

Econométrie des Séries Temporelles Univariées

Chapitre 1 : les processus ARMA

Gilles de Truchis

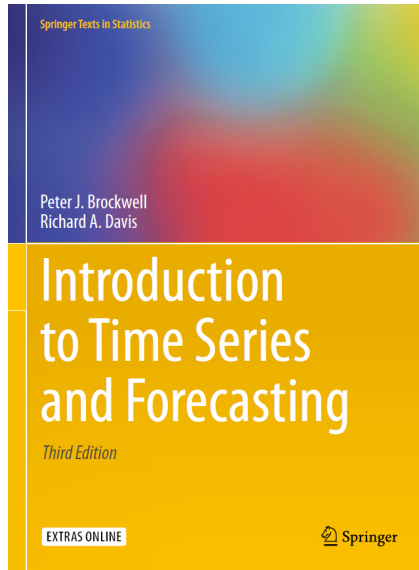
Master 1 ESA

Plan du chapitre

- 1 Introduction
- 2 Généralités
- 3 AR et MA

- 4 Stationnarité et inversibilité
- 5 Les processus ARMA
- 6 La prévision des ARMA

Références



Plan

- 1 Introduction
- 2 Généralités
- 3 AR et MA
- 4 Stationnarité et inversibilité
- 5 Les processus ARMA
- 6 La prévision des ARMA

Les séries temporelles

Definition (1)

Un **processus stochastique temporel** noté $\{X_t\}_{t \in \mathbb{Z}}$ ou simplement X_t est une séquence de variables aléatoires ordonnées dans le temps

Definition (2)

Une **série temporelle** notée $\{x_t\}_{t \in \mathbb{T}}$ ou x_t est un segment des réalisations d'un processus stochastique $\{X_t\}_{t \in \mathbb{Z}}$ avec $\mathbb{Z} \supseteq \mathbb{T}$

Definition (3)

Une **série temporelle infinie** notée $\{x_t\}_{t=-\infty}^{\infty}$ est un segment infini des réalisations d'un processus stochastique $\{X_t\}_{t \in \mathbb{Z}}$

Note 1 on suppose ici que le temps est discret ($t \in \mathbb{Z}$) mais $t \in \mathbb{R}$ est possible

Note 2 la terminologie séries temporelle dénote autant x_t que X_t

le symbol \supseteq dénote un sur-ensemble

Caractéristiques distributionnelles des séries temporelles

- Attention à ne pas confondre temps continu v.s. discret et variable continue v.s. discrète

e.g. 1 Soit l'évolution journalière des cas de contamination au variant Delta

⇒ x_t est en temps **discret** mais la réalisation d'une variable **discrète**

e.g. 2 Soit l'évolution du nombre neutrinos produits par les éruptions solaires

⇒ x_t est en temps **continu** mais la réalisation d'une variable **discrète**

e.g. 3 Soit l'évolution haute fréquence d'un cours de bourse

⇒ x_t est en temps **continu** mais la réalisation d'une variable **continue**

e.g. 4 Soit l'évolution trimestrielle du PIB en France

⇒ x_t est en temps **discret** mais la réalisation d'une variable **continue**

Analyse des séries temporelles

- Les séries temporelles observées révèlent de l'information
 - Sur le processus sous-jacent générateur des données (DGP)
 - ⇒ Qu'est-ce qu'un DGP en séries temporelles ?
- L'approche paramétrique consiste à choisir ex-ante un modèle
 - dont on pourra estimer les paramètres sous certaines hypothèses
 - ⇒ par des tests on jugera de la performance du modèle choisi
 - ⇒ on pourra également faire des prédictions des valeurs futures
- L'analyse des séries temporelles dépasse l'économie ou la finance
 - physique, hydrologie, climatologie, démographie, etc.

Modélisation des séries temporelles

Definition (4)

Pour une série observée x_t , un **modèle** est une spécification, au moins partielle, des distributions jointes de la séquence de variables aléatoires $\{X_t\}_{t \in \mathbb{Z}}$ dont on postule que les x_t sont des réalisations

- Dans le meilleur des cas, le modèle probabiliste est complet
 - Pour toute la séquence X_t , on spécifie toutes les distributions jointes du vecteur aléatoire $(X_1, X_2, \dots, X_n)'$ pour $n = 1, 2, \dots$, ou dit autrement

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad -\infty < x_1, \dots, x_n < \infty, \quad n = 1, 2, \dots$$

- Dans la pratique, obtenir un modèle complet est extrêmement difficile
 - \Rightarrow le nombre de paramètre à estimer peut dépasser le nombre de données

\Rightarrow Souvent on spécifie seulement les moments 1 et 2

$$\mathbb{E}(X_t) \text{ et } \mathbb{E}(X_{t+h}X_t), \quad t = 1, 2, \dots, \quad h = 0, 1, \dots$$

- cette approche implique en générale une perte d'information
- elle peut néanmoins se justifier dans le cadre de la prédiction (cf. S26)

Modélisation : un exemple simple

- Le bruit blanc indépendamment et identiquement distribué (*i.i.d.*)
- X_1, X_2, \dots , sont des variables aléatoires
 - dont la **moyenne est nulle**
 - dont les **distributions sont identiques**
 - dont les **distributions sont indépendantes**
- Plus formellement, pour $F(\cdot)$ une fonction de répartition :

Definition (5)

Une série temporelle X_t est un bruit blanc *i.i.d.* si $\mathbb{E}(X_t) = 0$, $\forall t$ et si

- X_t satisfait $\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \dots \mathbb{P}(X_n \leq x_n) = F(x_1) \dots F(x_n)$
- et $\mathbb{P}(X_{n+h} \leq x | X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+h} \leq x)$, $\forall h \geq 1$ (*indépendamment distribué*)

Approche globale de la modélisation des séries temporelles

- Le choix du modèle repose sur les caractéristiques de x_t

⇒ Comment identifier ces caractéristiques ?

1 Tracer la série temporelle et examiner son allure

- présence d'une tendance ?
- présence d'une composante saisonnière ?
- présence de changement radicaux dans la dynamique ?
- présence de points aberrants ?

2 Retirer les composantes déterministes (tendance et/ou saisonnalité)

3 Identifier un modèle, l'estimer et examiner les résidus pour le valider

4 Prédire la série à l'aide du modèle et évaluer la qualité des prévisions

⇒ Comparer avec des modèles alternatifs les points 3 et 4

Plan

- 1 Introduction
- 2 Généralités
- 3 AR et MA
- 4 Stationnarité et inversibilité
- 5 Les processus ARMA
- 6 La prévision des ARMA

Stationnarité stricte

- Se limiter à spécifier les moments 1 et 2 mène à la **stationnarité faible**
- Mais qu'est-ce que la stationnarité ?

Definition (6)

Un processus stochastique $\{X_t\}_{t \in \mathbb{Z}}$ est strictement stationnaire si

$$(X_1, X_2, \dots, X_n) \stackrel{d}{=} (X_{1+h}, X_{2+h}, \dots, X_{n+h})$$

pour $h, n \geq 1$

e.g. Un bruit blanc *i.i.d.* Gaussien est **strictement et faiblement stationnaire**

Note 1 Les processus strictement stationnaire ne sont pas forcément faiblement stationnaire

Note 2 Le terme stationnarité fait néanmoins référence à la forme faible dans la suite du cours

Moments inconditionnels

Definition (7)

On peut exprimer les moments inconditionnels centrés et non-centrés de X_t comme des espérances de $h(X_t)$, une fonction continue de X_t

$$\mathbb{E}(h(X_t)) = \int h(X_t)f(X_t)dX_t$$

avec $f(X_t)$ la fonction de densité inconditionnelle de X_t

- Pour calculer l'espérance de X_t on a $h(X_t) = X_t$

$$\mathbb{E}(X_t) = \mu_t$$

- Pour calculer la variance de X_t on a $h(X_t) = (X_t - \mathbb{E}(X_t))^2$

$$\mathbb{V}(X_t) = \sigma_t^2$$

Fonction d'autocovariance

- Pour comprendre la forme faible de la stationnarité introduisons la fonction d'autocovariance (ACovF)

Definition (8)

La fonction **d'autocovariance** de X_t est dérivée de la densité jointe de

$$(X_t, X_{t-1}, \dots, X_{t-h})$$

et se note $\gamma_X(t, t+h)$

$$\begin{aligned}\gamma_X(t, t+h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \mathbb{E}\left((X_t - \mu_t)(X_{t-h} - \mu_{t-h})\right) \\ &= \int \dots \int (X_t - \mu_t)(X_{t-h} - \mu_{t-h}) f(X_t, \dots, X_{t-h}) dX_t \dots dX_{t-h}\end{aligned}$$

avec $f(X_t, \dots, X_{t-h})$ la fonction de densité inconditionnelle de X_t, \dots

- $\gamma_X(t, t+h)$ décrit la dépendance de X_t à son passé jusqu'à l'horizon h

Stationnarité faible

Definition (9)

X_t est *faiblement stationnaire* si

- $\forall t \in \mathbb{Z}, \mathbb{E}(X_t) = \mu < \infty$

- $\forall t, h \in \mathbb{Z}, \text{Cov}(X_t, X_{t+h}) = \gamma_X(t, t+h) = \gamma_X(h) < \infty$

$\Rightarrow \forall t \in \mathbb{Z}, \mathbb{V}(X_t) = \sigma_X^2 < \infty$ car $\text{Cov}(X_t, X_{t+h}) = \mathbb{V}(X_t)$ pour $h = 0$

- Ici on cherche à résumer la stabilité en loi de X_t seulement à travers ses 2 premiers moments

\Rightarrow Cette définition est pertinente dans le cas Gaussien mais plus rarement en général

Note Dans le cas Gaussien, stationnarité stricte et faible sont équivalentes

Ergodicité

- C'est un concept complexe que nous définirons heuristiquement

Definition (10)

Soit X_t un processus strictement stationnaire. Si X_t est ergodique on a alors

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \text{Cov}(X_t, X_{t+j}) = 0$$

- Intuition : l'ergodicité impose que **la corrélation entre deux observations tend à s'annuler en moyenne** lorsque la distance temporelle augmente entre elles

Note Le concept d'ergodicité est utile pour invoquer la loi des grands nombres

Mélange

- Définissons deux ensembles informationnelles (σ -algèbres) :

$$\mathcal{F}_{-\infty}^t = \sigma(\dots, X_{t-1}, X_t) \text{ et } \mathcal{F}_t^\infty = \sigma(X_t, X_{t+1}, \dots)$$

ainsi que la dépendance entre 2 événements A et B comme

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

\Rightarrow la proba jointe de A et B – la proba de A et B considérés indépendants

Definition (11)

On définit comme coefficients de mélange

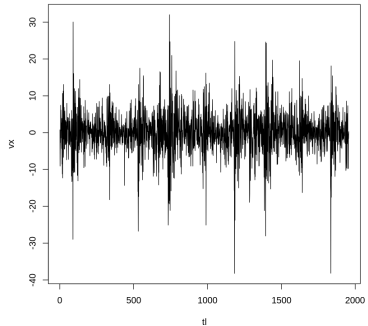
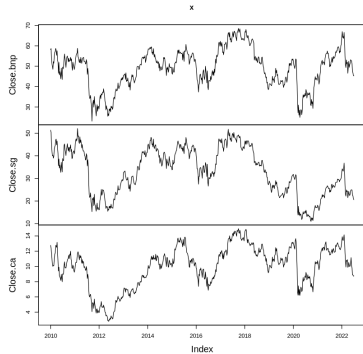
$$\alpha(\ell) = \sup_{A \in \mathcal{F}_{-\infty}^{t-\ell}, B \in \mathcal{F}_t^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

*et X_t **comme fortement mélangeant** si $\alpha(\ell) \rightarrow 0$ quand $\ell \rightarrow \infty$ et faiblement mélangeant sinon.*

Note 1 Un processus fortement mélangeant est ergodique (pas de réciproque)

Note 2 Les règles de mélanges sont importantes pour l'application des TCL

Analyse graphique de la stationnarité



Fonction d'Autocorrélation

Definition (12)

Si X_t est **faiblement stationnaire**, alors sa fonction d'autocovariance existe et on peut construire sa fonction d'autocorrélation

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\gamma_X(h)}{\sigma_X^2}$$

- A la différence de $\gamma_X(h)$, $\rho_X(h)$ est bornée sur $[-1, 1]$

Note La linéarité de l'opérateur de covariances nous assure par ailleurs que

$$\text{Cov}(aX + bY + c, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$$

si $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$, $\mathbb{E}(Z^2) < \infty$ et pour a , b et c des constantes

Note 1 $\gamma_X(h)$ existe pour tous les processus faiblement stationnaires mais pas forcément pour tous les processus strictement stationnaires

Note 2 $\rho_X(0) = \gamma_X(0)/\gamma_X(0) = 1$

Autocorrélations partielles

- $\rho_X(h)$ délivre parfois un résultat ambiguë
- Soit une dépendance non-nulle obtenue aux ordres 1 et 2 :

$$\rho_X(1) \neq 0 \text{ et } \rho_X(2) \neq 0$$

- Cela reflète-il une dépendance directe entre x_t et x_{t-2} ...

... ou une dépendance indirecte liée à une dépendance à l'ordre 1 ?

$$x_{t-2} \rightarrow x_{t-1} \text{ puis } x_{t-1} \rightarrow x_t$$

⇒ Pour lever cette ambiguïté il faudra purger $x_{t-1} \rightarrow x_t$

- Les autocorrélations partielles permettent cela comme nous le verrons par la suite

Exemple de processus stationnaire et non-stationnaire

- Les bruits blancs, notés $WN(.)$, sont des processus stationnaires

Definition (13)

Une série temporelle X_t est un **bruit blanc** si $\mathbb{E}(X_t) = 0, \forall t, \mathbb{E}(X_t^2) = \sigma_X^2$ et $Cov(X_t, X_{t+h}) = \gamma_X(h) = 0 \forall h \neq 0$

Note les processus *i.i.d.* sont des bruits blancs mais la réciproque ne tient pas

- Les marches aléatoires sont non-stationnaires (exercice : démontrez cela)

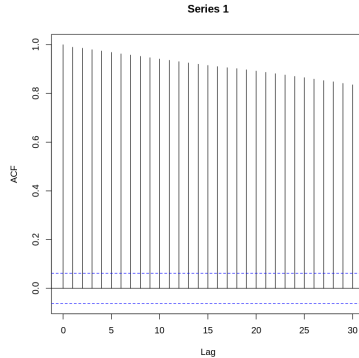
Definition (14)

Une série temporelle X_t définie comme une somme de bruits blancs *i.i.d.* est appelée **marche aléatoire**

$$S_t = X_1 + X_2 + \dots + X_t = \sum_{i=1}^t X_i, \quad \forall t > 0$$

et satisfait $\mathbb{E}(S_t) = 0, \mathbb{E}(S_t^2) = t\sigma_X^2$ et $\gamma_S(t, t+h) = t\sigma_X^2$

Analyse graphique d'une ACF



Autocorrelation et prédiction

- Supposons que X_t soit Gaussien : $X_t \sim \mathcal{N}(\mu_X, \sigma_X^2)$
- $\rho_X(h)$ va décrire la dépendance entre X_t et X_{t+h}
- On peut montrer que la **distribution conditionnelle** de

$$X_{n+h} | X_n = x_n$$

est donnée par

$$\mathcal{N}(\mu + \rho_X(h)(x_n - \mu), \sigma_X^2(1 - \rho_X(h)^2))$$

\Rightarrow il s'agit donc d'une distribution prédictive car x_n est observable mais pas X_{n+h}

Note cela suggère que $\mu + \rho_X(h)(x_n - \mu)$ est une prédiction de X_{n+h} car

$$\mathbb{E}(X_{n+h} | X_n = x_n) = \mu + \rho_X(h)(x_n - \mu)$$

- Mais que peut-on dire de la précision de cette prédiction ?

Le critère de l'erreur quadratique moyenne

- La prédiction la plus précise sera celle qui minimise la **variance de l'erreur de prédiction** (MSE)

$$\mathbb{E}\{X_{n+h} - m(X_n)\}^2$$

où $m(Y)$ est une certaine fonction (un prédicteur) de X_n

- Avec $X = X_{n+h}$, $Y = X_n$ et $f_{X|Y}(\cdot)$ la densité conditionnelle on a

$$\begin{aligned}\mathbb{E}\left((X - m(Y))^2\right) &= \int_{-\infty}^{\infty} (x - m(y))^2 f_{X|Y}(x|y) dx \\ &= \int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y) dx - 2m(Y) \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \\ &\quad + m(Y)^2 \int_{-\infty}^{\infty} f_{X|Y}(x|y) dx\end{aligned}$$

- C'est donc cette expression que l'on cherche à minimiser

Le meilleur prédicteur dans le cas Gaussien

- Résolvons ce problème d'optimisation :

Note $\int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y) dx$ ne compte pas et $\int_{-\infty}^{\infty} f_{X|Y}(x|y) dx = 1$ donc

$$\begin{aligned}\arg \min_{m(Y)} \mathbb{E}((X - m(Y))^2) &= \arg \min_{m(Y)} \left(-2m(Y) \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx + m(Y)^2 \right) \\ &= \arg \min_{m(Y)} \left(-2m(Y) \mathbb{E}(X|Y) + m(Y)^2 \right)\end{aligned}$$

- Les conditions de première ordre impliquent donc :

$$\nabla \mathbb{E}((X - m(Y))^2) = -2\mathbb{E}(X|Y) + 2m(Y) = 0 \iff m(Y) = \mathbb{E}(X_{n+h}|X_n)$$

- Ce résultat confirme que le meilleur prédicteur au sens de l'erreur quadratique moyenne (le plus précis) est

$$\mathbb{E}(X_{n+h}|X_n = x_n) = \mu + \rho_X(h)(x_n - \mu)$$

Note 1 En remplaçant on obtient que l'erreur quadratique moyenne est

$$\mathbb{E}(X_{n+h} - m(X_n))^2 = \sigma_X^2(1 - \rho_X(h)^2)$$

Le meilleur prédicteur linéaire

■ Dans un cadre non-Gaussien, c'est bien plus complexe...

... à moins qu'une hypothèse de linéarité soit faite $m(Y) = \ell(Y) = aY + b$

\Rightarrow le calcul précédent reste valide et $\ell(Y) = \mu + \rho_X(h)(x_n - \mu)$

Note 1 Le cadre Gaussien impose une structure linéaire qui peut se perdre sous d'autres hypothèses distributionnelles

Note 2 Imposer un prédicteur linéaire relâche l'hypothèse de Normalité **mais**

... soit on restreint implicitement les choix de distributions possibles

... soit on cherche un prédicteur linéaire dans un cadre non-linéaire (information omise)

Séries temporelles linéaires

⇒ Se limiter aux séries temporelles linéaires simplifie le cadre d'analyse

- Une série temporelle $\{X_t\}_{t \in \mathbb{Z}}$ est un **processus linéaire** si

$$X_t = \ell(\dots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \dots)$$

où $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ et $\ell(\cdot)$ est une fonction linéaire

Definition (15)

Un processus est donc linéaire s'il admet une représentation du type

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$$

avec $\psi_j \in \mathbb{R}$ des coefficients constant (filtre linéaire) et $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

- Introduisons l'opérateur retard $L^j X_t = X_{t-j}$
- Une réécriture plus compacte est alors possible

$$X_t = \Psi(L) \varepsilon_t$$

où $\psi(L) = \sum_{j=-\infty}^{\infty} \psi_j L^j$

Séries temporelles linéaires et convergence

- La **convergence absolue** des ψ_j est cruciale pour la convergence de X_t

⇒ En effet, on sait que $\mathbb{E}(|\varepsilon_t|) \leq \sigma$ et donc

$$\mathbb{E}(|X_t|) \leq \sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}(|\varepsilon_{t-j}|) \leq \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right) \sigma < \infty$$

- Elle assure également la convergence en moyenne quadratique de X_t

⇒ En effet, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ implique $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ (et l'ergodicité)

- Or, pour $S_n = \sum_{j=-n}^n \psi_j \varepsilon_{t-j}$ et $0 < m < n$ on a

$$\mathbb{E}\left((S_m - S_n)^2\right) = \mathbb{E}\left(\left(\sum_{m < |j| \leq n} \psi_j \varepsilon_j\right)^2\right) = \sigma_\varepsilon^2 \sum_{m < |j| \leq n} \psi_j^2 \rightarrow 0$$

car $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ est une condition équivalente à $\sum_{m < |j| \leq n} \psi_j^2 \rightarrow 0$

Note Ces calculs reposent sur l'inégalité de Jensen et l'inégalité triangulaire

Convergence et stationnarité

Theorem (1)

Soit Y_t une série stationnaire d'espérance nulle et d'autocovariance γ_Y . Si $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, alors

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \Psi(L)Y_t$$

est stationnaire d'espérance nulle et d'autocovariance

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h + k - j)$$

Note Si Y_t est strictement stationnaire, X_t l'est aussi

Convergence, stationnarité et linéarité

Theorem (2)

Si X_t est linéaire, i.e. $Y_t = \varepsilon_t \sim WN(0, \sigma_Y^2)$, on a alors

$$\gamma_X(h) = \sigma_Y^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}$$

car $\mathbb{E}(X_t) = \sum_{j=-\infty}^{\infty} \psi_j \mathbb{E}(Y_t) = 0$ et

$$\mathbb{E}(X_{t+h}, X_t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \mathbb{E}(Y_{t+h-j}, Y_{t-k})$$

avec $\mathbb{E}(Y_{t+h-j}, Y_{t-k}) = \gamma_Y(h-j+k) = \sigma_Y^2$ si $k = j-h$ et 0 sinon

Théorème de Wold

- Le Théorème 2 est le corollaire du théorème suivant

Theorem (3)

Si X_t est un processus stationnaire en covariance et σ_ε^2 la variance des erreurs projetée, alors X_t admet la représentation linéaire suivante

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

ou $\mu_t = \lim_{m \rightarrow \infty} P_{t-m}(X_t)$ est une composante déterministe (parfaitement prédictible) et possiblement constante ($\mu_t = \mu$)

Note 1 La **décomposition de Wold** est basée sur l'unicité des erreurs d'un projecteur linéaire $P_t(\cdot)$,

$$\varepsilon_t = X_t - P_{t-1}(X_t)$$

Note 2 Les erreurs sont non-corrélées, $\mathbb{E}(\varepsilon_t) = 0$ et $\sigma_\varepsilon^2 = \mathbb{E}(\varepsilon_t^2) \leq \mathbb{E}(X_t^2) < \infty$

Moyenne mobile infinie : $MA(\infty)$

- Soit X_t un processus linéaire

⇒ Sans restriction particulière, X_t dépend de son passé et son futur

- Si on suppose que pour tout $j < 0$, $\psi_j = 0$, le processus est dit **causal**

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

⇒ On parle de **représentation moyenne mobile infinie** ou $MA(\infty)$

- D'après le Théorème 2 on sait que

$$\gamma_X(h) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$$

et

$$\mathbb{V}(X_t) = \gamma_X(0) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2$$

- Par le Théorème 1 on sait que si $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, X_t est stationnaire

Plan

- 1 Introduction
- 2 Généralités
- 3 AR et MA
- 4 Stationnarité et inversibilité
- 5 Les processus ARMA
- 6 La prévision des ARMA

Moyenne mobile d'ordre q : $MA(q)$

- La représentation $MA(\infty)$ est très générale
- Supposons un cas particulier où X_t **dépend de q innovations passées**

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \Leftrightarrow X_t = \Theta(L) \varepsilon_t$$

$\Rightarrow X_t$ est donc corrélé avec ses q observations passées

- D'après les formules du processus $MA(\infty)$ on en déduit

$$\gamma_X(h) = \sigma_\varepsilon^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}, \quad h \leq q$$

et

$$\mathbb{V}(X_t) = \gamma_X(0) = \sigma_\varepsilon^2 \sum_{j=0}^q \theta_j^2$$

et surtout

$$\gamma_X(h) = \rho_X(h) = 0, \quad h > q$$

Note 1 Un $MA(q < \infty)$ est toujours stationnaire car $\gamma_X(t, t+h) = \gamma_X(h) < \infty$

Note 2 $c = \mu_X = \mathbb{E}(X_t) \neq 0$ pourrait intégrer le $MA(q)$ sans affecter sa structure

La convention de Box-Jenkins

- Box et Jenkins sont des économètres célèbres ayant développé une méthodologie d'analyse systématique des séries temporelles
- Elle est implémentée dans de nombreux logiciels avec une convention de signe négatif pour les écritures MA

⇒ Si X_t suit un $MA(q)$, il s'écrira alors

$$X_t = \varepsilon_t - \vartheta_1 \varepsilon_{t-1} - \vartheta_2 \varepsilon_{t-2} - \dots - \vartheta_q \varepsilon_{t-q}$$

ou $\vartheta_j = -\theta_j$

- En théorie c'est équivalent

Note Si vous estimez un $MA(1)$, sous cette convention le logiciel renverra, e.g., $\hat{\vartheta}_1 = 0.8$ et vous devrez lire $\theta_j = -0.8$

Moyenne mobile d'ordre 1 : MA(1)

- Dans le cas très spécifique où X_t dépend de 1 innovation passée

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$\Rightarrow X_t$ est donc corrélé avec une unique 1 observation passée

- D'après les formules du MA(q) on en déduit pour l'**ACovF**

$$\gamma_X(0) = \mathbb{V}(X_t) = \mathbb{E}(X_t X_t) = (1 + \theta_1^2) \sigma_\varepsilon^2$$

$$\gamma_X(1) = \mathbb{E}(X_t X_{t-1}) = \mathbb{E}((\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2})) = \theta_1 \sigma_\varepsilon^2$$

$$\gamma_X(j) = \mathbb{E}(X_t X_{t-j}) = 0, \quad j > 1$$

car $\mathbb{E}(\varepsilon_t \varepsilon_{t-j}) = 0$, $j \neq 0$ et σ_ε^2 sinon

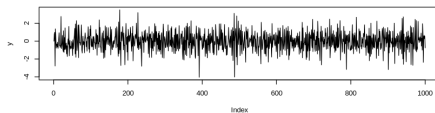
- Pour l'**ACF** on obtient donc

$$\rho_0 = 1$$

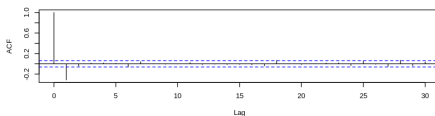
$$\rho_X(1) = \frac{\gamma_X(1)}{\gamma_X(0)} = \frac{\theta_1}{1 + \theta_1^2}$$

$$\rho_X(h) = 0, \quad h > 1$$

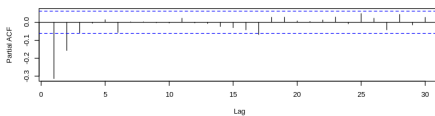
Analyse graphique d'un MA(1) : $X_t = \varepsilon_t - 0.4\varepsilon_{t-1}$



Series 1



Series y



Identification des MA(1) : problème

- A l'inverse, peut-on déduire de $\rho_X(1)$ la forme d'un MA(1) ?

⇒ D'après la formule de $\rho_X(1)$ les racines de $\rho_X(1)\theta_1^2 - \theta_1 + \rho_X(1) = 0$ sont

$$\theta_1^{(1)} = \frac{-1 + \sqrt{1 - 4\rho_X(1)^2}}{2\rho_X(1)}$$

$$\theta_1^{(2)} = \frac{-1 - \sqrt{1 - 4\rho_X(1)^2}}{2\rho_X(1)}$$

- On peut montrer qu'elles sont réelles que si $-1/2 \leq \rho_X(1) < 1/2$

Note 1 Si $\rho_X(0) = 1, \rho_X(1) = 0.6, \rho_X(h > 1) = 0$, $\rho_X(h)$ n'est pas la fonction d'autocorrélation d'un MA(1)

Note 2 Si $\rho_X(0) = 1, \rho_X(1) = 0.3, \rho_X(h > 1) = 0$, $\rho_X(h)$ est la fonction d'autocorrélation d'un MA(1)

Note 3 $\rho_X(1) = 0.3$ mène à deux solutions réelles $\theta_1^{(1)} = -1/3$ et $\theta_1^{(2)} = -3$

⇒ Ces 2 MA(1) ont la même fonction $\rho_X(h)$ et **ne sont pas identifiables !**

Inversibilité des MA(1) : intuition

- $\rho_X(1)$ décrit la dépendance entre X_t et X_{t-1} ...

... or le modèle MA(1) se formule comme une dépendance entre X_t et ε_{t-1}

- En utilisant l'opérateur retard L on voit pourtant une réécriture possible

$$X_t = (1 + \theta_1 L)\varepsilon_t \Rightarrow (1 + \theta_1 L)^{-1}X_t = \varepsilon_t$$

- La validité de cette écriture va dépendre de **l'inversibilité** de $(1 + \theta_1 L)$

Inversibilité des MA(1) : démonstration

- Repartons de $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$

$$\begin{aligned}
 \varepsilon_t &= X_t - \theta_1 \varepsilon_{t-1} \\
 &= X_t - \theta_1 (X_{t-1} - \theta_1 \varepsilon_{t-2}) = X_t - \theta_1 X_{t-1} + \theta_1^2 \varepsilon_{t-2} \\
 &= X_t - \theta_1 X_{t-1} + \theta_1^2 (X_{t-2} - \theta_1 \varepsilon_{t-3}) \\
 &= X_t - \theta_1 X_{t-1} + \theta_1^2 X_{t-2} - \theta_1^3 \varepsilon_{t-3} = \dots \\
 &= \sum_{j=0}^l (-\theta_1)^j X_{t-j} + \theta_1^{l+1} \varepsilon_{t-l-1}
 \end{aligned}$$

- Pour que ε_t soit défini, la somme doit converger et donc $|\theta| < 1$
- \Rightarrow le deuxième terme va alors tendre vers 0 quand $l \rightarrow \infty$ et on obtient

$$\varepsilon_t = \sum_{j=0}^l \theta_1^j X_{t-j} = \left(\sum_{j=0}^l \theta_1^j L^j \right) X_t$$

- Mais nous sommes partis de $X_t = (1 + \theta_1 L) \varepsilon_t \dots$

... pour arriver à une représentation valide de ε_t impliquant que

$$\varepsilon_t = (1 + \theta_1 L)^{-1} X_t$$

est valide si $|\theta_1| < 1$

Inversibilité des MA(1) : application

- On constate que si la condition d'inversibilité $|\theta_1| < 1$ est vérifiée ...

... alors $X_t = (1 + \theta_1 L)\varepsilon_t$,

$$\begin{aligned}\varepsilon_t &= (1 + \theta_1 L)^{-1} X_t \\ &= \sum_{j=0}^{\infty} \theta_1^j X_{t-j}\end{aligned}$$

et $X_t = -\sum_{j=1}^{\infty} \theta_1^j X_{t-j} + \varepsilon_t$ sont des écritures équivalentes

- Cette forme écrite sur les retards de X_t est dite $\text{AR}(\infty)$

Note 1 On verra que tout $\text{MA}(q)$ inversible admet une forme $\text{AR}(\infty)$

Note 2 En retenant uniquement les processus inversibles on résoud le problème d'identification

\Rightarrow Nous avons 2 solutions réelles $\theta_1^{(1)} = -1/3$ et $\theta_1^{(2)} = -3$ mais seul $\theta_1^{(1)}$ sera retenu

Processus auto-régressif d'ordre p : $AR(p)$

- Tout comme l'on a introduit les $MA(q)$ on peut construire des $AR(p)$

$$\tilde{X}_t = c + \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \dots + \phi_p \tilde{X}_{t-p} + \varepsilon_t$$

- Mais contrairement au MA , la présence de c a des conséquences car

$$\mu_{\tilde{X}} = \mathbb{E}(\tilde{X}_t) = c + \phi_1 \mu_{\tilde{X}} + \dots + \phi_p \mu_{\tilde{X}} \Rightarrow c = \mu_{\tilde{X}}(1 - \phi_1 - \dots - \phi_p)$$

$\Rightarrow \mathbb{E}(\tilde{X}_t) \neq c$ et travailler avec $X_t = \tilde{X}_t - \mu_{\tilde{X}}$ est plus simple

- En utilisant l'opérateur retard L on peut également écrire

$$(1 - \phi_1 L - \dots - \phi_p L^p)X_t = \varepsilon_t \Leftrightarrow \Phi(L)X_t = \varepsilon_t$$

Note 1 Pour les MA , la stationnarité des innovations suffisait à assurer celle de X_t mais cela ne tient plus pour les $AR(p)$

Note 2 La **stationnarité** de X_t va ici reposer sur l'inversibilité de $\Phi(L)$ car si

$$X_t = \Phi(L)^{-1} \varepsilon_t$$

est valide et que $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, en vertu du Th. 3, X_t est stationnaire

AR(1) et solution stationnaire

- L'AR(1) est le processus le plus simple de la classe des AR(p)

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \Leftrightarrow (1 - \phi_1 L)X_t = \varepsilon_t \Leftrightarrow X_t = (1 - \phi_1 L)^{-1} \varepsilon_t$$

- Comme pour l'inversibilité des MA, supposons Y_t un AR(1) stationnaire basé sur ε_t et

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \varepsilon_t \\ &= \varepsilon_t + \phi_1 (\phi_1 Y_{t-2} + \varepsilon_{t-1}) = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 (\varepsilon_{t-2} + \phi_1 Y_{t-3}) \\ &= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 Y_{t-3} = \dots \end{aligned}$$

$$Y_t = \sum_{j=0}^l \phi_1^j \varepsilon_{t-j} + \phi_1^{l+1} Y_{t-l-1}$$

- Comme pour le cas des MA on voit que la suite converge si $|\phi_1| < 1 \dots$

... et que le second terme tend vers 0 quand $l \rightarrow \infty$

$\Rightarrow X_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}$ est l'**unique solution stationnaire** de

$$X_t = \phi_1 X_{t-1} + \varepsilon_t$$

AR(1) causal vs non-causal

- Le résultat est cohérent avec le Th. 1

⇒ Si $|\phi_1| < 1$ on sait que $\sum_{j=0}^{\infty} |\phi_1^j| < \infty$ et donc X_t est stationnaire

- Par ailleurs, si X_t est stationnaire, $\mathbb{E}(X_t)^2 < \infty$ et quand $l \rightarrow \infty$

$$\mathbb{E}\left(X_t - \sum_{j=0}^l \phi_1^j \varepsilon_{t-j}\right)^2 = \phi_1^{2l+2} \mathbb{E}(X_{t-l-1})^2 \rightarrow 0$$

- Mais qu'en est-il du cas $|\phi_1| > 1$ puisque $\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}$ ne converge pas ?

- A partir de $X_{t+1} = \phi_1 X_t + \varepsilon_{t+1}$ on peut écrire $X_t = \phi_1^{-1} X_{t+1} - \phi_1^{-1} \varepsilon_{t+1}$

$$\begin{aligned} X_t &= -\phi_1^{-1} \varepsilon_{t+1} - \phi_1^{-2} \varepsilon_{t+2} + \phi_1^{-2} X_{t+2} \\ &= \dots \end{aligned}$$

$$= -\sum_{j=1}^l \phi_1^{-j} \varepsilon_{t+j} + \phi_1^{-l-1} X_{t+l+1} \Rightarrow X_t = -\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j} \text{ si } |\phi_1| > 1$$

⇒ C'est toujours la même solution stationnaire mais écrite vers le futur

AR(1) solution non-stationnaire

- Solutions causale et non-causale sont similaires sur le plan probabiliste

... dans le cadre (restreint) de la **stationnarité en covariance** !

⇒ Le **causal** est plus naturel, nous le retiendrons (on exclut $|\phi_1| > 1$)

- Qu'en est-il du cas $|\phi_1| = 1$?

$$X_t = \varepsilon_t + \phi_1(\phi_1 X_{t-2} + \varepsilon_{t-1}) = \dots = \sum_{j=0}^l \phi_1^j \varepsilon_{t-j} + \phi_1^{l+1} X_{t-l-1}$$

$$\Rightarrow X_t - \phi_1^{l+1} X_{t-l-1} = \sum_{j=0}^l \phi_1^j \varepsilon_{t-j}$$

- S'il existe une solution stationnaire, on devrait avoir $\gamma_X(0) < \infty$ mais

$$\mathbb{V}\left(\sum_{j=0}^l \phi_1^j \varepsilon_{t-j}\right) = \sum_{j=0}^l \phi_1^{2j} \mathbb{V}(\varepsilon_{t-j}) = \sum_{j=0}^l \sigma_\varepsilon^2 = (l+1)\sigma_\varepsilon^2$$

$$\mathbb{V}\left(X_t - \phi_1^{l+1} X_{t-l-1}\right) = 2\gamma_X(0) - 2\phi_1^{l+1} \gamma_X(l+1) \leq 2\gamma_X(0) + 2\gamma_X(l+1) \leq 4\gamma_X(0)$$

et donc $(l+1)\sigma_\varepsilon^2 \leq 4\gamma_X(0)$ ce qui implique $\gamma_X(0) = \infty$ si $l \rightarrow \infty$

AR(1) et ACF

- D'après Th. 2 on peut calculer $\gamma_X(h)$ et $\rho_X(h)$ pour $|\phi_1| < 1$
- C'est moins direct que pour les MA donc détaillons $\gamma_X(0), \gamma_X(1), \dots, \gamma_X(h)$

$$\gamma_X(0) = \mathbb{V}(X_t) = \mathbb{V}(\phi_1 X_{t-1} + \varepsilon_t) = \phi_1^2 \gamma_X(0) + \sigma_\varepsilon^2 \Rightarrow \gamma_X(0) = \sigma_\varepsilon^2 / (1 - \phi_1^2)$$

$$\gamma_X(1) = \mathbb{E}(X_t X_{t-1}) = \mathbb{E}(\phi_1 X_{t-1}^2 + \varepsilon_t X_{t-1}) = \phi_1 \gamma_X(0)$$

$$\begin{aligned} \gamma_X(2) &= \mathbb{E}(X_t X_{t-2}) = \mathbb{E}(\phi_1 X_{t-1} X_{t-2} + \varepsilon_t X_{t-2}) = \phi_1 \gamma_X(1) = \phi_1^2 \gamma_X(0) \\ &= \dots \end{aligned}$$

$$\gamma_X(h) = \mathbb{E}(X_t X_{t-h}) = \phi_1 \gamma_X(h-1) = \phi_1^h \gamma_X(0), \quad h \geq 1$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \phi_1^j \phi_1^{j+h} = \frac{\sigma_\varepsilon^2 \phi_1^h}{1 - \phi_1^2}$$

$$\text{car } \mathbb{E}(\varepsilon_t X_{t-h}) = 0, \quad \forall h$$

$$\Rightarrow \text{Pour } \forall h \geq 0, \text{ l'ACF d'un AR(1) est } \rho_X(h) = \gamma_X(h) / \gamma_X(0) = \phi_1^h$$

AR(p) et ACF

- Repartons de l'AR(p) : $X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t$ avec $\phi_j = 0$ si $j > p$
- On suppose X_t stationnaire et les autocovariances sont alors

$$\gamma_X(0) = \mathbb{E}(X_t X_t) = \sum_{j=1}^p \phi_j \mathbb{E}(X_t X_{t-j}) + \mathbb{E}(X_t \varepsilon_t) = \sum_{j=1}^p \phi_j \gamma_X(j) + \sigma_\varepsilon^2$$

$$\gamma_X(1) = \mathbb{E}(X_t X_{t-1}) = \sum_{j=1}^p \phi_j \mathbb{E}(X_{t-1} X_{t-j}) + \mathbb{E}(X_{t-1} \varepsilon_t) = \sum_{j=1}^p \phi_j \gamma_X(j-1)$$

$$\vdots$$

$$\gamma_X(h) = \mathbb{E}(X_t X_{t-h}) = \sum_{j=1}^p \phi_j \mathbb{E}(X_{t-h} X_{t-j}) + \mathbb{E}(X_{t-h} \varepsilon_t) = \sum_{j=1}^p \phi_j \gamma_X(j-h)$$

- Les ACFs sont alors $\rho_X(h) = \gamma_X(h)/\gamma_X(0) = \sum_{j=1}^p \phi_j \rho_X(h-j)$

Note On rappelle que $\gamma_X(h) = \gamma_X(-h)$ car $\gamma_X(\cdot)$ est symétrique

Les équations de Yule-Walker

- Elles facilitent le passage des ACF vers les ϕ_h si l'on connaît les

$$\rho_X(h) = \rho_h$$

- Pour p retards il faut construire un système à p équations

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

- En inversant le système on obtient alors

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} \quad (1)$$

Yule-Walker et autocorrélations partielles (PACF)

- Comme précisé au S20, l'interprétation de ρ_j peut-être ambiguë
- Si l'on veut connaître la PACF d'ordre j dans notre AR(p) on considère

$$X_t = \phi_{1,j}X_{t-j} + \cdots + \phi_{j,j}X_{t-j} + \varepsilon_t$$

la **corrélation partielle d'ordre j** étant $\phi_{j,j}$ (dépendance directe entre x_t et x_{t-j})

- Grâce à Yule-Walker si on résout le système à l'ordre j on retrouve $\phi_{j,j}$

\Rightarrow Pour $j = 1$ on a

$$\phi_{1,1} = (1)^{-1}\rho_1$$

\Rightarrow Pour $j = 2$ on a

$$\begin{pmatrix} \phi_{2,1} \\ \phi_{2,2} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$$

ou $\phi_{2,2} \neq \rho_2$ donne la corrélation partielle d'ordre 2

PACF pour un AR(2) : exercice

- Soit un AR(2) qu'on suppose stationnaire

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \Leftrightarrow (1 - \phi_1 L - \phi_2 L^2) X_t = \varepsilon_t$$

- Calculez l'ACF à l'ordre $h > 2$
- Calculez la PACF à l'ordre $j = 1, 2$ et 3

Note Aidez-vous de la structure de l'ACovF de l'AR(2)

$$\gamma_X(0) = \mathbb{E}(X_t X_t) = \mathbb{E}(\phi_1 X_t X_{t-1} + \phi_2 X_t X_{t-2} + X_t \varepsilon_t) = \phi_1 \gamma_X(1) + \phi_2 \gamma_X(2) + \sigma_\varepsilon^2$$

$$\gamma_X(1) = \mathbb{E}(X_t X_{t-1}) = \mathbb{E}(\phi_1 X_{t-1}^2 + \phi_2 X_{t-1} X_{t-2} + X_{t-1} \varepsilon_t) = \phi_1 \gamma_X(0) + \phi_2 \gamma_X(1)$$

$$\gamma_X(2) = \mathbb{E}(X_t X_{t-2}) = \mathbb{E}(\phi_1 X_{t-2} X_{t-1} + \phi_2 X_{t-2}^2 + X_{t-2} \varepsilon_t) = \phi_1 \gamma_X(1) + \phi_2 \gamma_X(0)$$

$$\begin{aligned} \gamma_X(h) &= \mathbb{E}(X_t X_{t-h}) = \mathbb{E}(\phi_1 X_{t-h} X_{t-1} + \phi_2 X_{t-h} X_{t-2} + X_{t-h} \varepsilon_t) \\ &= \phi_1 \gamma_X(h-1) + \phi_2 \gamma_X(h-2) \end{aligned}$$

PACF pour un AR(2) : solution

- Soit un AR(2) qu'on suppose stationnaire

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

- Calculez l'ACF à l'ordre $h > 2$

$$\rho_0 = 1$$

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \Rightarrow \rho_1 = \phi_1 / (1 - \phi_2)$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0 \Rightarrow \rho_2 = \phi_2 + \phi_1^2 / (1 - \phi_2)$$

$$\rho_h = \phi_1 \rho_{h-1} + \phi_2 \rho_{h-2}, \quad h > 2$$

- Calculez la PACF à l'ordre $j = 1, 2$ et 3

$$\phi_{1,1} = \rho_1$$

$$\phi_{2,2} = \phi_2 \text{ car } (\phi_{1,1} X_{t-1} + \phi_{2,2} X_{t-2}) \Leftrightarrow (\phi_1 X_{t-1} + \phi_2 X_{t-2})$$

$$\phi_{3,3} = 0 \text{ car } 3 > p = 2$$

\Rightarrow Pour un AR(p) on voit un **passage brutal à 0 des PACFs**

Note Pour les MA(q) ce ne sera pas le cas comme nous allons le voir

PACF pour un MA(1) : exemple

- Soit le MA(1) du S40 que l'on suppose inversible : $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$
- Pour la PACF à l'ordre j on considère un AR(j) :

$$X_t = \sum_{i=1}^j \phi_{i,j}^j X_{t-i} + \varepsilon_t$$

et les équations de Yule-Walker

$$\begin{pmatrix} \phi_{1,j} \\ \phi_{2,j} \\ \phi_{3,j} \\ \vdots \\ \phi_{j,j} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & 0 & \cdots & 0 \\ \rho_1 & 1 & \rho_1 & \cdots & 0 \\ 0 & \rho_1 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- En résolvant pour $j = 1, 2, \dots$ on obtient

$$\phi_{1,1} = \rho_1, \quad \phi_{2,2} = -\theta_1^2 / (1 + \theta_1^2 + \theta_1^4), \quad \phi_{3,3} = \theta_1^3 / (1 + \theta_1^2 + \theta_1^4 + \theta_1^6)$$

et pour tout j on a $\phi_{j,j} = -(-\theta_1)^j / (1 + \theta_1^2 + \dots + \theta_1^{2j})$

Note Si $\theta < 0$ tous les $\phi_{j,j} < 0$ et si $\theta > 0$ le signe de $\phi_{j,j}$ alternera selon que j est pair ou impair

PACF pour un MA(2) : exercice

- Soit un MA(2) que l'on suppose inversible : $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$
- Calculez l'ACF à l'ordre $h > 2$
- Calculez la PACF à l'ordre $j = 1$ et 2

Note Aidez-vous de la structure de l'ACovF du MA(2)

$$\gamma_X(0) = \mathbb{E}(X_t X_t) = (1 + \theta_1^2 + \theta_2^2) \sigma_\varepsilon^2$$

$$\begin{aligned}\gamma_X(1) &= \mathbb{E}(X_t X_{t-1}) = \mathbb{E}((\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2})(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3})) \\ &= (\theta_1 + \theta_1 \theta_2) \sigma_\varepsilon^2\end{aligned}$$

$$\begin{aligned}\gamma_X(2) &= \mathbb{E}(X_t X_{t-2}) = \mathbb{E}((\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2})(\varepsilon_{t-2} + \theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4})) \\ &= \theta_2 \sigma_\varepsilon^2\end{aligned}$$

$$\gamma_X(h) = \mathbb{E}(X_t X_{t-j}) = 0, \quad \forall h > 2$$

PACF pour un MA(2) : solution

- Soit un MA(2) que l'on suppose inversible : $X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$
- Pour l'ACF à l'ordre $h > 2$ on a $\rho_X(0) = 1$, $\rho_X(h) = 0$, $\forall h > 2$ et

$$\rho_X(1) = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_X(2) = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

- Pour la PACF à l'ordre $j = 1$ et 2 on utilise Yule-Walker

$$\begin{pmatrix} \phi_{1,j} \\ \phi_{2,j} \\ \phi_{3,j} \\ \phi_{4,j} \\ \vdots \\ \phi_{j,j} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & 0 & \cdots & 0 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & 0 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & 0 \\ 0 & \rho_2 & \rho_1 & 1 & \cdots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- En résolvant $j = 1, 2, \dots$ on obtient $\phi_{1,1} = \rho_1$, $\phi_{2,2} = \theta_2/(1 + \theta_1^2 + \theta_2^2 + \theta_2)$

$$\phi_{3,3} = -2\theta_2^2/(1 + \theta_1^2 + (\theta_2 - 1)\theta_2)(\theta_1^2 + (\theta_2 + 1)^2)$$

et pour tout $j > 3$ on a des expressions de plus en plus complexes

Plan

- 1 Introduction
- 2 Généralités
- 3 AR et MA
- 4 Stationnarité et inversibilité
- 5 Les processus ARMA
- 6 La prévision des ARMA

Conditions de stationnarité : intuition

- Repartons d'un AR(1) avec $\varepsilon_t \sim WN(0, \sigma^2)$: $(1 - \phi_1 L)X_t = \varepsilon_t$
- Th.1 et Th.3 : X_t sera stationnaire causal s'il peut s'exprimer comme

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

ce qui implique l'inversibilité de $\Phi(L) = (1 - \phi_1 L)$

- $\Phi(L)$ est un polynôme de degré 1 dont l'inverse est un polynôme infini

$$(1 - \phi_1 L)^{-1} = \sum_{j=0}^{\infty} \phi_1^j L^j = \Phi(L)^{-1}$$

comme obtenu au S43 (se voit aussi avec une division euclidienne)

$\Rightarrow \Phi(L)^{-1}$ converge si $|\phi_1| < 1$

- Une **condition en découle sur la racine de $\Phi(z)$** :

$$1 - \phi_1 z = 0 \Leftrightarrow z = \phi_1^{-1}$$

\Rightarrow Si $|\phi_1| < 1$, $|z| > 1$

Conditions de stationnarité : AR(p)

- Considérons le cadre plus général d'un AR(p) : $\Phi(L)X_t = \varepsilon_t$

$$\Phi(L) = 1 - \phi_1 L^1 - \dots - \phi_p L^p = 1 - \sum_{j=1}^p \phi_j L^j$$

- L'utilisation des conditions sur les racines va devenir cruciale ...

... car les conditions sur les coefficients sont indéterminables

⇒ Par le **Th. fondamental de l'algèbre** on sait que

$$1 - \sum_{j=1}^p \phi_j z^j = 0 \Leftrightarrow \prod_{j=1}^p (1 - \lambda_j z) = 0, \quad z \in \mathbb{C}$$

et on voit alors que les p racines sont $z = 1/\lambda_j$, $j = 1, 2, \dots, p$

⇒ Une écriture alternative de X_t est donc

$$(1 - z_1^{-1}L)(1 - z_2^{-1}L) \dots (1 - z_p^{-1}L)X_t = \varepsilon_t$$

où $|z_j| > 1 \Leftrightarrow |\lambda_j| < 1$ assure que $\exists (1 - z_j^{-1}L)^{-1}$ et donc $\exists \Phi(L)^{-1}$

Conditions de stationnarité : AR(2)

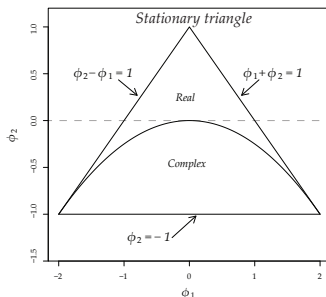
- Pour $p > 1$, les racines peuvent être complexes $z \in \mathbb{C}$
- Étudions le cas d'un AR(2), $(1 - \phi_1 L - \phi_2 L^2)X_t = \varepsilon_t$, de polynôme

$$(1 - \phi_1 z - \phi_2 z^2) = 0 \Rightarrow \lambda_1^{-1} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}, \quad \lambda_2^{-1} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$$

Note 1 On voit que selon le signe de $\phi_1^2 + 4\phi_2$, $z = \lambda_j^{-1} \in \mathbb{R}$ ou $z = \lambda_j^{-1} \in \mathbb{C}$

Note 2 On peut montrer également que $\phi_1 = \lambda_1 + \lambda_2$ et $\phi_2 = -\lambda_1 \lambda_2$

Note 3 On peut alors trouver les conditions sur les coefficients pour que $|z| > 1$



Stationnarité et AR(2) : exemple

- Soit un AR(2) : $(1 - 0.7L + 0.1L^2)X_t = \varepsilon_t$
- La factorisation du polynôme s'obtient à l'aide de $\lambda_1 = 0.5$ et $\lambda_2 = 0.2$

$$(1 - 0.7z + 0.1z^2) \Leftrightarrow (1 - \lambda_1 z)(1 - \lambda_2 z) = (1 - 0.5z)(1 - 0.2z)$$

- On voit alors que pour $|z| = 2$ et $|z| = 5$

$$(1 - \lambda_1 z)(1 - \lambda_2 z) = 0$$

et X_t est donc stationnaire puisque $|z| > 1$

- Une manière alternative de définir la condition de stationnarité est

$$\exists \phi_j, \quad j = 1, 2, \dots, p \Rightarrow \Phi(z) = (1 - \phi_1 z - \dots - \phi_p z^p) \neq 0 \quad \forall |z| \leq 1$$

Note 1 Cette définition impose une forme causale ($|z| \leq 1$ vs $|z| = 1$)

Note 2 Comme on peut avoir $z \in \mathbb{C}$, on dit que les racines sont en dehors du cercle unitaire défini par

$$\text{cis}(\vartheta) = e^{i\vartheta} = \cos \vartheta + i \sin \vartheta$$

Conditions d'inversibilité : intuition

- Repartons d'un MA(1) et de la démonstration du S40

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} = (1 + \theta_1 L) \varepsilon_t$$

- On a vu que l'écriture AR(∞) venait de l'inversion de $(1 + \theta_1 L)$

$$(1 + \theta_1 L)^{-1} = \sum_{j=0}^{\infty} \theta_1^j L^j = \Theta(L)^{-1}$$

obtenu par récursion (ou simple division euclidienne) et

- $\Theta(L)^{-1}$ converge si $|\theta_1| < 1$ et une **condition sur la racine en découle**

$$1 - \theta_1 z = 0 \Leftrightarrow z = \theta_1^{-1}$$

\Rightarrow Si $|\theta_1| < 1$, $|z| > 1$ et $\varepsilon_t = (1 + \theta_1 L)^{-1} X_t = \sum_{j=0}^{\infty} \theta_1^j X_{t-j}$ ou

$$X_t = - \sum_{j=1}^{\infty} \theta_1^j X_{t-j} + \varepsilon_t$$

sont des écritures équivalentes valides

Conditions d'inversibilité : MA(q)

- Considérons le cadre plus général d'un MA(q) : $X_t = \Theta(L)\varepsilon_t$

$$\Theta(L) = 1 + \theta_1 L^1 + \dots + \theta_q L^q = 1 + \sum_{j=1}^q \theta_j L^j$$

- L'utilisation des conditions sur les racines va devenir cruciale ...

... car les conditions sur les coefficients sont indéterminables

⇒ Par le **Th. fondamental de l'algèbre** on sait que

$$1 + \sum_{j=1}^q \theta_j z^j = 0 \Leftrightarrow \prod_{j=1}^q (1 - \lambda_j z) = 0, \quad z \in \mathbb{C}$$

et on voit alors que les q racines sont $z = 1/\lambda_j$, $j = 1, 2, \dots, p$

⇒ Une écriture alternative de $\Theta(L)$ est donc

$$(1 - z_1^{-1}L)(1 - z_2^{-1}L) \dots (1 - z_q^{-1}L)$$

où $|z_j| > 1 \Leftrightarrow |\lambda_j| < 1$ assure que $\exists (1 - z_j^{-1}L)^{-1}$ et donc $\exists \Theta(L)^{-1}$

Conditions d'inversibilité : MA(2)

- Soit un MA(2), dont on souhaite étudier les conditions d'inversibilité

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \Rightarrow (1 + \theta_1 L + \theta_2 L^2) = (1 - z_1^{-1} L)(1 - z_2^{-1} L)$$

- On sait que

$$(1 + \theta_1 z + \theta_2 z^2) = 0 \Rightarrow \lambda_1^{-1} = \frac{-\theta_1 + \sqrt{\theta_1^2 - 4\theta_2}}{2\theta_2}, \quad \lambda_2^{-1} = \frac{-\theta_1 - \sqrt{\theta_1^2 - 4\theta_2}}{2\theta_2}$$

- On peut alors déterminer pour quelles z

$$(1 - \lambda_1 z)(1 - \lambda_2 z) = 0$$

et ce MA(2) est inversible ou non selon que $|z| > 1$ ou $|z| < 1$

- Une manière alternative de définir la condition d'inversibilité est

$$\exists \theta_j, \quad j = 1, 2, \dots, q \Rightarrow \Theta(z) = (1 + \theta_1 z + \dots + \theta_q z^q) \neq 0 \quad \forall |z| \leq 1$$

Note Comme on peut avoir $z \in \mathbb{C}$, on dit que les racines sont en dehors du cercle unitaire défini par

$$\text{cis}(\vartheta) = e^{i\vartheta} = \cos \vartheta + i \sin \vartheta$$

Plan

- 1 Introduction
- 2 Généralités
- 3 AR et MA
- 4 Stationnarité et inversibilité
- 5 Les processus ARMA
- 6 La prévision des ARMA

Processus auto-régressif moyenne mobile d'ordre 1 : ARMA(1, 1)

- En combinant AR(1) et MA(1) on obtient un ARMA(1, 1)

$$X_t - \phi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1} \Leftrightarrow \Phi(L)X_t = \Theta(L)\varepsilon_t$$

- D'après les Th.1 et 3, X_t sera stationnaire si

$$X_t = \Phi(L)^{-1} \Theta(L) \varepsilon_t = \Psi(L) \varepsilon_t$$

est une écriture valide

- D'après S43, $\Phi(L)^{-1}$ admet un développement en séries infinie

$$\Phi(L)^{-1} = \sum_{j=0}^{\infty} \phi_1^j L^j$$

qui converge si $|\phi_1| < 1$ et qui implique l'écriture MA(∞) suivante

$$X_t = \varepsilon_t + (\phi_1 + \theta_1) \sum_{j=1}^{\infty} \phi_1^{j-1} \varepsilon_{t-j} = \Psi(L) \varepsilon_t$$

et démontre l'existence d'une solution stationnaire causale (unique)

ARMA(p, q)

- En combinant AR(p) et MA(q) on obtient un ARMA(p, q)

$$X_t - \sum_{j=1}^p \phi_j X_{t-j} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \Leftrightarrow \Phi(L)X_t = \Theta(L)\varepsilon_t$$

- D'après les Th.1 et 3, X_t sera stationnaire si on peut écrire

$$X_t = \Phi(L)^{-1}\Theta(L)\varepsilon_t = \Psi(L)\varepsilon_t$$

- D'après S57, $\Phi(L)^{-1}$ admet une factorisation en p polynôme et

$$\Phi(z)^{-1} = \sum_{j=0}^{\infty} \tilde{\lambda}_j z^j = \tilde{\Lambda}(z)$$

avec $\sum_{j=0}^{\infty} |\tilde{\lambda}_j| < \infty$ **si les $|z|$ ne sont pas sur le cercle unitaire**
(développement en séries de Laurent au voisinage du cercle unitaire)

- On peut alors définir $\tilde{\Lambda}(L)$ comme un filtre linéaire et écrire

$$X_t = \Phi(L)^{-1}\Theta(L)\varepsilon_t = \tilde{\Lambda}(L)\Theta(L)\varepsilon_t = \Psi(L)\varepsilon_t$$

ce qui démontre l'existence d'une solution stationnaire causale (unique)

Stationnarité des ARMA(p, q)

Definition (16)

X_t est un ARMA(p, q) et admet une unique solution stationnaire causale si

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

où $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ et s'il existe $\phi_j, j = 1, 2, \dots, p$ tel que

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0, \quad \forall |z| \leq 1$$

et si les polynômes $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et

$$\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

n'ont pas de racines communes (condition d'identification)

Inversibilité des ARMA(p, q)

Definition (17)

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-1}$$

est un ARMA(p, q) inversible s'il existe $\theta_j, j = 1, 2, \dots, q$ tel que

$$\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q \neq 0, \quad \forall |z| \leq 1$$

de sorte que X_t admet la représentation

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

et si les polynômes $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et

$$\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

n'ont pas de racines communes (condition d'identification)

Variance des ARMA(p, q) stationnaires inversibles

- D'après le Th. 2 on sait que pour $X_t = \Psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$

$$\gamma_X(h) = \mathbb{E}(X_{t+h}X_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}$$

- Considérons le cas d'un ARMA(1, 1) : $X_t - \phi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$

$$\begin{aligned} \gamma_X(0) &= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2 \\ &= \mathbb{V}\left(\varepsilon_t + (\phi_1 + \theta_1) \sum_{j=1}^{\infty} \phi_1^{j-1} \varepsilon_{t-j}\right) \\ &= \sigma_\varepsilon^2 + \sigma_\varepsilon^2 (\phi_1 + \theta_1)^2 \sum_{j=1}^{\infty} \phi_1^{2(j-1)} \\ &= \sigma_\varepsilon^2 \left(1 + (\phi_1 + \theta_1)^2 \sum_{j=0}^{\infty} \phi_1^{2j}\right) \\ &= \sigma_\varepsilon^2 \left(1 + \frac{(\phi_1 + \theta_1)^2}{1 - \phi_1^2}\right) \text{ car } \sum_{j=0}^{\infty} \phi_1^{2j} = 1/(1 - \phi_1^2) \text{ si } |\phi_1| < 1 \end{aligned}$$

ACovF des ARMA(p, q) stationnaires inversibles

- Toujours dans le cas d'un ARMA(1, 1) : $X_t - \phi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$

$$\begin{aligned}\gamma_X(1) &= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j-1} \\ &= \sigma_\varepsilon^2 \left(\theta_1 + \phi_1 + (\phi_1 + \theta_1)^2 \phi_1 \sum_{j=0}^{\infty} \phi_1^{2j} \right) \\ &= \sigma_\varepsilon^2 \left(\theta_1 + \phi_1 + \frac{(\phi_1 + \theta_1)^2 \phi_1}{1 - \phi_1^2} \right) \\ \gamma_X(h) &= \phi_1^{h-1} \gamma_X(1), \quad h > 1\end{aligned}$$

- Il en découle les autocorrélations

$$\begin{aligned}\rho_0 &= 1 \\ \rho_1 &= \frac{(\theta_1 + \phi_1)(1 + \theta_1 \phi_1)}{1 + 2\theta_1 \phi_1 + \theta_1^2} \\ \rho_h &= \phi_1 \rho_{h-1} = \phi_1^{h-1} \rho_1\end{aligned}$$

Comportement des ACF et PACF des ARMA(p, q)

- Si X_t suit un MA(q) inversible
 - $\text{ACF}(h) = 0$ si $h > q$
 - $\text{PACF}(h) \neq 0$ mais $\text{PACF}(h) \rightarrow 0$ quand $h \rightarrow \infty$
- Si X_t suit un AR(p) stationnaire
 - $\text{ACF}(h) \neq 0$ mais $\text{ACF}(h) \rightarrow 0$ quand $h \rightarrow \infty$
 - $\text{PACF}(h) = 0$ si $h > p$
- Si X_t suit un ARMA(p, q) stationnaire et inversible
 - $\text{ACF}(h) \neq 0$ mais $\text{ACF}(h) \rightarrow 0$ quand $h \rightarrow \infty$
 - $\text{PACF}(h) \neq 0$ mais $\text{PACF}(h) \rightarrow 0$ quand $h \rightarrow \infty$

Analyse des ACFs

- L'ACF d'un ARMA(p, q) peut prendre des **formes très diverses**
- Considérons le cas simple d'un AR(2) : $(1 - z_1^{-1}L)(1 - z_2^{-1}L)X_t = \varepsilon_t$
- On suppose $|z_1| > 1$, $|z_2| > 1$ et le lien avec ϕ_1 et ϕ_2 est donné S58

$$\phi_1 = z_1^{-1} + z_2^{-1} \text{ et } \phi_2 = -z_1^{-1}z_2^{-1}$$

- Un calcul compliqué (non demandé) permet d'obtenir l'ACovF

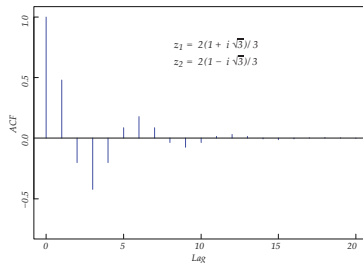
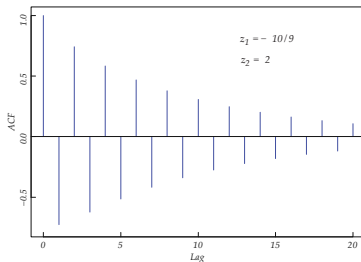
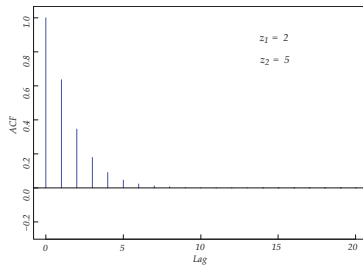
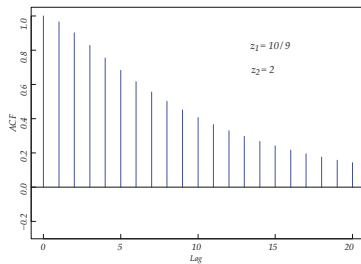
$$\gamma(h) = \frac{\sigma_\varepsilon^2 z_1^2 z_2^2}{(z_1 z_2 - 1)(z_2 - z_1)} \left((z_1^2 - 1)^{-1} z_1^{1-h} - (z_2^2 - 1)^{-1} z_2^{1-h} \right)$$

- Si $z \in \mathbb{C}$, $z_1 = r e^{i\vartheta}$, $z_2 = r e^{-i\vartheta}$ et on peut alors raffiner le calcul

$$\gamma(h) = \sigma_\varepsilon^2 \frac{r^4 r^{-h} \sin(h\theta + \psi)}{(r^2 - 1)(r^4 - 2r^2 \cos(2\theta) + 1) \sin \theta}$$

- Nous verrons au C2 comment l'analyse graphique peut aiguiller le choix de modélisation

Analyse graphique des ACFs



Plan

- 1 Introduction
- 2 Généralités
- 3 AR et MA
- 4 Stationnarité et inversibilité
- 5 Les processus ARMA
- 6 La prévision des ARMA

Prédictions linéaires

- Soit X_t un processus stationnaire causal avec $\mu_X = 0$ et $\gamma_X(h)$ connue
- Comme au S26, on cherche un prédicteur linéaire $P_n(X_{n+h})$ de

$$X_{n+h} | X_n, \dots, X_1, \quad h > 0$$

\Rightarrow $P_n(X_{n+h})$ sera de la forme $P_n(X_{n+h}) = \underbrace{\alpha_1}_{\text{orange}} X_n + \dots + \underbrace{\alpha_n}_{\text{bleu}} X_1$

- On cherche donc les $\alpha_1, \dots, \alpha_n$ qui minimisent

$$\sigma_\epsilon^2 = \mathbb{E}\left((X_{n+h} - P_n(X_{n+h}))^2\right), \quad \text{où } \epsilon = X_{n+h} - P_n(X_{n+h})$$

i.e., les $\alpha_0, \dots, \alpha_n$ qui minimise l'erreur quadratique moyenne (MSE)

- La solution est donnée par $\mathbb{E}(X_{n+h} - \sum_{i=1}^n \alpha_i X_{n+1-i}) = 0$ ou

$$\mathbb{E}((X_{n+h} - \sum_{i=1}^n \alpha_i X_{n+1-i}) X_{n+1-j}) = 0, \quad j = 1, \dots, n \quad (2)$$

de manière équivalente car un bon prédicteur ne laisse pas d'information inutilisée

$$\mathbb{E}(\epsilon \times \text{Variables utilisées pour prédire}) = 0$$

Prédictions linéaires et autocovariance

- On peut réécrire $P_n(X_{n+h}) = \sum_{i=1}^n \alpha_i X_{n+1-i}$ vectoriellement

$$P_n(X_{n+h}) = \boldsymbol{\alpha}'_n \mathbf{X}_n, \quad \boldsymbol{\alpha}_n = (\alpha_1, \dots, \alpha_n)'$$

avec $\mathbf{X}_n = (X_n, \dots, X_1)$

- Une réécriture vectorielle de l'Eq. (2) est aussi possible via

$$\Gamma_n \boldsymbol{\alpha}_n = \boldsymbol{\gamma}_n(h), \quad \boldsymbol{\gamma}_n(h) = (\gamma_X(h), \gamma_X(h+1), \dots, \gamma_X(h+n-1))'$$

et $\Gamma_n = [\gamma_X(i-j)]_{i,j=1}^n$ avec intérêt que

$$\begin{aligned} \sigma_\epsilon^2 &= \mathbb{E}\left((X_{n+h} - P_n(X_{n+h}))^2\right) \\ &= \gamma_X(0) - 2 \sum_{i=1}^n \alpha_i \gamma_X(h+i-1) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \gamma_X(i-j) \alpha_j \\ &= \gamma_X(0) - \boldsymbol{\alpha}'_n \boldsymbol{\gamma}_n(h) \end{aligned}$$

et donc la variance de l'erreur de prédiction est simple à obtenir

Note Si $\mu_X \neq 0$, on peut le retirer du modèle puis l'ajouter à $P_n(\cdot)$ car la structure linéaire reste inchangée

Prédictions d'un AR(1) et Yule-Walker : exemple

- Soit $X_t = \phi_1 X_{t-1} + \varepsilon_t$, avec $|\phi_1| < 1$ et X_1, \dots, X_n
- On souhaite prédire en X_{n+1} et l'on sait à présent que

$$P_n(X_{n+1}) = \alpha'_n \mathbf{X}_n$$

- Une application direct de $\Gamma_n \alpha_n = \gamma_n(h)$ nous donne

$$\begin{pmatrix} 1 & \phi_1 & \phi_1^2 & \cdots & \phi_1^{n-1} \\ \phi_1 & 1 & \phi_1 & \cdots & \phi_1^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_1^{n-1} & \phi_1^{n-2} & \phi_1^{n-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_1^2 \\ \vdots \\ \phi_1^n \end{pmatrix}$$

d'après YW (S46) et sachant que $\gamma(0)$ étant des deux côtés il disparaît

- On voit alors que $\alpha_n = (\phi_1, 0, \dots, 0)'$ est une **solution évidente**

\Rightarrow le meilleur prédicteur linéaire est $P_n(X_{n+1}) = \phi_1 X_n$

$$\Rightarrow \sigma_\varepsilon^2 = \gamma(0) - \alpha'_n \gamma_n(1) = \gamma(0)(1 - \alpha'_n \rho_n(1)) = \sigma_\varepsilon^2 (1 - \phi_1^2)^{-1} (1 - \phi_1^2) = \sigma_\varepsilon^2$$

Prédictions d'un AR(p)

- Soit un AR(p) stationnaire

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

- Vous devez pouvoir montrer via l'Eq.(2) que

$$P_n(X_{n+1}) = \phi_1 X_n + \dots + \phi_p X_{n+1-p}$$

est le meilleur prédicteur linéaire de X_{n+1}

- Pour cela il suffit de remplacer X_{n+h} par son expression dans

$$\mathbb{E}(X_{n+h} - \sum_{i=1}^n \alpha_i X_{n+1-i})$$

et de déterminer la séquence α_n solution de

$$\mathbb{E}(X_{n+h} - \sum_{i=1}^n \alpha_i X_{n+1-i}) = 0$$

Astuce On peut passer par $\Gamma_n \alpha_n = \gamma_n(h)$ puisqu'en multipliant par X_{t-j} de chaque côté et en prenant l'espérance on obtient $\Gamma_p \phi_p = \gamma_p(h)$ et donc

$$\alpha_n = \phi_p$$

Prédictions de variables aléatoires quelconques

- Les propriétés de l'opérateur $P_n(.|.)$ sont préservées pour

$$Y \text{ et } \mathbf{W} = (W_n, \dots, W_1)$$

des variables aléatoires quelconques

$\Rightarrow P_n(Y|\mathbf{W})$ est toute solution de

$$\Gamma \boldsymbol{\alpha} = \boldsymbol{\gamma}$$

avec $\boldsymbol{\gamma} = Cov(Y, \mathbf{W})$ et $\Gamma = Cov(\mathbf{W}, \mathbf{W})$

- Ces formules sont donc générales et s'appliquent
 - aux valeurs manquantes
 - aux modèles ARMA
 - à tout autre modèle linéaire

Estimation d'une valeur manquante : exemple

- Soit $X_t = \phi_1 X_{t-1} + \varepsilon_t$, avec $|\phi_1| < 1$ et X_1, \dots, X_n
- On suppose que $X_2 = Y$ est une valeur manquante à “prédire”
- La structure AR(1) nous indique que $\mathbf{W} = (W_1, W_3)$ et donc

$$\Gamma = Cov(W_1, W_3) = \begin{pmatrix} 1 & \phi_1^2 \\ \phi_1^2 & 1 \end{pmatrix}$$

car la dépendance entre X_1 et X_3 n'est pas directe alors que

$$\boldsymbol{\gamma} = (\phi_1, \phi_1)'$$

car la dépendance entre X_1 et X_2 puis X_3 et X_2 , est directe

\Rightarrow La solution de $\Gamma \boldsymbol{\alpha} = \boldsymbol{\gamma}$ est alors

$$\boldsymbol{\alpha} = (1 + \phi^2)^{-1} \begin{pmatrix} \phi_1 \\ \phi_1 \end{pmatrix}$$

et le meilleur prédicteur linéaire de X_2 est donc

$$P_n(X_2|\mathbf{W}) = \boldsymbol{\alpha}' \mathbf{W} = \phi(1 + \phi^2)^{-1}(X_1 + X_3)$$

Prédictions par récursion : Durbin-Levinson

- En théorie, $\forall h, \Gamma_n \alpha_n = \gamma_n(h)$ permet d'obtenir $P_n(X_{n+h})$
 - En pratique, c'est numériquement lourd pour n grand et $h > 1...$
- \Rightarrow Lier $P_n(X_{n+1})$ à $P_{n+1}(X_{n+2}), \dots, P_{n+h-1}(X_{n+h})$ serait donc utile
- Pour $h = 1$ on sait que $\Gamma_n^{-1} \gamma_n = \phi_n$ donne

$$P_n(X_{n+1}) = \phi_n' \mathbf{X}_n = \phi_{n,1} X_n + \dots + \phi_{n,n} X_1$$

avec $\gamma_n = (\gamma(1), \dots, \gamma(n))'$ et $\sigma_{\epsilon,n}^2 = \gamma(0) - \phi_n' \gamma_n$

Solution L'algorithme de Durbin-Levinson (ADL) se base sur

$$\phi_{n,n} = \left(\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right) \sigma_{\epsilon,n-1}^{-2}$$

$$\begin{pmatrix} \phi_{n,1} \\ \phi_{n,2} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \phi_{n-1,2} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{n,n} \begin{pmatrix} \phi_{n,1} \\ \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}$$

et $\sigma_{\epsilon,n}^2 = \sigma_{\epsilon,n-1}^2 \times (1 - \phi_{n,n}^2)$, où $\phi_{1,1} = \gamma(1)/\gamma(0)$ et $\sigma_{\epsilon,0}^2 = \gamma(0)$

Prédictions par récursion : les erreurs de prédiction in-sample (innovations)

- Une alternative intéressante à Durbin-Levinson se base sur

$$\epsilon_n = X_n - P_{n-1}(X_n) \equiv X_n - \widehat{X}_n$$

- Sous forme matricielle, les innovations $\epsilon_n = (\epsilon_1, \dots, \epsilon_n)'$ s'écrivent

$$\epsilon_n = A_n \mathbf{X}_n, \quad A_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\alpha_1 & 1 & 0 & \dots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ -\alpha_{n-1} & -\alpha_{n-2} & -\alpha_{n-3} & \dots & 1 \end{pmatrix}$$

et donc $\widehat{\mathbf{X}}_n := (X_1, P_1(X_2), \dots, P_{n-1}(X_n))'$ peut s'écrire

$$\widehat{\mathbf{X}}_n = \mathbf{X}_n - \epsilon_n = A_n^{-1} \epsilon_n - \epsilon_n = \Theta_n (\mathbf{X}_n - \widehat{\mathbf{X}}_n)$$

avec $\Theta_n = (A_n^{-1} - I_n)$, où I_n est une matrice identité et

$$\Theta_n = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \theta_{1,1} & 0 & 0 & \dots & 0 \\ \theta_{2,2} & \theta_{2,1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \dots & 0 \end{pmatrix}$$

Prédictions par récursion : l'algorithme des innovations

Solution L'algorithme des innovations (AI)

- Les coefficients $\theta_{n,1}, \dots, \theta_{n,n}$ s'obtiennent aussi récursivement

$$\theta_{n,n-k} = \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j \right) \nu_k^{-1}$$

avec $\mathbb{E}(X_i X_j) = \kappa(i, j)$, $\nu_0 = \kappa(1, 1)$, et

$$\nu_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j$$

\Rightarrow on commence par ν_0 puis on résout pour θ_{11} , ν_1 , puis θ_{22} , θ_{21} , ν_2 , etc.

Note 1 ADL repose sur la représentation $\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$

Note 2 AI repose sur $\hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j})$

Prédictions par récursion à l'horizon h

- L'approche récursive permet d'itérer au-delà de $h = 1$: $P_{n+h-1}(X_{n+h})$
- On utilise le fait que l'erreur de prévision ϵ_{n+h} , $h \geq 1$, est **imprévisible**

$$\begin{aligned}
 P_n(X_{n+h} - P_{n+h-1}(X_{n+h})) &= 0 \\
 \Rightarrow P_n(X_{n+h}) &= P_n(P_{n+h-1}(X_{n+h})) \\
 &= P_n(\hat{X}_{n+h}) \\
 &= P_n\left(\sum_{j=1}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j})\right)
 \end{aligned}$$

car $P_n(\cdot)$ est un opérateur linéaire, et on obtient finalement

$$P_n(X_{n+h}) = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j})$$

car si $j < h$, $\epsilon_{n+h-j} \equiv$ erreurs de prévisions futures

Note Pour $j > h$, $\epsilon_{n+h-j} \equiv$ erreurs de prévisions observables et

$$\nu_n = \mathbb{E}(X_{n+h} - P_{n+h-1}(X_{n+h}))^2 = \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 \nu_{n+h-j-1}$$

Prédictions par récursion des ARMA(p, q)

- Soit un $\Phi(L)X_t = \Theta(L)\varepsilon_t$ un ARMA(p, q) stationnaire inversible

- Ansley (1979) a démontré une écriture générale pour \hat{X}_{n+1}

$$\hat{X}_{n+1} = \begin{cases} \frac{\sum_{j=1}^n \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j})}{\sum_{j=1}^p \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j})} & 1 \leq n < m \\ & n \geq m \end{cases} \quad (3)$$

avec $m = \max(p, q)$ et $\theta_{n,j} \rightarrow \theta_j$ pour $j = 1, \dots, q$ si $n \rightarrow \infty$

- L'approche d'Ansley (1979) repose sur des ACovF de

$$W_t = \sigma_\varepsilon^{-1} \Phi(L)X_t \text{ si } t > m \text{ et } W_t = \sigma_\varepsilon^{-1} X_t \text{ sinon}$$

- Concernant la MSE, on peut montrer que

$$\sigma_\varepsilon^2 = \mathbb{E}(X_{n+1} - \hat{X}_{n+1})^2 = \sigma_\varepsilon^2 \mathbb{E}(W_{n+1} - \widehat{W}_{n+1})^2 = \sigma_\varepsilon^2 r_n$$

où $r_n \rightarrow 1$ si $n \rightarrow \infty$

Prédictions d'un $AR(p)$ et d'un $MA(q)$

\Rightarrow Pour un $AR(p)$, si $n \geq p$, d'après Eq.(3), on a directement que

$$\hat{X}_{n+1} = \sum_{j=1}^p \phi_j X_{n+1-j}$$

\Rightarrow Pour un $MA(q) \equiv ARMA(1, q)$ avec $\phi_1 = 0$, si $n \geq 1$, d'après Eq.(3), on a directement que

$$\hat{X}_{n+1} = \sum_{j=1}^{\min(n, q)} \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j})$$

où les coefficients $\theta_{n,j} \rightarrow \theta_j$ pour $j = 1, \dots, q$ si $n \rightarrow \infty$

Prédictions d'un ARMA(1, 1)

- Appliquons l'approche d'Ansley (1979) à un ARMA(1, 1)

$$X_t - \phi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- Si $|\phi_1| < 1$ et $|\theta_1| < 1$ et $n \geq 1$, d'après Eq.(3) on a

$$\hat{X}_{n+1} = \phi_1 X_n + \theta_{n,1} (X_n - \hat{X}_n)$$

- Pour $q = 1$, le calcul de $\theta_{n,1}$ est possible via le S68 qui donne

$$\gamma_X(0) = \sigma_\varepsilon^2 \left(1 + \frac{(\phi_1 + \theta_1)^2}{1 - \phi_1^2} \right) = \sigma_\varepsilon^2 (1 - \phi_1^2)^{-1} (1 + 2\theta_1 \phi_1 + \theta_1^2)$$

et en découle le calcul des

$$\kappa_W(i, j) = \begin{cases} (1 - \phi_1^2)^{-1} (1 + 2\theta_1 \phi_1 + \theta_1^2) & i = j = 1 \\ 1 + \theta_1^2 & i = j \geq 2 \\ \theta_1 & |i - j| = 1, i \geq 1 \\ 0 & \text{sinon} \end{cases}$$

- On obtient $r_0 = (1 - \phi_1^2)^{-1} (1 + 2\theta_1 \phi_1 + \theta_1^2)$, $r_n = 1 + \theta_1^2 - r_{n-1}^{-1} \theta_1^2$ et

$$\theta_{n1} = r_{n-1}^{-1} \theta_1$$

par application de l'AI où r_n remplace ν_n

Prédictions par récursion à l'horizon h d'un ARMA(p, q)

- L'approche d'Ansley (1979) permet une écriture simple pour $P_n(X_{n+h})$

$$P_n(X_{n+h}) = \sum_{i=1}^p \phi_i P_n(X_{n+h-i}) + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

si $h > m - n$

- Si $n > m = \max(p, q)$, comme souvent, alors $\forall h \geq 1$..

$$P_n(X_{n+h}) = \sum_{i=1}^p \phi_i P_n(X_{n+h-i}) + \sum_{j=h}^q \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

Prédictions dans un cas asymptotique : $n \rightarrow \infty$

- Dans les équations d'Ansley (1979), $n \rightarrow \infty$ simplifie les choses car

$$\hat{X}_{n+h} = \sum_{j=1}^p \phi_j X_{n+h-j} + \sum_{j=h}^q \theta_j (X_{n+h-j} - \hat{X}_{n+h-j})$$

e.g. 1 Pour un MA(2), $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$, on a directement que

$$\hat{X}_{n+1} = \sum_{j=1}^q \theta_j (X_{n+1-j} - \hat{X}_{n+1-j}) = \theta_1 \varepsilon_n + \theta_2 \varepsilon_{n-1}$$

$$\hat{X}_{n+2} = \theta_2 \varepsilon_n$$

$$\hat{X}_{n+h} = 0, \text{ pour } h \geq 3$$

e.g. 2 Pour un ARMA(1,2), $X_t = \phi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$, on a

$$\hat{X}_{n+1} = \phi_1 X_n + \theta_1 \varepsilon_n + \theta_2 \varepsilon_{n-1}$$

$$\hat{X}_{n+2} = \phi_1 \hat{X}_{n+1} + \theta_2 \varepsilon_n$$

$$\hat{X}_{n+h} = \phi_1 \hat{X}_{n+h-1}, \text{ pour } h \geq 3$$

Prédictions et Inférence

- Soit un $\text{ARMA}(p, q)$ stationnaire causal

$$\Phi(L)X_t = \Theta(L)\varepsilon_t$$

dont les prévisions à l'ordre $h > 0$ sont \hat{X}_{n+h}

- On sait que les erreurs de prévisions

$$\epsilon = X_{n+h} - \hat{X}_{n+h}$$

sont d'espérance nulle et de variance

$$\sigma_\epsilon^2 = \gamma(0) - \boldsymbol{\alpha}'_n \boldsymbol{\gamma}_n(h)$$

⇒ Sous une hypothèse de loi, on peut construire **l'intervalle de confiance**

$$\hat{X}_{n+h} \pm \Phi_{1-\alpha/2} \sigma_\epsilon$$

pour un seuil de risque α

- Si ε_t est supposé Gaussien, X_{n+h} sera aussi Gaussien,

$$\frac{X_{n+h} - \hat{X}_{n+h}}{\sigma_\epsilon} \xrightarrow{d} \mathcal{N}(0, 1)$$

et $\Phi_{1-\alpha/2}$ représentera le fractile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, 66(1), 59-65.