fonction $1/\Psi$ est analytique sur \mathcal{B} . Si l'écriture est causale, elle est donc également inversible si Ψ n'a pas de racine dans \mathcal{B} . En supposant $\psi_0 = 1$ sans perte de généralité, on a alors la relation

$$\sum_{k>0} \zeta_k z^k = \frac{1}{\Psi(z)} = \frac{1}{1 + \sum_{k\geq 1} \psi_k z^k} = 1 - \sum_{k>1} \psi_k z^k + \left(\sum_{k>1} \psi_k z^k\right)^2 - \dots, \ z \in \mathcal{B}.$$

On retrouve par identification

$$\zeta_0 = 1 \ \text{et} \ \zeta_k = \sum_{p \geq 1} (-1)^p \sum_{k_1 + \ldots + k_p = k} \psi_{k_1} \ldots \psi_{k_p} \ , \ k = 1, 2, \ldots$$

On vérifie bien que la condition $\sum_{k\geq 0} |\zeta_k| < \infty$ est vérifiée puisque Θ n'a pas de racine sur \mathcal{B} et donc que $1/\Psi(z)$ converge absolument sur \mathcal{B} .

Proposition 3.2. Un processus linéaire $X_t = m + \sum_{k \in \mathbb{Z}} \psi_k \epsilon_{t-k}, t \in \mathbb{Z}$ est stationnaire, d'espérance m et de fonction d'auto-covariance

$$\gamma(h) = \sigma^2 \sum_{k \in \mathbb{Z}} \psi_k \psi_{k-h} , h \in \mathbb{Z},$$

 $o\dot{u} \ \sigma^2 = \text{var}(\epsilon_t).$

Preuve. On écrit

$$\operatorname{var}(X_t) \leq \mathbb{E}\left(\sum_{k \in \mathbb{Z}} \psi_k \epsilon_{t-k}\right)^2 \leq \mathbb{E}\left(\sum_{k \in \mathbb{Z}} |\psi_k| |\epsilon_{t-k}|\right)^2 = \mathbb{E}\left(\lim_{N \to \infty} \left(\sum_{k=-N}^N |\psi_k \epsilon_{t-k}|\right)^2\right).$$

Par le théorème de convergence monotone,

$$\operatorname{var}(X_t) \leq \sigma^2 \lim_{N \to \infty} \mathbb{E}\left(\sum_{j \mid k = -N}^{N} |\psi_j| |\psi_k|\right) \leq \sigma^2 \left(\sum_{k \in \mathbb{Z}} |\psi_k|\right)^2 < \infty.$$

On calcule facilement $\mathbb{E}(X_t) = m$ et pour $h \in \mathbb{Z}$,

$$cov(X_t, X_{t+h}) = \mathbb{E}\left(\sum_{k \in \mathbb{Z}} \psi_k \epsilon_{t-k} \sum_{k \in \mathbb{Z}} \psi_k \epsilon_{t+h-k}\right) = \sum_{j,k \in \mathbb{Z}} \psi_j \psi_k \mathbb{E}\left(\epsilon_{t-j} \epsilon_{t+h-k}\right),$$

où $\mathbb{E}(\epsilon_{t-j}\epsilon_{t+h-k})$ vaut σ^2 si j=k-h et 0 sinon.

Si l'écriture est causale, on retrouve la décomposition de Wold avec $(\epsilon_t)_{t\in\mathbb{Z}}$ le processus des innovations si on impose $\psi_0=1$

$$X_t = m + \sum_{k>1} \psi_k \epsilon_{t-k} + \epsilon_t.$$

La fonction d'auto-covariance s'écrit alors

$$\gamma(h) = \gamma(-h) = \sigma^2 \sum_{k>h} \psi_k \psi_{k-h} , h \ge 0.$$

3.1 Moyennes mobiles

Définition 3.3. Une moyenne mobile d'ordre $q \ge 1$, notée MA(q) est un processus de la forme

$$X_t = m + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} , t \in \mathbb{Z},$$

 $o\dot{u} \; \theta_1, ..., \theta_q \; sont \; des \; r\'eels, \; \theta_q \neq 0 \; et \; (\epsilon_t)_{t \in \mathbb{Z}} \; est \; un \; bruit \; blanc.$

Une moyenne mobile (MA) est donc un processus linéaire causal pour lequel les coefficients ψ_k sont nuls au-delà d'un certain rang q, l'ordre du processus.

Proposition 3.4. Soit $X_t = m + \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, t \in \mathbb{Z}$ un processus MA(q) avec $\theta_0 = 1$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc de variance σ^2 . Alors $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, de fonction d'auto-covariance

$$\gamma(h) = \sigma^2 \sum_{k=0}^{q-h} \theta_k \theta_{k+h} , h = 0, 1, ..., q$$

 $et \gamma(h) = 0 \ pour \ h > q.$

Preuve. Exercice.

Exercice. Montrer que la corrélation d'ordre 1 d'un processus MA(q) est strictement inférieure à 1 en valeur absolue. Calculer la valeur maximale en fonction de q.

Une moyenne mobile a donc sa fonction d'auto-covariance nulle à partir d'un certain rang. C'est en fait une caractérisation des processus MA.

Proposition 3.5. Soit $q \ge 1$ et $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire d'espérance m dont la fonction l'auto-covariance γ vérifie $\gamma(q) \ne 0$ et $\gamma(h) = 0$ pour h > q. Alors, il existe des coefficients $\theta_1, ..., \theta_q$ tels que

$$X_t = m + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} , t \in \mathbb{Z},$$

 $o\dot{u}$ $(\epsilon_t)_{t\in\mathbb{Z}}$ est le bruit blanc d'innovations de $(X_t)_{t\in\mathbb{Z}}$.

Preuve. On suppose m=0 sans perte de généralité. Soit $\epsilon_t=X_t-\Pi_{H_{t-1}}(X_t), t\in\mathbb{Z}$ l'innovation, on sait que le passé H_t du processus au temps t vérifie $H_t=H_{t-1}\oplus\mathrm{vect}\{\epsilon_t\}$ avec H_{t-1} et $\mathrm{vect}\{\epsilon_t\}$ orthogonaux. Par récurrence, on a donc pour tout t

$$H_t = H_{t-q} \oplus E_t^q$$
,

où $E_t^q = \text{vect}\{\epsilon_{t-q+1},...,\epsilon_t\}$ et H_{t-q} et E_t^q sont orthogonaux. On déduit

$$X_t = \Pi_{H_{t-1}}(X_t) + \epsilon_t = \Pi_{H_{t-q-1}}(X_t) + \Pi_{E_{t-1}^q}(X_t) + \epsilon_t.$$

Comme $\mathbb{E}(X_t X_{t-h}) = 0$ pour h > q, on sait que $\Pi_{H_{t-q-1}}(X_t) = 0$. Le résultat est donc vérifié pour $\theta_1, ..., \theta_q$ les coefficients de la régression linéaire de X_t par $\epsilon_{t-1}, ..., \epsilon_{t-q}$. De plus, la covariance à l'ordre q, $\operatorname{cov}(X_t, X_{t-q}) = \theta_q \sigma^2$ est non nulle, la moyenne mobile est donc d'ordre q.

Attention, ce résultat n'implique pas qu'un processus MA(q) vérifiant

$$X_t = m + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} , t \in \mathbb{Z}$$

avec $(\epsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc, ait pour innovation ϵ_t . En réalité, ce n'est l'innovation de X_t que si cette représentation est inversible. Dans le cas contraire, le résultat implique en revanche qu'il existe des coefficients $\widetilde{\theta}_1, ..., \widetilde{\theta}_q$ et un bruit blanc $(\widetilde{\epsilon}_t)_{t\in\mathbb{Z}}$ (le processus des innovations) tels que

$$X_t = m + \widetilde{\epsilon}_t + \widetilde{\theta}_1 \widetilde{\epsilon}_{t-1} + \dots + \widetilde{\theta}_q \widetilde{\epsilon}_{t-q} , \ t \in \mathbb{Z}.$$

C'est la décomposition de Wold du processus.

Exercice. Montrer que la représentation $X_t = \epsilon_t + \theta \epsilon_{t-1}, t \in \mathbb{Z}$ est inversible si et seulement si $|\theta| < 1$.

Soit le polynôme de degré $q, \Theta: x \mapsto 1 + \theta_1 x + \dots + \theta_q x^q$, on peut donc écrire

$$X_t = m + \Theta(B)\epsilon_t, \ t \in \mathbb{Z}.$$

Théorème 3.6. Soit $X_t = m + \Theta(B)\epsilon_t, t \in \mathbb{Z}$ un processus MA(q) avec Θ un polynôme de degré q tel que $\Theta(0) = 1$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc. Alors, la représentation est inversible si et seulement si Θ n'a aucune racine dans la boule unité complexe $\mathcal{B} = \{z \in \mathbb{C} : |z| \leq 1\}$. Dans ce cas, $(\epsilon_t)_{t \in \mathbb{Z}}$ est le bruit blanc des innovations de $(X_t)_{t \in \mathbb{Z}}$.

Preuve. On suppose m=0 sans perte de généralité. Si Θ n'a pas de racine sur \mathcal{B} , alors $1/\Theta$ est analytique sur \mathcal{B} et

$$\sum_{k>1} \zeta_k z^k = \frac{1}{\Theta(z)} = \frac{1}{1 + \sum_{k=1}^q \theta_k z^k} = 1 - \sum_{k=1}^q \theta_k z^k + \left(\sum_{k=1}^q \theta_k z^k\right)^2 - \dots$$

on retrouve par identification $\zeta_0=1$ et $\zeta_k=\sum_{p\geq 1}\sum_{k_1+\ldots+k_p=k}(-1)^p\theta_{k_1}\ldots\theta_{k_p}$. La condition $\sum_{k\geq 0}|\zeta_k|<\infty$ tient au fait que $1/\Theta$ converge absolument sur \mathcal{B} . Réciproquement, si $\epsilon_t=\sum_{k\geq 0}\zeta_kX_{t-k}, t\in\mathbb{Z}$ avec $\sum_{k\geq 0}|\zeta_k|<\infty$. En posant $\theta_0=1$,

$$\epsilon_t = \sum_{k>0} \zeta_k \Theta(B) \epsilon_{t-k} = \sum_{k>0} \zeta_k \left(\sum_{j=0} \theta_j B^{j+k} \epsilon_t \right) = \sum_{k>0} \left(\sum_{j=0}^k \zeta_{k-j} \theta_j \right) \epsilon_{t-k}.$$

Par identification, $\zeta_0 \theta_0 = 1$ et $\sum_{j=0}^k \zeta_{k-j} \theta_j = 0$ pour tout $k \ge 1$. Or, pour tout $z \in \mathcal{B}$,

$$\Theta(z) \times \sum_{k>0} \zeta_k z^k = \sum_{k>0} \left(\sum_{j=0}^k \zeta_{k-j} \theta_j \right) z^k = 1.$$

Comme $|\sum_{k\geq 0} \zeta_k z^k| \leq \sum_{k\geq 0} |\zeta_k| < \infty$ pour $z\in \mathcal{B}$, on conclut que $\Theta(z)$ n'a pas de racine sur \mathcal{B} . Enfin, si la représentation est inversible, alors $\zeta_0 = 1/\Theta(0) = 1$ et

$$X_t = -\sum_{k>1} \zeta_k X_{t-k} + \epsilon_t , \ t \in \mathbb{Z}.$$

Pour tout $h \ge 1$, $cov(\epsilon_t, X_{t-h}) = \sum_{k=1}^q \theta_k cov(\epsilon_t, \epsilon_{t-h-k}) = 0$. On déduit

$$X_t - \Pi_{H_{t-1}}(X_t) = X_t - \sum_{k \ge 1} \zeta_k X_{t-k} = \epsilon_t , \ t \in \mathbb{Z}.$$
 Exercice. Soit $z \in \mathbb{C}, |z| < 1$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc.

1. Montrer que $\xi_z := (I - zB)^{-1}(I - z^{-1}B) = \sum_{k \geq 0} \tau_k B^k$ où $\tau_k = \tau_k(z)$ vérifie

$$\tau_0 = 1$$
 et $\tau_k = z^{k-2}(z^2 - 1)$, $k = 1, 2,$

- 2. Montrer que $\eta_t = \xi_z \epsilon_t, t \in \mathbb{Z}$ est un bruit blanc. En déduire que ξ_z et ξ_z^{-1} préservent la "blancheur".
- 3. Soit le processus $\mathrm{MA}(q)$ $X_t = \Theta(B)\epsilon_t, t \in \mathbb{Z}$, où Θ est un polynôme de degré q sans racine unitaire tel que $\Theta(0)=1$. On note $z_1,...,z_q$ les racines de Θ comptées avec multiplicité dont $z_1,...,z_r$ les racines de module < 1. A l'aide des opérateurs $\xi_{z_1},...,\xi_{z_r}$, représenter le processus sous forme causale en faisant apparaître son bruit blanc des innovations (indication: écrire $\Theta(z)$ $(1-z_1^{-1}z)...(1-z_q^{-1}z)).$
- 4. Calculer l'innovation de $X_t = \epsilon_t + \frac{3}{2}\epsilon_{t-1} \epsilon_{t-2}, t \in \mathbb{Z}$.

Processus auto-régressifs 3.2

Définition 3.7. Un processus auto-régressif d'ordre p > 1, noté AR(p) est un processus stationnaire vérifiant

$$X_t = m + \phi_1(X_{t-1} - m) + \dots + \phi_n(X_{t-n} - m) + \epsilon_t, \ t \in \mathbb{Z},$$

où $m, \phi_1, ..., \phi_q$ sont des réels, $\phi_p \neq 0$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc.

Soit Φ le polynôme $z \mapsto 1 - \phi_1 z - \dots - \phi_p z^p$, on peut écrire

$$\Phi(B)(X_t - m) = \epsilon_t , t \in \mathbb{Z}.$$

Un processus de cette forme n'est pas toujours stationnaire. Autrement dit, tous les polynômes de degré p ne définissent pas un processus auto-régressif par cette écriture.

Proposition 3.8. Soit $\Phi: z \mapsto 1 - \phi_1 z - \dots - \phi_p z^p$ un polynôme de degré $p \geq 1$, $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc et m un réel. Le processus $(X_t)_{t\in\mathbb{Z}}$ vérifiant $\Phi(B)(X_t-m)=\epsilon_t, t\in\mathbb{Z}$ est un processus AR(p) (il est stationnaire) si et seulement si Φ n'a pas de racine de module 1 dans \mathbb{C} . De plus, $(\epsilon_t)_{t\in\mathbb{Z}}$ est le processus des innovations si et seulement si Φ n'a pas de racine dans la boule unité $\mathcal{B} = \{z \in \mathbb{C} : |z| \leq 1\}$.

Preuve. Admis pour le moment.

Exercice. Soit $\phi \in \mathbb{R} \setminus \{-1,1\}$. Exprimer le processus AR(1) $X_t = \phi X_{t-1} + \epsilon_t, t \in \mathbb{Z}$ sous la forme $X_t = \sum_{k \in \mathbb{Z}} \psi_k \epsilon_{t-k}$ avec $\sum_{k \in \mathbb{Z}} |\psi_k| < \infty$.

Exercice. Soit $\phi \in \{-1,1\}$ et $X_t = \phi X_{t-1} + \epsilon_t, t \in \mathbb{Z}$ avec $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc tel que ϵ_t est orthogonal à H_{t-1} pour tout t. Montrer que pour tout $h \ge 1$,

$$X_{t+h} = \phi^h X_t + \sum_{h=0}^{h-1} \phi^k \epsilon_{t+h-k}.$$

En déduire que la variance de X_t ne peut pas être constante au cours du temps et donc que le processus n'est pas stationnaire.

Les processus auto-régressifs sont caractérisés par une fonction d'auto-corrélation partielle nulle à partir d'un certain rang.

Proposition 3.9. Soit $p \ge 1$ et $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire d'espérance m dont la fonction l'autocorrélation partielle α vérifie $\alpha(p) \ne 0$ et $\alpha(h) = 0$ pour h > p. Alors, il existe des coefficients $\phi_1, ..., \phi_p$ tels que

$$X_t = m + \phi_1(X_{t-1} - m) + \dots + \phi_n(X_{t-n} - m) + \epsilon_t, \ t \in \mathbb{Z},$$

 $o\dot{u}$ $(\epsilon_t)_{t\in\mathbb{Z}}$ est le bruit blanc d'innovations de $(X_t)_{t\in\mathbb{Z}}$.

Preuve. On suppose m=0 sans perte de généralité. On sait $\alpha(h)=\beta_h^{(h)}$ où $\beta_h^{(h)}$ est le coefficient de X_{t-h} de la régression linéaire

$$X_t = \beta_1^{(h)} X_{t-1} + \dots + \beta_h^{(h)} X_{t-h} + \eta_{t,h} = \Pi_{H_{t-1}^h}(X_t) + \eta_t^{(h)}.$$

Comme $\alpha(h)=0$ pour h>p, on déduit que $\Pi_{H^{p+1}_{t-1}}(X_t)=\Pi_{H^p_{t-1}}(X_t)$ et par suite $\Pi_{H^h_{t-1}}(X_t)=\Pi_{H^p_{t-1}}(X_t)$ pour tout $h\geq p$. On a donc,

$$\Pi_{H_{t-1}}(X_t) = \lim_{h \to \infty} \Pi_{H_{t-1}^h}(X_t) = \Pi_{H_{t-1}^q}(X_t),$$

ce qui permet de conclure.

Théorème 3.10. Soit $\Phi(B)(X_t - m) = \epsilon_t, t \in \mathbb{Z}$ une représentation AR(p) d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ avec Φ un polynôme de degré $p \geq 1$ sans racine unitaire tel que $\Phi(0) = 1$ et $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc. Alors, ϵ_t est l'innovation de X_t pour tout $t \in \mathbb{Z}$ si et seulement si Φ n'a pas de racine dans la boule unité complexe \mathcal{B} . Dans ce cas, la représentation $X_t = m + \Phi(B)^{-1}\epsilon_t, t \in \mathbb{Z}$ est causale.

Preuve. On suppose m=0. La preuve consiste à construire la décomposition de Wold du processus. Soit R l'ensemble des racines de Φ , pour tout $t \in \mathbb{Z}$,

$$\Phi(B)X_t = \prod_{z \in R, |z| > 1} (I - z^{-1}B) \prod_{z \in R, |z| < 1} (I - z^{-1}B)X_t = \epsilon_t$$

avec la convention $\prod_{z \in \emptyset} (I - z^{-1}B) = I$. Soit $\xi_z = (I - z^{-1}B)(I - zB)^{-1}$, on peut écrire

$$\prod_{z \in R, |z| > 1} (I - z^{-1}B) \prod_{z \in R, |z| < 1} (I - zB) \prod_{z \in R, |z| < 1} \xi_z X_t := \widetilde{\Phi}(B) \prod_{z \in R, |z| < 1} \xi_z X_t = \epsilon_t.$$

On sait que $\xi_z^{-1}=\xi_{z^{-1}}$ préserve la propriété de blancheur, on a donc

$$\widetilde{\Phi}(B)X_t = \prod_{z \in R, |z| < 1} \xi_z^{-1} \epsilon_t := \widetilde{\epsilon}_t `, \ t \in \mathbb{Z},$$

où $\widetilde{\epsilon}_t$ est un bruit blanc. Par construction, $\widetilde{\Phi}$ a toutes ses racines de module strictement supérieur à 1, $\widetilde{\Phi}(B)^{-1}$ est donc causal. On retrouve donc la décomposition de Wold de $(X_t)_{t\in\mathbb{Z}}$:

$$X_t = \widetilde{\Phi}(B)^{-1}\widetilde{\epsilon}_t , t \in \mathbb{Z}$$

ce qui permet de conclure.

Si Φ a toutes ses racines de module strictement supérieur à 1, on dira que la représentation $\Phi(B)(X_t-m) = \epsilon_t, t \in \mathbb{Z}$ est canonique, auquel cas ϵ_t est l'innovation au temps t. Les équations de Yule-Walker lient les coefficients d'un processus AR(p) à sa fonction d'auto-covariance (ou auto-corrélation).

Proposition 3.11 (Equations de Yule-Walker). Soit un processus AR(p) de représentation canonique

$$X_t = m + \phi_1(X_{t-1} - m) + \dots + \phi_n(X_{t-n} - m) + \epsilon_t, \ t \in \mathbb{Z},$$

et de fonction d'auto-covariance γ . Alors,

$$\gamma(0) = \sigma^2 + \sum_{j=1}^p \phi_j \gamma(j)$$
 et $\gamma(h) = \sum_{j=1}^p \phi_j \gamma(h-j)$, $h \ge 1$.

Preuve. Comme la représentation est canonique, ϵ_t est l'innovation et est donc orthogonal au passé X_{t-1}, X_{t-2}, \dots Pour tout $h \geq 0$,

$$cov(X_t, X_{t-h}) = cov(\phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t, X_{t-h}) = \sum_{j=1}^p \phi_j \gamma(h-j) + cov(\epsilon_t, X_{t-h}).$$

On conclut en remarquant que $cov(\epsilon_t, X_{t-h}) = \sigma^2$ si h = 0 et 0 sinon.

En divisant par $\gamma(0)$ de part et d'autre de l'égalité, on remarque que les équations de Yule-Walker pour $h \geq 1$ restent valables pour la fonction d'auto-corrélation ρ , ce qui permet par exemple d'estimer les paramètres $\phi_1, ..., \phi_p$ à partir de l'ACF empirique. On peut représenter ces équations sous forme d'un système linéaire

$$\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}.$$

La variance σ^2 de l'innovation est obtenue par l'équation de Yule-Walker pour h = 0, les équations pour h > p ne sont pas utiles en pratique.

3.3 Processus ARMA

Définition 3.12. Un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ est un processus auto-regressif moyenne mobile d'ordre (p,q), où ARMA(p,q), s'il existe des polynômes Φ,Θ sans racine commune, de degrés p,q respectivement tels que $\Phi(0) = \Theta(0) = 1$, un réel m et un bruit blanc $(\epsilon_t)_{t\in\mathbb{Z}}$ tels que

$$\Phi(B)(X_t - m) = \Theta(B)\epsilon_t , \ t \in \mathbb{Z}.$$

Les processus MA(p) et AR(q) sont bien sûr des cas particuliers. L'existence d'un processus ARMA (donc stationnaire) associé aux polynômes Φ , Θ dépend uniquement des racines de Φ .

Proposition 3.13. Soient Φ, Θ des polynômes de degrés p,q respectivement tels que $\Phi(0) = \Theta(0) = 1$, m un réel et $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc. Si Φ n'a pas de racine unitaire, il existe un processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire tel que

$$\Phi(B)(X_t - m) = \Theta(B)\epsilon_t , t \in \mathbb{Z}.$$

Preuve. Si Φ n'a pas de racine unitaire, $\Phi(B)$ est inversible et

$$X_t = m + \Phi(B)^{-1}\Theta(B)\epsilon_t, \ t \in \mathbb{Z}.$$

Le filtre $\Phi(B)^{-1}\Theta(B)$ est linéaire, le processus est donc stationnaire.

Définition 3.14. Soient Φ, Θ des polynômes tels que $\Phi(0) = \Theta(0) = 1$, m un réel et $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc. La représentation $\Phi(B)(X_t - m) = \Theta(B)\epsilon_t, t \in \mathbb{Z}$ d'un processus ARMA est:

- causale s'il existe $(\psi_k)_{k\geq 0}$ sommable telle que $X_t=m+\sum_{k\geq 0}\psi_k\epsilon_{t-k}, t\in\mathbb{Z}$.
- inversible s'il existe $(\zeta_k)_{k\geq 0}$ sommable telle que $\epsilon_t = \sum_{k>0} \zeta_k(X_{t-k} m), t \in \mathbb{Z}$.
- canonique si elle est causale, inversible et que Φ et Θ n'ont pas de racine en commun.

La condition supplémentaire d'absence de racine commune à Φ et Θ permet simplement d'avoir une représentation minimale du processus. Toute racine non-unitaire z commune à Φ et Θ peut se simplifier en appliquant le filtre $(I - z^{-1}B)^{-1}$ de part et d'autre de l'égalité.

Proposition 3.15. La représentation $\Phi(B)(X_t - m) = \Theta(B)\epsilon_t, t \in \mathbb{Z}$ d'un processus ARMA est causale (resp. inversible) si Φ (resp. Θ) n'a pas de racine complexe de module inférieur ou égal à 1.

Preuve. Si Φ n'a pas de racine dans \mathcal{B} , alors $\Phi(B)^{-1}$ est un filtre linéaire causal, tout comme $\Phi(B)^{-1}\Theta(B)$. L'écriture

$$X_t - m = \Phi(B)^{-1}\Theta(B)\epsilon_t$$
, $t \in \mathbb{Z}$

est donc causale. La preuve pour l'inversibilité est identique.

L'intérêt de la représentation canonique d'un processus ARMA est qu'elle fait intervenir le bruit blanc des innovations du processus recentré.

Proposition 3.16. Si la représentation $\Phi(B)X_t = \Theta(B)\epsilon_t, t \in \mathbb{Z}$, avec $\Phi(0) = \Theta(0) = 1$, d'un processus ARMA centré est canonique, alors $(\epsilon_t)_{t \in \mathbb{Z}}$ est le processus des innovations de $(X_t)_{t \in \mathbb{Z}}$.

Preuve. La causalité nous donne que

$$X_t \in E_t = \overline{\text{vect}\{\epsilon_s, s \le t\}}, \ \forall t \in \mathbb{Z},$$

et donc que $H_t \subseteq E_t$. De même, l'inversibilité entraı̂ne $\epsilon_t \in H_t$ pour tout $t \in \mathbb{Z}$. Comme $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc, ϵ_t est orthogonal à E_{t-1} et donc également à H_{t-1} . On a donc bien

$$X_t = \prod_{H_{t-1}} (X_t) + \psi_0 \epsilon_t , \ t \in \mathbb{Z},$$

avec $\psi_0 = \Theta(0)/\Phi(0) = 1$.

La réciproque est vraie à l'exception du cas où Θ a une ou plusieurs racines unitaires (de module 1). Dans ce cas, $(\epsilon_t)_{t\in\mathbb{Z}}$ peut être le bruit blanc des innovations sans que la représentation soit inversible. C'est le cas par exemple du processus MA(1) $X_t = \epsilon_t + \epsilon_{t-1}, t \in \mathbb{Z}$ qui n'est pas inversible mais qui a bien $(\epsilon_t)_{t\in\mathbb{Z}}$ comme bruit blanc des innovations.

Proposition 3.17. Soient Φ, Θ des polynômes tels que $\Phi(0) = \Theta(0) = 1$, m un réel et $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc. Si Φ et Θ n'ont pas de racine unitaire, alors tout processus ARMA tel que $\Phi(B)(X_t - m) = \Theta(B)\epsilon_t$, $t \in \mathbb{Z}$ admet une représentation canonique.

Preuve. Quitte à simplifier de part et d'autre de l'égalité, on peut supposer que Φ et Θ n'ont pas de racine commune. Soit $z_1,...,z_p$ les racines complexes de Φ (comptées avec multiplicité), on construit le polynôme $\widetilde{\Phi}$ tel que $\widetilde{\Phi}(0)=1$ et ayant pour racines z_j si $|z_j|>1$ et $1/z_j$ si $|z_j|<1$ pour j=1,...,p. On construit $\widetilde{\Theta}$ de la même façon. Soit

$$\widetilde{\epsilon}_t = \widetilde{\Phi}(B)\Phi(B)^{-1}\widetilde{\Theta}(B)\Theta(B)^{-1}\epsilon_t, \ t \in \mathbb{Z}.$$

Soit $\tau_1, ..., \tau_r$ les racines de Φ de module inférieur à 1 suivies de l'inverse des racines de Θ de module supérieur à 1 (comptées avec multiplicité), on peut écrire

$$\widetilde{\epsilon}_t = \xi_{\tau_1} \dots \xi_{\tau_r} \epsilon_t,$$

où $\xi_z = (I - z^{-1}B)(I - zB)^{-1}, z \in \mathbb{C}, |z| \neq 1$. Comme ξ_z préserve la blancheur, le processus $(\widetilde{\epsilon}_t)_{t \in \mathbb{Z}}$ est un bruit blanc. On vérifie alors

$$\widetilde{\Phi}(B)(X_t - m) = \widetilde{\Theta}(B)\widetilde{\epsilon}_t , \ t \in \mathbb{Z}.$$

Comme $\widetilde{\Phi}$ et $\widetilde{\Theta}$ ont toutes leurs racines en dehors de la boule unité complexe par construction, cette représentation est canonique.

Exercice. Calculer la variance de l'innovation d'un ARMA $\Phi(B)X_t = \Theta(B)\epsilon_t, t \in \mathbb{Z}$ où Φ, Θ n'ont pas de racine unitaire, en fonction des racines de Φ, Θ et de $\sigma^2 = \text{var}(\epsilon_t)$ (indication: en utilisant les notations de la preuve précédente, montrer que $\text{var}(\tilde{\epsilon}_t) = \sigma^2 \text{RV}(\xi_{\tau_1}) \dots \text{RV}(\xi_{\tau_r})$).

Contrairement aux AR et MA, les ARMA n'ont pas de caractérisations immédiates par leurs ACF ou PACF. Il existe une méthode pour déterminer les ordres p,q d'un processus ARMA basé sur un calcul de déterminants.

Proposition 3.18 (Méthode du coin). Soit $(X_t)_{t\in\mathbb{Z}}$ un processu stationnaire de fonction d'auto-covariance γ . Pour $i, j \geq 1$, on définit la matrice

$$R_{ij} = \begin{bmatrix} \gamma(i) & \gamma(i-1) & \dots & \gamma(i-j+1) \\ \gamma(i+1) & \gamma(i) & \dots & \gamma(i-j+2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(i+j-1) & \gamma(i+j-2) & \dots & \gamma(i) \end{bmatrix}.$$

 $Si(X_t)_{t\in\mathbb{Z}}$ est un ARMA(p,q), alors $det(R_{pq}) \neq 0$ et $det(R_{ij}) = 0$ pour tout i > p, j > q.

Preuve (partielle). Soit un processus ARMA(p,q) centré avec

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \ t \in \mathbb{Z},$$

où $(\epsilon_t)_{t\in\mathbb{Z}}$ est le processus des innovations. Pour i>q, on trouve

$$cov(X_t, X_{t-i}) = \gamma(i) = \phi_1 \gamma(i-1) + ... \phi_p \gamma(i-p).$$

Pour j > p, le vecteur $(1, -\phi_1, ..., -\phi_p, 0, ..., 0) \in \mathbb{R}^j$ appartient donc au noyau de R_{ij} , celle-ci n'est donc pas inversible. On admettra que det (R_{pq}) est non nul.

Ce résultat fournit une méthode pour estimer les ordres d'un processus ARMA: on construit des estimateurs \hat{R}_{ij} à partir de l'auto-covariance empirique du processus et on identifie les ordres p,q comme les seuils au-delà desquels les déterminants sont "proches" de zéro. Si on représente la matrice

$$\begin{bmatrix} \det(\widehat{R}_{11}) & \det(\widehat{R}_{12}) & \det(\widehat{R}_{13}) & \dots \\ \det(\widehat{R}_{21}) & \det(\widehat{R}_{22}) & \det(\widehat{R}_{23}) & \dots \\ \det(\widehat{R}_{31}) & \det(\widehat{R}_{32}) & \det(\widehat{R}_{33}) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

un "coin" de valeurs proches de zéro devrait apparaître en bas à droite (d'où le nom de méthode du coin).

4 Extensions des processus ARMA

4.1 Processus ARMA saisonniers

Les processus ARMA saisonniers, ou SARMA, permettent de modéliser des phénomènes périodiques par des processus stationnaires sans utiliser de décomposition avec tendance et composante saisonnière déterministes.

Définition 4.1. Un processus auto-régressif moyenne mobile saisonnier d'ordre $(p,q) \times (P,Q)$ de période $s \geq 2$, noté $SARMA_s(p,q,P,Q)$, est un processus stationnaire tel que

$$\Phi_s(B^s)\Phi(B)(X_t-m) = \Theta_s(B^s)\Theta(B)\epsilon_t, \ t \in \mathbb{Z},$$

où $m \in \mathbb{R}$, Φ , Φ _s, Θ , Θ _s sont des polynômes de degrés respectifs p, P, q, Q qui valent 1 en 0 et tels que Φ _s Φ et Θ _s Θ n'ont pas de racine en commun.

Les processus SARMA sont des cas particuliers de processus ARMA pour lesquels les polynômes qui les caractérisent ont une forme particulière (un SARMA_s(p,q,P,Q)) est un cas carticulier d'ARMA(sP+p,sQ+q)). Le but est de simplifier le modèle pour minimiser le nombre de paramètres, tout en permettant de modéliser des dépendances périodiques, à travers les polynômes Φ_s et Θ_s . En dehors de la moyenne m et la variance σ^2 de l'innovation, un SARMA_s(p,q,P,Q) comporte p+P+q+Q coefficients à estimer alors qu'un ARMA(sP+p,sQ+q) en comporte s(P+Q)+p+q.

Exercice. Montrer que la représentation $\Phi_s(B^s)\Phi(B)(X_t-m) = \Theta_s(B^s)\Theta(B)\epsilon_t, t \in \mathbb{Z}$ d'un processus SARMA_s(p,q,P,Q) est canonique si et seulement si Φ_s, Φ, Θ_s et Θ n'ont pas de racine de module strictement inférieur à 1.

4.2 ARIMA et SARIMA

Définition 4.2. Un processus $(X_t)_{t\in\mathbb{Z}}$ est un processus auto-régressif moyenne mobile intégré d'ordre (p,d,q), noté ARIMA(p,d,q) si $(I-B)^dX_t, t\in\mathbb{Z}$ est un processus ARMA(p,q). Un tel processus admet une représentation de la forme

$$(I-B)^d \Phi(B)(X_t-m) = \Theta(B)\epsilon_t, \ t \in \mathbb{Z},$$

où $m \in \mathbb{R}$, $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc et Φ , Θ sont les polynômes de degrés p, q respectivement sans racine commune (et Φ n'a pas de racine unitaire).

Un ARIMA(p,d,q) avec $d \ge 1$ n'est pas un processus stationnaire puisque le polynôme $z \mapsto (I-z)^d \Phi(z)$ a z=1 comme racine. Les ARIMA servent à modéliser des séries non-stationnaires. La même généralisation s'applique aux processus SARMA.

Définition 4.3. Un processus auto-régressif moyenne mobile intégré saisonnier d'ordre $(p, d, q) \times (P, D, Q)$ de période s, noté $SARIMA_s(p, d, q, P, D, Q)$ est un processus ayant pour représentation

$$(I - B^s)^D (I - B)^d \Phi_s(B^s) \Phi(B)(X_t - m) = \Theta(B)\epsilon_t , \ t \in \mathbb{Z},$$

où $m \in \mathbb{R}$, $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc et Φ , Φ_s , Θ , Θ_s sont les polynômes de degrés p, P, q, Q tels que $\Phi_s\Phi$ et $\Theta_s\Theta$ n'ont pas de racine en commun et Φ_s , Φ n'ont pas de racine unitaires.

Pour la même raison que précédemment un SARIMA_s (p,d,q,P,D,Q) n'est pas stationnaire dès que $d \ge 1$ ou $D \ge 1$. Etant donné le nombre potentiellement important de paramètres à estimer, on cherchera le plus souvent une modélisation la plus simple possible, typiquement $d,D \le 1$ et $p,P,q,Q \le 2$.

Exercice. Soit le processus

$$X_t = M_t + S_t + Z_t$$
, $t \in \mathbb{Z}$

où M_t est un polynôme de degré d en t, S_t une série périodique de période s et Z_t un ARMA(p,q). Montrer que $(X_t)_{t\in\mathbb{Z}}$ est un processus SARIMA, en précisant son ordre et sa période.

La modélisation SARIMA permet d'intégrer d'éventuelles dépendances périodiques dans la composante stochastique Z_t par les polynômes Φ_s , Θ_s . On observe souvent en pratique des phénomènes de dépendances de période s même après avoir enlever les tendances et composantes saisonnières.

4.3 Modélisation avec variables exogènes

Les modèles ARMAX (ARMA avec variables exogènes), parfois appelés REGARMA, sont des modèles de régression linéaire avec erreurs corrélées qui suivent un ARMA. Ces modèles sont utiles lorsque les résidus issus d'une régression par moindres carrés sont (trop) corrélés. La méthode s'apparente à une régression linéaire généralisée où la matrice de covariance est paramétrique (les paramètres sont les coefficients du processus ARMA). Dans le cadre Gaussien, les paramètres sont estimés par maximum de vraisemblance en même temps que les coefficients de la régression.

Proposition 4.4. Soit un modèle de régression

$$y = X\beta + u$$

où $X \in \mathbb{R}^{n \times k}$, $\beta \in \mathbb{R}^k$ et $u = (u_1, ..., u_n)$ est la réalisation d'un processus ARMA(p,q) de représentation canonique $\Psi(B)u_t = \Theta(B)\epsilon_t, t \in \mathbb{Z}$ avec $(\epsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc Gaussien de variance σ^2 . On note $\Omega = \Omega(\Psi, \Theta, \sigma^2)$ l'inverse de la matrice de covariance (la matrice de précision) de u, l'estimateur sans biais optimal de β est donné par

$$\widehat{\beta} = (\mathbf{X}^{\top} \Omega \mathbf{X})^{-1} \mathbf{X}^{\top} \Omega y.$$

Preuve. Admis (régression linéaire généralisée).

Exercice. Calculer la matrice de précision Ω d'un processus AR(1),

$$u_0 = 0$$
, $u_t = \phi u_{t-1} + \epsilon_t$, $t = 1, ..., n$,

avec $|\phi| < 1$.

4.4 Modèles GARCH

Des séries temporelles hétéroscédastiques conditionnellement au passé ont été introduites pour modéliser l'évolution de cours de la bourse en finance. Dans ces modèles, on s'intéresse essentiellement à la volatilité σ_t , qui désigne la variance au temps t conditionnellement au passé.

Définition 4.5. Un processus $(X_t)_{t\in\mathbb{Z}}$ est auto-régressif conditionnellement hétéroscédastique d'ordre p, noté ARCH(p), s'il admet une représentation de la forme

$$\left\{ \begin{array}{l} X_t = \sigma_t \epsilon_t \\ \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 \end{array} \right., \ t \in \mathbb{Z},$$

où $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc de variance 1 indépendant du passé X_{t-1}, X_{t-2}, \dots et $\alpha_0, \dots, \alpha_p$ sont des coefficients positifs.

Le nom provient du fait que la variance de X_t conditionnellement au passé X_{t-1}, X_{t-2}, \dots dépend du temps, alors que la variance non conditionnelle var (X_t) est constante lorsqu'elle existe.

Exercice. Calculer le processus des innovations de $(X_t^2)_{t\in\mathbb{Z}}$ où $(X_t)_{t\in\mathbb{Z}}$ suit un ARCH(p). Sous quelles conditions $(X_t^2)_{t\in\mathbb{Z}}$ est-il stationnaire?

Définition 4.6. Un processus $(X_t)_{t\in\mathbb{Z}}$ est auto-régressif conditionnellement hétéroscédastique généralisé d'ordre (p,q), noté GARCH(p,q), s'il admet une représentation de la forme

$$\left\{ \begin{array}{l} X_t = \sigma_t \epsilon_t \\ \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2 \end{array} \right., \ t \in \mathbb{Z},$$

où $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc de variance 1 non-corrélé au passé X_{t-1}, X_{t-2}, \ldots et $\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ sont des coefficients positifs.

S'il est stationnaire, un processus GARCH(p,q) est un bruit blanc.

Proposition 4.7. Un processus GARCH(p,q) stationnaire $(X_t)_{t\in\mathbb{Z}}$ est centré, d'auto-covariance $\gamma(h)=0$ pour $h\geq 1$ et de variance

$$\gamma(0) = \frac{\alpha_0}{1 - \alpha_1 - \dots - \alpha_p - \beta_1 - \dots - \beta_q}.$$

Preuve. Comme ϵ_t est indépendant du passé $X_{t-1}, X_{t-2}, ..., \mathbb{E}(X_t) = \mathbb{E}(\sigma_t)\mathbb{E}(\epsilon_t) = 0$ pour tout $t \in \mathbb{Z}$. De même, pour tout h > 1,

$$\gamma(h) = \mathbb{E}(X_t X_{t-h}) = \mathbb{E}(\sigma_t X_{t-h} \epsilon_t) = \mathbb{E}(\sigma_t X_{t-h}) \mathbb{E}(\epsilon_t) = 0.$$

En notant $\eta_t = X_t^2 - \sigma_t^2 = \sigma_t^2 (\epsilon_t^2 - 1),$

$$X_{t}^{2} = \eta_{t} + \sigma_{t}^{2} = \alpha_{0} + \sum_{i=1}^{p} \alpha_{i} X_{t-i}^{2} + \sum_{j=1}^{q} \beta_{j} \sigma_{t-j}^{2} + \eta_{t}$$

$$= \alpha_{0} + \sum_{i=1}^{p} \alpha_{i} X_{t-i}^{2} + \sum_{j=1}^{q} \beta_{j} X_{t-j}^{2} - \sum_{j=1}^{q} \beta_{j} \eta_{t-j} + \eta_{t}$$

Soit $m = \mathbb{E}(X_t^2)$ qui est constante par stationnarité, on a donc par passage à l'espérance

$$m = \alpha_0 + \left(\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j\right) m,$$

ce qui permet de conclure.

Exercice. Montrer qu'un processus GARCH(p,q) $(X_t)_{t\in\mathbb{Z}}$ stationnaire avec $(\epsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc Gaussien est tel que $(X_t^2)_{t\in\mathbb{Z}}$ est un processus ARMA.

5 Inférence

D'un point de vue théorique, l'inférence dans des modèles de séries temporelles est plus technique que dans les modèles d'échantillonage iid. La consistance des estimateurs classiques (estimateurs empiriques pour la moyenne, la variance et la fonction d'auto-covariance et estimateurs du maximum de vraisemblance Gaussienne pour les coefficients d'un ARMA par exemple) dépend de la vitesse de décroissance de la fonction d'auto-covariance. Pour les ARMA, la fonction d'auto-covariance décroît exponentiellement vite et on peut montrer que ces estimateurs sont consistants.

Certains tests traitent de la structure de dépendance, comme les tests de blancheur (tester si un processus est un bruit blanc) et des tests de stationnarité. Il est également possible de tester la nullité de l'auto-covariance empirique en un lag donné (pour déterminer l'ordre d'un MA par exemple) ou l'auto-corrélation partielle (pour déterminer l'ordre d'un AR). On peut aussi construire des tests de significativité sur les coefficients d'un SARMA.

5.1 Estimation des paramètres

On suppose qu'on observe la réalisation d'une trajectoire $X = (X_1, ..., X_n)^{\top}$ issue d'un processus linéaire Gaussien de représentation canonique

$$X_t = m + \sum_{k \ge 0} \psi_k \epsilon_{t-k} \ , \ t \in \mathbb{Z}$$

où $m \in \mathbb{R}$, $(\psi_k)_{k \geq 0}$ est une suite réelle sommable et $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc Gaussien de variance σ^2 . Comme d'habitude, on note γ sa fonction d'auto-covariance, définie sur \mathbb{Z} .

Proposition 5.1. Si $\sum_{k\in\mathbb{Z}}\psi_k\neq 0$ et $\sum_{h\geq 0}h\gamma(h)<\infty$, l'estimateur $\widehat{m}=n^{-1}\sum_{i=1}^nX_i$ vérifie

$$\sqrt{n}(\widehat{m}-m) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0,\tau^2),$$

$$o\dot{u} \ \tau^2 = \sigma^2 \big(\sum_{k \ge 0} \psi_k \big)^2.$$

Preuve. Puisqu'on suppose la normalité du processus, il suffit de remarquer que

$$\widehat{m} \sim \mathcal{N}\left(0, \frac{1}{n} \sum_{h=0}^{n-1} (n-h)\gamma(h)\right)$$

et le reste de la preuve est simple. Le résultat reste valable dans pour un processus non Gaussien mais la preuve est admise dans ce cas. \Box

L'auto-covariance empirique à l'ordre h, dont on rappelle l'expression

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{i=h+1}^{n} (X_i - \widehat{m})(X_{i-h} - \widehat{m}) , \ 0 \le h \le n$$

est une transformation quadratique du vecteur X. On peut donc déterminer ses moments dans le cas Gaussien.

Exercice. Déterminer la matrice S_h telle que $\widehat{\gamma}(h) = n^{-1}X^{\top}S_hX$. Calculer le biais de l'estimateur $\widehat{\gamma}(h)$.

Un long calcul (que l'on s'épargnera) permet de calculer la variance théorique de $\widehat{\gamma}(h)$ en fonction de $m, \gamma(0), ..., \gamma(n-h)$. On peut montrer également que $\widehat{\gamma}$ est asymptotiquement Gaussien, ce qui permet au final de construire des intervalles de confiance asymptotiques. Si le processus n'est pas Gaussiens, il est aussi possible de construire des intervalles de confiances sur $\gamma(h)$ en estimant les moments jusqu'à l'ordre 4, sous des conditions plus fortes de stationnarité et d'intégrabilité du processus.

Estimer entièrement la loi d'un processus linéaire est impossible dans le cas général du fait qu'il y a un nombre potentiellement inifini de paramètres à estimer: $\sigma^2, m, \psi_0, \psi_1, \ldots$ On peut en revanche s'en sortir si une suppose une structure ARMA et donc qu'on se place dans un vrai modèle paramétrique. On suppose donc maintenant que $(X_t)_{t\in\mathbb{Z}}$ est un processus ARMA(p,q) de représentation canonique

$$\Phi(B)(X_t - m) = \Theta(B)\epsilon_t, \ t \in \mathbb{Z},$$

où $m \in \mathbb{R}$, $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc Gaussien de variance σ^2 . On note $\phi = (\phi_1, ..., \phi_p)$ et $\theta = (\theta_1, ..., \theta_q)$ les coefficients des polynômes Φ et Θ .

Exercice. Montrer que $\sqrt{n}(\widehat{m}-m)$ converge en loi vers une loi normale non dégénérée dans le modèle ARMA Gaussien qui admet une représentation canonique. Exprimer la variance asymptotique τ^2 en

fonction de σ^2 , Φ , Θ .

Comme précédemment, X est un vecteur Gaussien d'espérance $\mathsf{m}=(m,...,m)^{\top}\in\mathbb{R}^n$ et de matrice de covariance

$$\Sigma = \Sigma_n(\phi, \theta, \sigma^2) = (\gamma(i-j))_{i,j=1,\dots,n}.$$

Une forme explicite de Σ en fonction des paramètres s'obtient à partir de la représentation en $\mathrm{MA}(\infty)$

$$X_t = m + \Phi(B)^{-1}\Theta(B)\epsilon_t = m + \sum_{k>0} \psi_k \epsilon_{t-k} , t \in \mathbb{Z},$$

par la formule $\gamma(h) = \gamma(-h) = \sigma^2 \sum_{k > h} \psi_k \psi_{k-h}, h \ge 0$. La vraisemblance du modèle est donnée par

$$V(m, \phi, \theta, \sigma^2, X) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\Big(-\frac{1}{2}(X - \mathsf{m})^\top \Sigma^{-1}(X - \mathsf{m})\Big).$$

L'expression compliquée de Σ en fonction des paramètres ϕ, θ, σ^2 à laquelle se rajoute les conditions sur Φ, Θ pour que l'écriture soit canonique, rend difficile le calcul des estimateurs du maximum de vraisemblance (il n'existe pas de forme explicite simple dans le cas général). On a donc recours en général à des méthodes numériques pour les calculer.

On retiendra que les estimateurs des paramètres d'un modèle ARMA sont asymptotiquement Gaussiens et que l'on peut estimer la variance asymptotique. Cela suffit à construire des intervalles de confiances de niveau de confiance asymptotique donné, et donc de tester leurs significativités. Déterminer les ordres p, q relève de la sélection de modèle. Il existe des critères pour comparer les différents modèles comme le critère AIC, qui dans un modèle ARMA(p,q) Gaussien vaut

$$\mathrm{AIC} = 2(p+q+2) - 2\ln\left(V(m,\phi,\theta,\sigma^2,X)\right) = (X-\mathsf{m})^{\top}\Sigma^{-1}(X-\mathsf{m}) - \ln\left(\det(\Sigma)\right) + 2(p+q) + cste.$$

Pour comparer des modèles emboîtés, le critère AIC revient à tester la significativité des paramètres en plus du modèle le plus grand. Pour ajuster un modèle SARMA, l'approche est la même (un processus SARMA est un cas particulier de processus ARMA) avec une structure particulière imposée sur les polynômes Φ, Θ . Enfin, les modèles SARIMA non-stationnaires n'ont pas de paramètres supplémentaires à estimer, seulement les ordres d, D des différenciations. Ces ordres sont déterminés à l'aide de tests de stationnarité. Si les résultats des tests concluent à la non-stationnarité du modèle, on peut appliquer des filtres de différenciation I-B ou $I-B^s$ pour tenter de rendre le processus stationnaire.

5.2 Tests de stationnarité

Il existe plusieurs tests de stationnarité sur des séries temporelles qui ne considèrent pas les mêmes hypothèses nulle et alternative.

Le test de Dickey-Fuller est basé sur le coefficient ϕ dans l'ajustement par un AR(1) $X_t = a + \phi X_{t-1} + \epsilon_t, t \in \mathbb{Z}$. On peut le voir comme un test de Student de l'hypothèse $H_0: \phi = 1$ contre $H_1: \phi < 1$ dans la régression (avec corrélations) de $(X_2, ..., X_n)$ par $(X_1, ..., X_{n-1})$. Si la série montre une tendance affine

en temps, une version différente consiste à tester la nullité simultanée des coefficients b, ϕ associés à X_{t-1} et t dans la régression linéaire

$$X_t = a + bt + \phi X_{t-1} + \epsilon_t, \ t \in \mathbb{Z}.$$

Le test de Dickey-Fuller augmenté (ADF) est basé sur l'écriture AR(p) du processus recentré $X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t, t \in \mathbb{Z}$. En posant $Y_t = \Delta X_t = X_t - X_{t-1}$, le jeu d'écriture

$$Y_{t} = (\phi_{1} - 1)X_{t-1} + \dots + \phi_{p}X_{t-p} + \epsilon_{t}$$

$$= (\phi_{1} - 1)X_{t-1} + \dots + (\phi_{p-1} + \phi_{p})X_{t-p+1} + \phi_{p}Y_{t-p} + \epsilon_{t}$$

$$= (\phi_{1} - 1)X_{t-1} + \dots + (\phi_{p-2} + \phi_{p-1} + \phi_{p})X_{t-p+2} + (\phi_{p-1} + \phi_{p})Y_{t-p+1} + \phi_{p}Y_{t-p} + \epsilon_{t}$$

$$\vdots$$

$$= (\phi_{1} + \dots + \phi_{p} - 1)X_{t-1} + (\phi_{2} + \dots + \phi_{p})Y_{t-2} + (\phi_{3} + \dots + \phi_{p})Y_{t-3} + \dots + \phi_{p}Y_{t-p} + \epsilon_{t}$$

montre que $-\Phi(1) = \phi_1 + ... + \phi_p - 1$ est le coefficient associé à X_{t-1} de la régression linéaire de ΔX_t par $X_{t-1}, \Delta X_{t-1}, ..., \Delta X_{t-p+1}$. Le test de Dickey-Fuller augmenté est un test de Student sur la la nullité de ce coefficient, on teste donc en réalité si le polynôme Φ a pour racine 1, une condition suffisante mais pas nécessaire pour la non-stationnarité. Il existe aussi une version avec tendance affine du temps. Les tests de Dickey-Fuller et Dickey-Fuller augmentés sont en quelque sorte des tests de non-stationnarité: l'hypothèse nulle implique la non-stationnarité du processus.

Le test de Kwiatkowski, Phillips, Schmidt et Shin (KPSS) est un test de stationnarité, cette fois. Il cherche à modéliser le processus observé comme la somme d'une marche aléatoire et d'un processus stationnaire

$$X_t = \xi_t + \eta_t$$
 où $\xi_t = \xi_{t-1} + \epsilon_t$, $t \in \mathbb{Z}$

avec $(\epsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance σ^2 et $(\eta_t)_{t\in\mathbb{Z}}$ un processus stationnaire. Le test KPSS s'écrit alors comme un test sur la variance σ^2 avec pour hypothèse nulle $H_0: \sigma^2 = 0$, contre l'alternative $H_1: \sigma^2 > 0$. Le processus n'est stationnaire que si H_0 est vraie, une marche aléatoire n'étant pas stationnaire.

Les tests ADF et KPSS sont complémentaires dans le sens où ils testent des hypothèses contraires. Pour tester la stationnarité, on combine en général les résultats des deux tests.

5.3 Tests de blancheur

Les tests de blancheurs consistent à tester si un processus est un bruit blanc. Ils sont utilisés sur les résidus (d'une régression ou d'une modélisation SARIMA par exemple) pour valider le modèle. Grossièrement, on peut donc tester si un modèle est un SARIMA avec un ordre (p,d,q,P,D,Q) et une période s donnés en testant si les résidus issus de la modélisation sont un bruit blanc, en faisant attention à d'éventuels problèmes de sur-ajustement.

Il existe plusieurs tests de blancheur:

• Le test de Durbin-Watson est construit à partir de la statistique

$$DW := \frac{\sum_{i=2}^{n} (X_i - X_{i-1})^2}{\sum_{i=1}^{n} X_i^2}.$$

Il permet de détecter une corrélation non-nulle à l'ordre 1 seulement, soit l'hypothèse nulle H_0 : $\gamma(1) = 0$. celle-ci est rejetée si DW est suffisamment éloigné de 2, sa valeur théorique moyenne si les X_i sont iid.

• Le test de Box-Pierce utilise la statistique

$$BP_h := n \sum_{i=1}^h \widehat{\rho}(i)^2$$

où h est un lag donné, choisi par le statisticien. Il permet de tester l'hypothèse de non-corrélation jusqu'à l'ordre h, soit $H_0: \gamma(1) = \dots = \gamma(h) = 0$. Si les X_i sont indépendants, BP_h converge en loi vers un χ^2 à h degrés de liberté. Le test est donc asymptotique et sa région de rejet est calculée à partir du quantile de du $\chi^2(h)$.

• Le test de Ljung-Box est une correction du test de Box-Pierce qui est meilleur en terme de niveau et puissance que le test de Box-Pierce, sur toutes les tailles d'échantillons (il faut donc l'utiliser de préférence). La statistique de test utilisée est

$$LB_h := n(n-2) \sum_{i=1}^h \frac{\widehat{\rho}(i)}{n-i}.$$

Comme pour le test de Box-Pierce, le test est asymptotique et la statistique est comparée au quantile du $\chi^2(h)$.

On remarque que la validité théorique de ces tests sous l'hypothèse nulle nécessitent que les X_i soient iid et pas seulement non-corrélés.

Dans les modèles ARMAX, le test de Breusch-Godfrey permet de tester la blancheur des résidus du modèle de régression (ou plus exactement la non-corrélation jusqu'à l'ordre h). Soit un modèle de régression

$$y_i = \beta_1 x_{1,i} + ... + \beta_k x_{k,i} + u_i, i = 1, ..., n$$

où $u = (u_1, ..., u_n)$ est le vecteur des bruits. On note $\hat{u} = (\hat{u}_1, ..., \hat{u}_n)$ le vecteur des résidus de la régression par moindres carrés ordinaires dans ce modèle. Le test de Breusch-Godfrey est construit à partir du R^2 du modèle de régression des résidus \hat{u}_i sur les $x_{i,i}$ et les résidus aux temps précédents

$$\widehat{u}_i = \mu_1 x_{1,i} + \ldots + \mu_k x_{k,i} + \nu_1 \widehat{u}_{i-1} + \ldots + \nu_h \widehat{u}_{i-h} + \epsilon_i \ , \ i = h+1, \ldots, n.$$

Soit γ_u la fonction d'auto-corrélation du processus $(u_t)_{t\in\mathbb{Z}}$. Sous $H_0: \gamma_u(1) = \dots = \gamma_u(h) = 0$, on a

$$(n-h)R^2 \xrightarrow[n\to\infty]{loi} \chi^2(h).$$

5.4 Ajustement d'un modèle SARIMA

Pour résumer, l'ajustement en pratique d'un modèle sur une série observée se fait de la manière suivante:

1. Représenter la série graphiquement pour détecter des tendances et phénomènes saisonniers. En présence d'une composante saisonnière, le plus simple est de déterminer la période s par le contexte ou à défaut, à partir de la représentation graphique.

- 2. Tester la stationnarité en effectuant les tests de Dickey-Fuller (ou Dickey-Fuller augmenté) et KPSS.
- 3. Si on retient l'hypothèse de non-stationnarité, appliquer des filtres de la forme I-B ou $I-B^s$ pour tenter de rendre la série stationnaire. On peut aussi appliquer un transformation (point par point) de type $Y_t = f(X_t)$ pour rendre la série stationnaire si par exemple elle montre une variance non constante.
- 4. Une fois la série stationnarisée, représenter les ACF et PACF empiriques pour détecter des comportements particulier sur les covariances. Privilégier une modélisation AR(p) si $\widehat{\rho}(h)$ est proche de zéro pour h > p ou MA(q) si $\widehat{\alpha}(h)$ est proche de zéro pour h > q. Si aucun de ces phénomènes n'est visible, mais que les corrélations et corrélations partielles décroissent rapidement, ajuster un modèle ARMA(p,q). Les ordres p,q peuvent être estimés par la méthode du coin.
- 5. Une fois le modèle posé, estimer les coefficients (par exemple par maximum de vraisemblance sous la normalité du processus). Vérifier la significativité des coefficients estimés à l'aide des intervalles de confiance.
- 6. Tester la blancheur des résidus par un ou plusieurs tests de types Ljung-Box, Durbin-Watson, Breush-Godfrey, etc...
- 7. Si plusieurs modèles sont retenus, on peut les comparer par un critère de vraisemblance penalisée de type AIC ou BIC.

6 Prévision

On s'intéresse maintenant au problème de prédire les valeurs futures de la séries à partir des données passées. Le meilleur prédicteur de X_{t+h} en fonction de $X_t, X_{t-1}, ...$ (sous entendu celui qui minimise l'erreur quadratique) est une certaine fonction $f_h(X_t, X_{t-1}, ...) = \mathbb{E}(X_{t+h}|X_t, X_{t-1}, ...)$ du passé au temps t. Rechercher le meilleur prédicteur parmi toutes les fonctions mesurables possibles est impossible en pratique, on se restreindra donc à des prédicteurs linéaires.

6.1 Meilleur prédicteur linéaire

Par définition, le meilleur prédicteur linéaire de X_{t+h} en fonction du passé au temps t, X_t, X_{t-1}, \dots est donné par

$$\widehat{X}_{t+h} = \Pi_{H_t}(X_{t+h}).$$

Dans le cas où le processus $(X_t)_{t\in\mathbb{Z}}$ est Gaussien, \widehat{X}_{t+h} est l'espérance conditionnelle $\mathbb{E}(X_{t+h}|X_t,X_{t-1},...)$, il n'y a donc pas de perte à se restreindre à une prévision linéaire.

Proposition 6.1. Soit un processus linaire de représentation causale et inversible

$$X_t = m + \sum_{k \ge 0} \psi_k \epsilon_{t-k} = m + \sum_{k \ge 1} \zeta_k (X_{t-k} - m) + \epsilon_t , \ t \in \mathbb{Z},$$

où $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc de variance σ^2 . Le meilleur prédicteur linéaire de X_{t+h} est donné par $\widehat{X}_{t+1} = m + \sum_{k\geq 1} \zeta_k(X_{t+1-k} - m)$ si h = 1 et

$$\widehat{X}_{t+h} = m + \sum_{h=1}^{h-1} \zeta_k (\widehat{X}_{t+h-k} - m) + \sum_{h=h}^{\infty} \zeta_k (X_{t+h-k} - m) , h \ge 2.$$

L'erreur quadratique de prédiction vaut $\tau_h^2 := \mathbb{E} \left(X_{t+h} - \widehat{X}_{t+h} \right)^2 = \sigma^2 \sum_{k=0}^{h-1} \psi_k^2$.

Preuve. Par linéarité,

$$\widehat{X}_{t+h} = \Pi_{H_t}(X_{t+h}) = m + \sum_{k \ge 1} \zeta_k \Pi_{H_t}(X_{t+h-k} - m),$$

où $\Pi_{H_t}(X_{t+h-k}-m)=X_{t+h-k}-m$ si $h\geq k$. En utilisant que $\epsilon_t,\epsilon_{t-1},\ldots\in H_t$ et $\epsilon_{t+1},\epsilon_{t+2},\ldots\in H_t^\perp,$

$$X_{t+h} - \widehat{X}_{t+h} = m + \sum_{k>0} \psi_k \epsilon_{t+h-k} - m - \sum_{k>0} \psi_k \Pi_{H_t}(\epsilon_{t+h-k}) = \sum_{k=0}^{h-1} \psi_k \epsilon_{t+h-k},$$

ce qui permet de conclure.

On voit clairement que l'erreur quadratique de prédiction τ_h^2 est croissante en h, ce qui traduit qu'il est plus difficile de prédire à un horizon h plus lointain. Si le processus $(X_t)_{t\in\mathbb{Z}}$ est Gaussien, les prédicteurs \widehat{X}_{t+h} le sont également par linéarité. On peut donc déduire un intervalle de prédiction de niveau de confiance $1-\alpha$:

$$\left[\widehat{X}_{t+h} - q_{1-\frac{\alpha}{2}}\tau_h, \widehat{X}_{t+h} + q_{1-\frac{\alpha}{2}}\tau_h\right]$$

où q est le quantile de la loi normale centré réduite.

Le meilleur prédicteur linéaire X_{t+h} n'est pas connu en pratique, mais les coefficients ψ_k et ζ_k peuvent être estimés en ajustant un modèle de ARMA par exemple. Par ailleurs, le passé lointain n'est pas observé puisque les observations $X_1, ..., X_n$ possèdent un instant initial. Lorsqu'on dispose de suffisamment de données, ce n'est pas un problème insurmontable car les coefficients de la représentation $AR(\infty)$ ou $MA(\infty)$ d'un processus ARMA (sans racine unitaire) décroissent exponentiellement vite. Les observations du passé lointain deviennent donc assez vite négligeables par rapport au passé immédiat.

En pratique, une fois le modèle ajusté et les paramètres $m, \sigma^2, \psi_k, \zeta_k$ estimés, on construit les prédicteurs par

$$\widetilde{X}_{n+1} = \widehat{m} + \sum_{k=1}^{n} \widehat{\zeta}_k (X_{n+1-k} - \widehat{m})$$

et

$$\widetilde{X}_{n+h} = \widehat{m} + \sum_{k=1}^{h-1} \widehat{\zeta}_k (\widetilde{X}_{n+h-k} - \widehat{m}) + \sum_{k>h} \widehat{\zeta}_k (X_{n+h-k} - \widehat{m}), \ h \ge 2,$$

en posant $X_i = 0$ pout $i \leq 0$. On obtient le même résultat si on travaille sur les données recentrées pour s'abstraire de la constante. La variance de l'erreur de prédiction peut être estimée également, ce qui permet de construire des intervalles de confiance du type

$$\left[\widetilde{X}_{t+h} - q_{1-\frac{\alpha}{2}}\widehat{\tau}_h, \widetilde{X}_{t+h} + q_{1-\frac{\alpha}{2}}\widehat{\tau}_h\right]$$

qui sont asymptotiquement exacts si le modèle est Gaussien et bien spécifié. On peut même aller plus loin et affiner les bornes des intervalles en tenant compte de l'erreur d'estimation des coefficients et de

la normalité jointe asymptotiquement des estimateurs (l'équivalent de ce qui se fait en régression pour la prévision).

Pour ajuster un modèle ARIMA avec un degré de différentiation $d \geq 1$ (et donc un processus non stationnaire), il suffit d'ajuster le modèle sur le processus différencié $Y_t = (I - B)^d X_t$, $t \in \mathbb{Z}$. On construit ensuite les prédicteurs de X_t à partir des Y_t obtenus. Pour d = 1,

$$\widetilde{X}_{n+1} = X_n + \widetilde{Y}_{n+1}$$
 et $\widetilde{X}_{n+h} = \widetilde{X}_{n+h-1} + \widetilde{Y}_{n+h}$, $h \ge 2$.

On peut généraliser aux ordres supérieurs $d \geq 2$ (exercice). Le même principe s'applique pour des SARIMA avec les opérateurs $(I - B^s)^D$.

Exercice. Construire un intervalle de prévision à l'horizon $h \ge 1$ pour un processus AR(1) Gaussien centré $X_t = \phi X_{t-1} + \epsilon_t, t \in \mathbb{Z}$, puis pour un processus MA(1) Gaussien centré $X_t = \theta \epsilon_{t-1} + \epsilon_t, t \in \mathbb{Z}$ avec $-1 < \phi, \theta < 1$.

6.2 Lissage exponentiel

Le lissage exponentiel est une technique dite déterministe de prévision qui consite à lisser la série en appliquant un filtre linéaire $F = \sum_{k \geq 1} \psi_k B^k$. Les coefficients ψ_k ont une forme particulière et décroissent exponentiellement vite, pour privilégier le passé immédiat dans la prévision. On cherche alors à ajuster la série brute $X_1, ..., X_n$ à la série lissée, de manière à calibrer le filtre le plus adapté à la prévision. On mentionne quatre formes de lissages exponentiels, adaptés à des comportements différents aux niveaux de la tendance et saisonalité de la série observée. Ces méthodes ont l'avantage d'être simple à implémenter, rapide en temps de calcul et ne nécessitent pas de modélisation.

Le lissage exponentiel simple d'une série temporelle $X_1,...,X_n$ est la série $m_t,t=0,...,n$ initialisée en $m_0 \in \mathbb{R}$ et définie par la relation de récurrence

$$m_t = \alpha X_t + (1 - \alpha) m_{t-1}, t = 1, ..., n$$

où $\alpha \in]0,1[$ est le facteur de lissage. Les paramètres m_0,α peuvent être calibrés sur les données, par les moindres carrés par exemple. Le lissage exponentiel simple est adapté aux séries qui ne montrent pas de tendance ni de composante saisonnière. La prévision à l'horizon $h \geq 1$ est donnée par

$$\widehat{X}_{n+h} = \widehat{X}_{n+1} = m_n , h \ge 1.$$

Elle ne dépend pas de h, on prédit donc la série future comme constante. Si on note $\widehat{X}_t = m_{t-1}$ la série ajustée au temps t à partir des données jusqu'au temps t-1 et $\widehat{\epsilon}_t = X_t - \widehat{X}_t$ l'erreur de modélisation, on a

$$\widehat{X}_{t+1} = \alpha X_t + (1 - \alpha)\widehat{X}_t = X_t + (\alpha - 1)\widehat{\epsilon}_t.$$

Proposition 6.2. La prévision \widehat{X}_{n+1} par lissage exponentiel simple est le meilleur prédicteur linéaire dans le modèle ARIMA(0,1,1) centré $X_t = X_{t-1} + (\alpha - 1)\epsilon_{t-1} + \epsilon_t, t \in \mathbb{Z}$ calculé sur les données $(m_0, X_1, ..., X_n)$.

Preuve. Exercice. \Box

En itérant la formule de récurrence:

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha) \hat{X}_t
= \alpha X_t + \alpha (1 - \alpha) X_{t-1} + (1 - \alpha)^2 \hat{X}_{t-1}
\vdots
= \alpha X_t + \alpha (1 - \alpha) X_{t-1} + \dots + \alpha (1 - \alpha)^{t-1} X_1 + \alpha (1 - \alpha)^t m_0.$$

on remarque que le lissage exponentiel simple revient à appliquer un filtre linéaire avec des coefficients à décroissance géométrique. Une valeur de α proche de 1 accorde un poids plus importante au passé immédiat et rend vite négligeable le passé plus lointain. Au contraire, le choix de la valeur initiale m_0 a d'autant plus d'impact sur le lissage que α est proche de zéro.

Le lissage exponentiel double est construit à partir des séries

$$m_t = \alpha X_t + (1 - \alpha) m_{t-1}$$

$$\ell_t = \alpha (m_t - m_{t-1}) + (1 - \alpha) \ell_{t-1}, \ t = 1, ..., n,$$

initialisées en $m_0, \ell_0 \in \mathbb{R}$, avec $\alpha \in]0,1[$ le facteur de lissage. Ici aussi, les paramètres α, m_0, ℓ_0 peuvent être calibrés sur les données de manière à maximiser l'ajustement. Ce type de lissage est adapté à des processus non-stationnaires que l'on peut approcher localement par une droite. La prévision à l'horizon h est alors donnée par

$$\hat{X}_{n+h} = m_t + \left(h - 1 + \frac{1}{\alpha}\right)\ell_n , \ h \ge 1.$$

Celle-ci croît linéairement avec h, on prédit donc les valeurs futures de la série par une droite. En notant $\hat{X}_t = m_{t-1} + \ell_{t-1}/\alpha$ la série ajustée au temps t à partir des données jusqu'au temps t-1 et $\hat{\epsilon}_t = X_t - \hat{X}_t$ l'erreur de modélisation, on peut réécrire

$$\hat{X}_{t+1} = 2X_t - X_{t-1} + 2(\alpha - 1)\hat{\epsilon}_t + (1 - \alpha)^2 \hat{\epsilon}_{t-1}.$$

Proposition 6.3. La prévision \widehat{X}_{n+1} par lissage exponentiel double est le meilleur prédicteur linéaire dans le modèle ARIMA(0,2,2) centré $X_t = 2X_{t-1} - X_{t-2} + (1-\alpha)^2 \epsilon_{t-2} + 2(1-\alpha)\epsilon_{t-1} + \epsilon_t, t \in \mathbb{Z}$ calculé sur les données $(m_0 + \ell_0/\alpha, X_1, ..., X_n)$.

Preuve. Exercice.
$$\Box$$

Une variance du lissage exponentiel double est le lissage de Holt défini par les séries

$$m_t = \alpha X_t + (1 - \alpha)(m_{t-1} + \ell_{t-1})$$

$$\ell_t = \beta(m_t - m_{t-1}) + (1 - \beta)\ell_{t-1},$$

initialisée en $m_0, \ell_0 \in \mathbb{R}$, avec $\alpha, \beta \in]0, 1[$ des facteurs de lissage. Il est plus flexible que le lissage exponentiel double de par l'ajout d'un second facteur de lissage β et donc d'un degré de liberté supplémentaire. La série ajustée est $\hat{X}_t = m_{t-1} + \ell_{t-1}$, et la prévision à l'horizon h est donnée par

$$\widehat{X}_{n+h} = m_n + h\ell_n , h \ge 1,$$

qui est affine en h comme pour le lissage exponentiel double. Par un raisonnement similaire, on montre que la prévision \widehat{X}_{n+1} correspond au meilleur prédicteur linéaire dans le modèle ARIMA(0,2,2) centré $X_t = 2X_{t-1} - X_{t-2} + (1-\alpha)\epsilon_{t-2} + (\alpha+\alpha\beta-2)\epsilon_{t-1} + \epsilon_t, t \in \mathbb{Z}$, calculé sur les données $(m_0 + \ell_0, X_1, ..., X_n)$.

Le lissage de Holt-Winters est un type de lissage exponentiel plus général adapté aux séries qui ont un comportement saisonnier. Soient $\alpha, \beta, \gamma \in]0,1[$ des facteurs de lissage, on définit les séries

$$m_{t} = \alpha(X_{t} - s_{t-r}) + (1 - \alpha)(m_{t-1} + \ell_{t-11})$$

$$\ell_{t} = \beta(m_{t} - m_{t-1}) + (1 - \beta)\ell_{t-1}$$

$$s_{t} = \gamma(X_{t} - m_{t}) + (1 - \gamma)s_{t-r}$$

initialisées en m_0, ℓ_0 et $s_0, ..., s_{r-1}$ (avec r la période de la série). La série ajustée est alors définie par $\hat{X}_t = m_{t-1} + \ell_{t-1} + s_{t-1}$, et la prévision à l'horizon h,

$$\hat{X}_{n+h} = c_n + h\ell_n + s_{n+h-r} , h \ge 1.$$

La prévision comporte une composante affine en h et une composante saisonnière s_{n+h-r} . Par un calcul compliqué, on peut montrer que la prévision de Holt-Winters correspond au meilleur prédicteur linéaire dans un modèle $SARIMA_r(0,1,r+1,0,1,0)$ avec des coefficients particuliers qui dépendent de $\alpha, \beta, \gamma, m_0, \ell_0, s_0, ..., s_{r-1}$.