



MASTER
Econométrie et Statistique Appliquée
Université d'Orléans

Econométrie des Séries Temporelles Univariées

Gilles de Truchis

Master 1 ESA

January 21, 2025



Table of contents

Introduction

AR et MA

Stationnarité et inversibilité

Les processus ARMA

La prévision des ARMA

Introduction à l'estimation

MLE

Sélection de modèle

ARIMAX

Tests de validation

Faits stylisés et rappels

DS versus TS

Intégration fractionnaire

ARMA Saisonniers

Théorie limite standard

Théorie limite non-standard

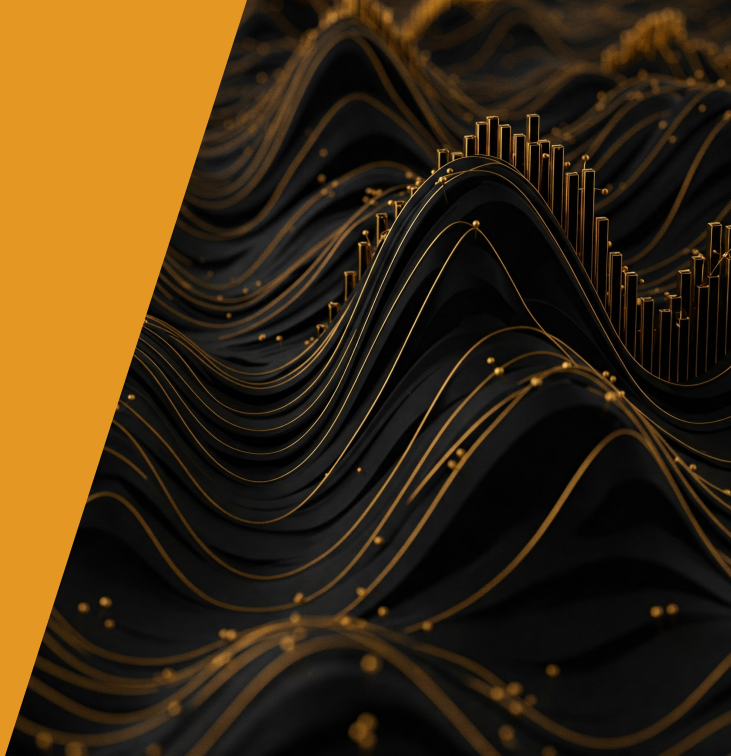
Tests de racine unitaire

Les régressions factices

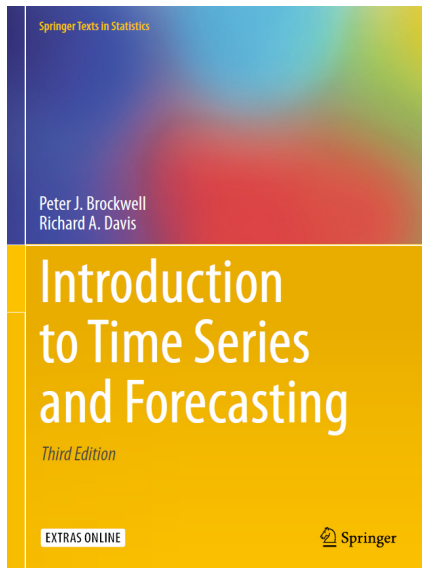
Conclusion

Chapter 1

Les Processus ARMA



Références



Les séries temporelles

Definition 1 (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 (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 (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

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 (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
 - ⇒ le nombre de paramètre à estimer peut dépasser le nombre de données

⇒ 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. S25)



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 (5)

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

$$X_t \text{ satisfait } \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \dots \mathbb{P}(X_n \leq x_n) = F(x_1) \dots F(x_n)$$

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

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

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 (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 (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 (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 (9)

X_t est **faiblement stationnaire** si

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

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

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

- Ici on cherche à résumer la stabilité en loi de X_t seulement à travers ses 2 premiers moments
⇒ 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 (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 (11)

On définit comme coefficients de mélange

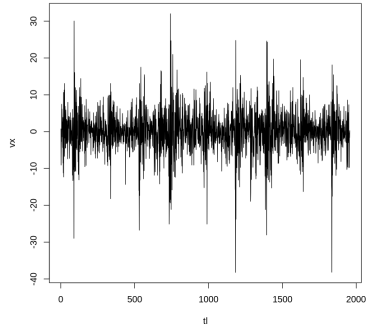
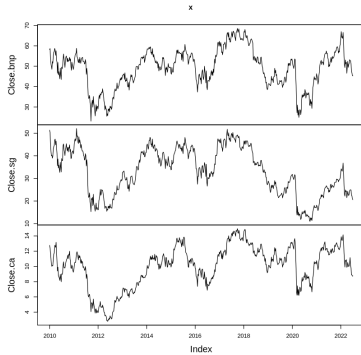
$$\alpha(\ell) = \sup_{A \in \mathcal{F}_{-\infty}^{t-\ell}, B \in \mathcal{F}_t^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

*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 (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 $WN(\cdot)$, sont des processus stationnaires

Definition 13 (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)

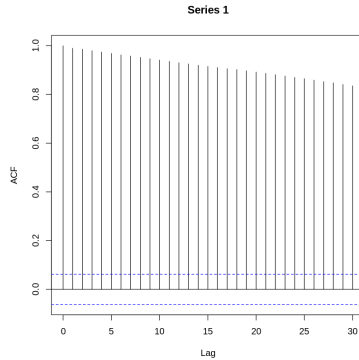
Definition 14 (14)

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 : $X_t \sim \mathcal{N}(\mu_X, \sigma_X^2)$
- $\rho_X(\mathbf{h})$ 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

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

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

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

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

- 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 $m(Y) = \ell(Y) = aY + b$

⇒ 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 (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 (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 (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 (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

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

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

Moyenne mobile d'ordre q : MA(q)

- La représentation MA(∞) 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(∞) 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

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

⇒ 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

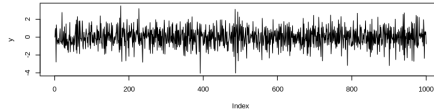
- Pour **l'ACF** on obtient donc

$$\rho_0 = 1$$

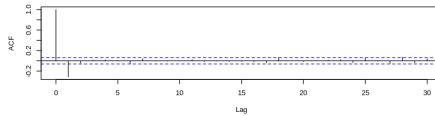
$$\rho_X(1) = \frac{\gamma_X(1)}{\gamma_X(0)} = \frac{\theta_1}{1 + \theta_1^2}$$

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

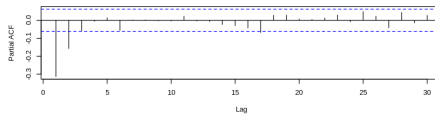
Analyse graphique d'un MA(1) : $X_t = \varepsilon_t - 0.4\varepsilon_{t-1}$



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 $|\theta| < 1$

⇒ 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 $X_t = (1 - \theta_1) \varepsilon_t \dots$

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

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

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 $AR(\infty)$

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

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

⇒ 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 $X_t = \phi_1^{-1} X_{t+1} - \phi_1^{-1} \varepsilon_{t+1}$

$$X_t = -\phi_1^{-1} \varepsilon_{t+1} - \phi_1^{-2} \varepsilon_{t+2} + \phi_1^{-2} X_{t+2}$$

= ...

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

⇒ 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 $\gamma_X(0), \gamma_X(1), \dots, \gamma_X(h)$

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

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

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

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

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

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

\Rightarrow Pour $\forall h \geq 0$, 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

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

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

⋮

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

- 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

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

- En inversant le système on obtient alors

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

Yule-Walker et autocorrélations partielles (PACF)

- Comme précisé au S97, l'interprétation de ρ_j peut-être ambiguë
- Si l'on veut connaître la PACF d'ordre j dans notre AR(p) on considère

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

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 $\phi_{j,j}$

⇒ Pour $j = 1$ on a

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

⇒ 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 $h > 2$

$$\rho_0 = 1$$

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

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

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

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

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

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

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

⇒ 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 S38 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

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

- 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

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 S41 (se voit aussi avec une division euclidienne)

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

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

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

\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)

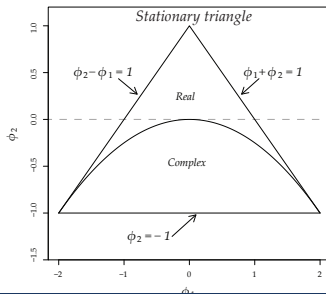
- Pour $p > 1$, les racines peuvent être complexes $z \in \mathbb{C}$
- Étudions le cas d'un AR(2), $(1 - \phi_1 L - \phi_2 L^2)X_t = \varepsilon_t$, de polynôme

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

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

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

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 S38

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

- 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**

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

⇒ 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

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

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

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

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 S41, $\Phi(L)^{-1}$ 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)

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

- 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 S54, $\Phi(L)^{-1}$ 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 (16)

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

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

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 (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
 - $ACF(h) = 0$ si $h > q$
 - $PACF(h) \neq 0$ mais $PACF(h) \rightarrow 0$ quand $h \rightarrow \infty$
- Si X_t suit un AR(p) stationnaire
 - $ACF(h) \neq 0$ mais $ACF(h) \rightarrow 0$ quand $h \rightarrow \infty$
 - $PACF(h) = 0$ si $h > p$
- Si X_t suit un ARMA(p, q) stationnaire et inversible
 - $ACF(h) \neq 0$ mais $ACF(h) \rightarrow 0$ quand $h \rightarrow \infty$
 - $PACF(h) \neq 0$ mais $PACF(h) \rightarrow 0$ 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 S55

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

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

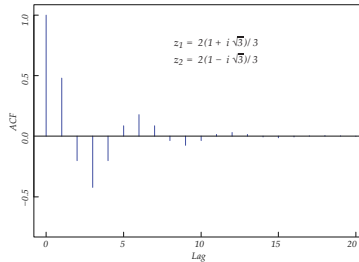
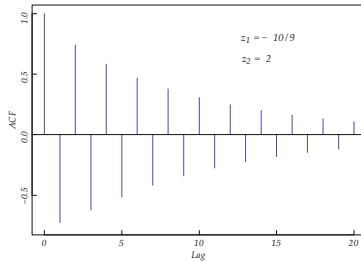
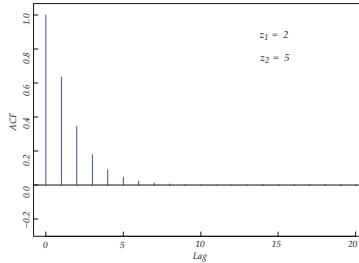
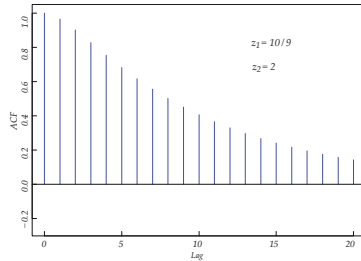
$$\gamma(h) = \frac{\sigma_\varepsilon^2 z_1^2 z_2^2}{(z_1 z_2 - 1)(z_2 - z_1)} \left((z_1^2 - 1)^{-1} z_1^{1-h} - (z_2^2 - 1)^{-1} z_2^{1-h} \right)$$

- Si $z \in \mathbb{C}$, $z_1 = re^{i\theta}$, $z_2 = re^{-i\theta}$ 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



Prédictions linéaires

- Soit X_t un processus stationnaire causal avec $\mu_X = 0$ et $\gamma_X(h)$ connue
- Comme au S25, on cherche un prédicteur linéaire $P_n(X_{n+h})$ de

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

⇒ $P_n(X_{n+h})$ sera de la forme $P_n(X_{n+h}) = \alpha_1 X_n + \dots + \alpha_n X_1$

- 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

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

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

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

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

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

$$\Gamma_n \alpha_n = \gamma_n(\mathbf{h}), \quad \gamma_n(\mathbf{h}) = (\gamma_X(\mathbf{h}), \gamma_X(\mathbf{h} + 1), \dots, \gamma_X(\mathbf{h} + n - 1))'$$

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

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

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

$$P_n(X_{n+1}) = \alpha_n' X_n$$

- 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 (S44) 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**

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

⇒ $\sigma_\varepsilon^2 = \gamma(0) - \alpha_n' \gamma_n(1) = \gamma(0)(1 - \alpha_n' \rho_n(1)) = \sigma_\varepsilon^2 (1 - \phi_1^2)^{-1} (1 - \phi_1^2) = \sigma_\varepsilon^2$

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 $P_n(\cdot|\cdot)$ sont préservées pour

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

des variables aléatoires quelconques

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

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

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

- 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

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

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

⇒ 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 $h > 1$...
- ⇒ 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

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

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

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

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

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

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

⇒ on commence par ν_0 puis on résout pour θ_{11}, ν_1 , puis $\theta_{22}, \theta_{21}, \nu_2$, etc.

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

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

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**

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

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

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

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

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

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 \widehat{X}_{n+1}

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

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

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

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

Prédictions d'un AR(p) et d'un MA(q)

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

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

⇒ 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

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

- Pour $q = 1$, le calcul de $\theta_{n,1}$ est possible via le S64 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 $P_n(X_{n+h})$

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

si $h > m - n$

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

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

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

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

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

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

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

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

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

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

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

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

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

Prédictions et Inférence

- Soit un ARMA(p, q) stationnaire causal

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

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

- On sait que les erreurs de prévisions

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

sont d'espérance nulle et de variance

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

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

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

pour un seuil de risque α

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

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

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

Chapter 2

Estimation et Sélection de Modèles



Estimation des modèles linéaires

Objectif Identifier un modèle linéaire approprié pour une série $\{\tilde{X}_t\}_{t=1}^n$

- Cela implique de résoudre plusieurs problèmes itérativement
 - estimer μ_X (afin de travailler avec $X_t = \tilde{X}_t - \mu_X$)
 - estimer les coefficients du modèle (dans la classe des $\text{ARMA}(p, q)$)
 - sélectionner l'ordre des retards optimaux p et q
 - estimer la variance du bruit blanc σ_ε^2
- Dans un second temps, le modèle sélectionné devra être
 - soumis à des tests de diagnostic (e.g. sphéricité des erreurs)
 - être utilisé pour de la prédiction

Estimation de μ_X

- L'estimateur de la moyenne μ_X d'un processus stationnaire X_t est

$$\bar{X}_n = n^{-1} \sum_{j=1}^n X_j$$

- \bar{X}_n , la moyenne empirique, est un estimateur non-biaisé car

$$\mathbb{E}(\bar{X}_n) = n^{-1}(\mathbb{E}X_1 + \dots + \mathbb{E}X_n) = \mu_X$$

- La MSE de \bar{X}_n est donnée par

$$\begin{aligned} \mathbb{E}(\bar{X}_n - \mu_X)^2 &= \mathbb{V}(\bar{X}_n) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = n^{-2} \sum_{i-j=-n}^n (n - |i-j|) \gamma_X(i-j) \\ &= n^{-1} \sum_{h=-n}^n (1 - n^{-1}|h|) \gamma_X(h) \end{aligned}$$

- Si $X_t \sim \text{ARMA}(p, q)$, $\gamma_X(h)$ est absolument sommable et

$$\lim_{n \rightarrow \infty} \mathbb{V}(\bar{X}_n) = n^{-1} \sum_{h=-\infty}^{\infty} |\gamma_X(h)|$$

Inférence autour de μ_X

- Sous certaines conditions on peut montrer que

$$n^{1/2}(\bar{X}_n - \mu_X) \sim \mathcal{N}\left(0, \sum_{|h|<n} (1 - n^{-1}|h|)\gamma(h)\right)$$

e.g. Si X_t est Gaussien ou linéaire ce résultat est valide

- Une approximation de l'intervalle de confiance (IC) à 95% est alors

$$\left(\bar{X}_n - 1.96 \frac{\nu^{1/2}}{n^{1/2}}, \bar{X}_n + 1.96 \frac{\nu^{1/2}}{n^{1/2}}\right)$$

où $\nu = \sum_{h=-\infty}^{\infty} \gamma_X(h)$, généralement inconnu, devra être estimé aussi

- Considérons un exemple : soit un AR(1) avec $|\phi_1| < 1$,

$$X_t - \mu_X = \phi_1(X_{t-1} - \mu_X) + \varepsilon_t, \quad \gamma_X(h) = \phi_1^{|h|} \sigma_\varepsilon^2 (1 - \phi_1^2)^{-1}$$

et donc $\nu = (1 + 2 \sum_{h=1}^{\infty} \phi^h) \sigma_\varepsilon^2 (1 - \phi_1^2)^{-1} = \sigma_\varepsilon^2 (1 - \phi_1)^{-2}$

- On peut alors construire l'IC à 95% pour μ_X :

$$\bar{x}_n \pm 1.96 \sigma n^{-1/2} (1 - \phi_1)^{-1}$$

Estimateurs de $\gamma_X(\cdot)$ et $\rho_X(\cdot)$

- Les estimateurs de l'ACovF et l'ACF sont données par

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)$$

$$\text{et } \hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$$

- Ces estimateurs sont biaisés mais si on replace n^{-1} par

$$(n-h)^{-1}$$

ils sont presque pas biaisés pour n grand

- On peut montrer que pour $\boldsymbol{\rho} = (\rho(1), \dots, \rho(k))'$ les ACF empiriques sont approximativement Normales

$$\hat{\boldsymbol{\rho}} \xrightarrow{a.a.d} \mathcal{N}(\boldsymbol{\rho}, n^{-1}W)$$

avec W une matrice de covariance dont les coefficients sont donnés par la formule de Bartlett

$$w_{ij} = \sum_{k=1}^{\infty} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)) \\ \times (\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k))$$

Inférence autour de $\gamma_X(\cdot)$ et $\rho_X(\cdot)$

- Prenons l'exemple d'un AR(1) et comparons-le avec l'ACF empirique

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

avec $|\phi_1| < 1$ de sorte que nous savons que $\rho_X(h) = \phi_1^{|h|}$ et donc

$$\begin{aligned} w_{ii} &= \sum_{k=1}^i \phi_1^{2i} (\phi_1^{-k} - \phi_1^k)^2 + \sum_{k=i+1}^{\infty} \phi_1^{2k} (\phi_1^{-i} - \phi_1^i)^2 \\ &= (1 - \phi_1^{2i})(1 + \phi_1^2)(1 - \phi_1^2)^{-1} - 2i\phi_1^{2i} \end{aligned}$$

- Considérons le niveau annuel en "pieds" du Lac Huron (1875-1972)
- Supposons que l'estimation du modèle AR(1) donne

$$x_t - \bar{x} = 0.791(x_{t-1} - \bar{x}) + \varepsilon_t$$

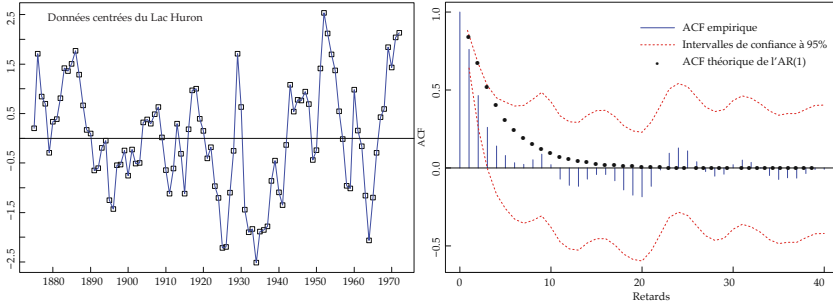
- On peut alors calculer $\hat{\rho}(h)$ et le comparer avec

$$\rho_X(h) = \phi_1^{|h|}$$

les intervalles de confiance à 95% étant donnés par

$$\hat{\rho}(h) \pm 1.96n^{-1/2}w_{ii}^{1/2}, \quad i = 1, \dots, h$$

Analyse graphique de $\rho_X(\cdot)$ et $\hat{\rho}(h)$



- L'allure des données :
 - ⇒ stationnarité et faible de dépendance donc l'AR(1) est un bon candidat
- Analyse de l'ACF empirique et ses intervalles de confiance :
 - ⇒ l'ACF théorique touche les intervalles aux retards 2 à 4
 - ⇒ cela suggère une certaine incompatibilité entre les données et le modèle

Analyse graphique de $\rho_X(\cdot)$ et $\hat{\rho}(h)$

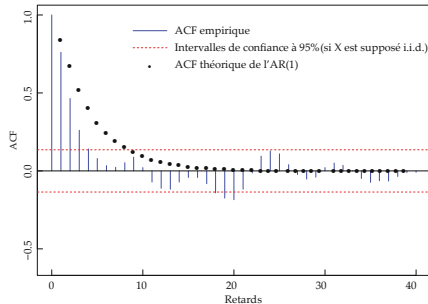
- Dans l'hypothèse où X_t serait une séquence i. i. d. ($0, \sigma_X^2 < \infty$)

$$\hat{\rho} \xrightarrow{a.a.d} \mathcal{N}(\rho, n^{-1}), \quad \rho = 0$$

ce qui implique que 95% des fois, $\hat{\rho}$ devrait tomber dans l'intervalle

$$\pm 1.96n^{-1/2}$$

ce qui ici revient à ± 0.1990 car $n = 97$



Les estimateurs de Yule-Walker d'un AR

- Soit un processus AR(p) stationnaire : $X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t$
 - On a vu au Ch. 1 qu'en multipliant par X_{t-j} , de chaque côté ...
- ... et en prenant l'espérance, on obtenait les équations de Yule-Walker

⇒ Exprimées en terme d'ACovF sous forme matricielle on a

$$\Gamma_p \phi_p = \gamma_p$$

avec $\Gamma_p = [\gamma(i-j)]_{i,j=1}^p$ et $\gamma_p = (\gamma(1), \dots, \gamma(p))'$

- Dans la pratique on voudra remplacer $\gamma(j)$ par $\hat{\gamma}(j)$ et construire

$$\hat{\phi}_p = \hat{R}_p^{-1} \hat{\rho}_p$$

l'estimateur de Yule-Walker des p coefficients AR ainsi que

$$\hat{\sigma}_\varepsilon^2 = \hat{\gamma}(0) (1 - \hat{\rho}_p' \hat{R}_p^{-1} \hat{\rho}_p)$$

où $\hat{\rho}_p = (\rho(1), \dots, \rho(p))' = \gamma_p / \gamma(0)$

Inférence et équations de Yule-Walker

- On peut montrer que pour $n \rightarrow \infty$ sous certaines conditions

$$\hat{\phi}_p \xrightarrow{a.a.d} \mathcal{N}(\phi, n^{-1} \sigma_\varepsilon^2 \Gamma_p^{-1})$$

- En remplaçant σ_ε^2 par $\hat{\sigma}_\varepsilon^2$, on peut écrire que l'intervalle

$$\hat{\phi}_j \pm \Phi_{1-\alpha/2} n^{-1/2} \hat{\sigma}_\varepsilon$$

contient ϕ_j avec une probabilité de $(1 - \alpha)$ avec

$$\Phi_{1-\alpha/2}$$

le quantile à $(1 - \alpha)$ d'une distribution normale centrée réduite

Les estimateurs de Yule-Walker en présence de composantes MA

- La limite de Yule-Walker se rencontre pour les ARMA($p, q > 0$)

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

- En effet, les équations à solutionner pour $p > 0$ et $q > 0$ sont

$$\hat{\gamma}(h) - \phi_1 \hat{\gamma}(h-1) - \dots - \phi_p \hat{\gamma}(h-p) = \sigma_\varepsilon^2 \sum_{j=h}^q \theta_j \psi_{j-h}$$

où les $\Psi_z = \Theta(z)\Phi(z)^{-1}$ sont les coefficients de la forme MA(∞)

- On voit que malgré son écriture simple, ce problème est non-linéaire

⇒ On ne peut garantir l'existence et l'unicité de la solution

Note Pour un processus MA pur, il est possible d'utiliser Yule-Walker

Les estimateurs de Yule-Walker d'un pure MA

- Soit un MA(1) avec $|\theta_1| < 1$: $X_t = \varepsilon_t + \theta_1\varepsilon_{t-1}$
- On sait que (cf. Ch1) $\hat{\gamma}(0) = \hat{\sigma}_\varepsilon^2(1 + \hat{\theta}_1^2)$ et

$$\hat{\rho}(1) = \frac{\hat{\theta}_1}{1 + \hat{\theta}_1^2}$$

- On a vu qu'on peut en déduire $\hat{\theta}_1 \in \mathbb{R}$ si $|\hat{\rho}(1)| \leq 1/2$ ($\hat{\theta}_1 \in \mathbb{C}$ sinon)

$$\hat{\theta}_1 = (2\hat{\rho}(1))^{-1}(1 - (1 - 4\hat{\rho}^2(1))^{1/2})$$

et

$$\hat{\sigma}_\varepsilon^2 = (1 + \hat{\theta}_1^2)\hat{\gamma}(0)$$

Note Si $|\hat{\rho}(1)| = 0.5$ on trouve $|\hat{\theta}_1| = 1$ et le processus n'est pas inversible

PACF : estimation et inférence

- Comme énoncé au C1, Yule-Walker nous permet d'obtenir les PACFs
- Les PACFs empiriques peuvent également s'obtenir via la régression

$$x_t = \hat{\phi}_{1,j}x_{t-j} + \dots + \hat{\phi}_{j,j}x_{t-j} + \varepsilon_t$$

avec l'estimateur OLS (ou MLE Gaussien)

$$\hat{\phi}_{j,j}$$

donnant la corrélation partielle d'ordre j

- La théorie limite de ces estimateurs nous révèle que pour $j > p$

$$\hat{\phi}_{j,j} \xrightarrow{a.a.d} \mathcal{N}(0, n^{-1})$$

et donc qu'un intervalle de confiance à 95% construit autour de

$$\phi_{j,j} = 0$$

est simplement donné par (formule de Quenouille)

$$\pm \frac{1.96}{\sqrt{n}}$$

Rappels MLE

- Vous connaissez déjà l'estimateur du Maximum de Vraisemblance (MLE)
- Pour un ensemble de paramètres ϑ

- le MLE est convergent

$$\hat{\vartheta} \xrightarrow{p} \vartheta_0$$

- le MLE est asymptotiquement efficace

$$\mathbb{V}(\hat{\vartheta}) = I_n^{-1}(\vartheta_0)$$

- le MLE est asymptotiquement normalement distribué (pour des lois exp)

$$\sqrt{n}(\hat{\vartheta} - \vartheta_0) \xrightarrow{d} \mathcal{N}(0, I_n^{-1}(\vartheta_0))$$

où $I_n^{-1}(\vartheta_0)$ représente la matrice d'information de Fisher

- sous certaines hypothèse de régularité

Note l'élément crucial dans la construction de la vraisemblance étant le **choix de la distribution**

Densité conditionnelle Gaussienne et AR(1)

- Soit un ARMA(p, q) dont on suppose $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

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

- Si on sait que $p = 1$ et $q = 0$, on a un AR(1) pour lequel on sait que

$$X_1 \sim \mathcal{N}(0, \sigma_\varepsilon^2(1 - \phi_1^2)^{-1})$$

$$\Rightarrow f_{X_1}(x_1; \phi_1) = (\sigma_\varepsilon \sqrt{2\pi})^{-1} \exp\left(-\frac{1}{2} \frac{x_1^2}{\sigma_\varepsilon^2}\right)$$

$$X_2|X_1 \sim \mathcal{N}(\phi_1 X_1, \sigma_\varepsilon^2)$$

$$\Rightarrow f_{X_2|X_1}(x_2|x_1; \phi_1) = (\sigma_\varepsilon \sqrt{2\pi})^{-1} \exp\left(-\frac{1}{2} \frac{(x_2 - \phi_1 x_1)^2}{\sigma_\varepsilon^2}\right)$$

⋮

$$X_n|X_{n-1} \sim \mathcal{N}(\phi_1 X_{n-1}, \sigma_\varepsilon^2)$$

$$\Rightarrow f_{X_n|X_{n-1}}(x_n|x_{n-1}; \phi_1) = (\sigma_\varepsilon \sqrt{2\pi})^{-1} \exp\left(-\frac{1}{2} \frac{(x_n - \phi_1 x_{n-1})^2}{\sigma_\varepsilon^2}\right)$$

Maximum de Vraisemblance Gaussien Exact et AR(1)

- Des densités conditionnelles découle la fonction de vraisemblance

$$L_n(\vartheta; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{(1 - \phi_1^2)^{1/2}}{(\sigma_\varepsilon \sqrt{2\pi})^n} \exp\left(-\frac{(1 - \phi_1^2)\mathbf{x}_1^2}{2\sigma_\varepsilon^2}\right) \prod_{t=2}^n \exp\left(-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}\right)$$

- Dont l'écriture sous forme de log-vraisemblance donne

$$\ell_n(\vartheta; \mathbf{x}_1, \dots, \mathbf{x}_n) = -\frac{1}{2} \ln\left(\frac{(1 - \phi_1^2)}{(\sigma_\varepsilon \sqrt{2\pi})^n}\right) - \frac{1}{2\sigma_\varepsilon^2} \left((1 - \phi_1^2)\mathbf{x}_1^2 + \sum_{t=2}^n \varepsilon_t^2 \right)$$

avec $\vartheta = (\phi_1, \sigma_\varepsilon^2)'$

- Le second terme fait intervenir : $RSS_c := \sum_{t=2}^n \varepsilon_t^2$
- On peut construire un OLS conditionnel basé sur la minisation de RSS_c
- Ou un OLS non-conditionnel basé sur $RSS_c + (1 - \phi_1^2)\mathbf{x}_1^2$

Note Seul le MLE permet de gérer la non-linéarité engendrée par $q > 0$

$$\hat{\vartheta} = \arg \max_{\vartheta \in \mathbb{R}} \ell_n(\vartheta; \mathbf{x}_1, \dots, \mathbf{x}_n)$$

Maximum de Vraisemblance Gaussien Exact et AR(p)

- Dans le cas d'un AR(p), il nous faut isoler

$$f_{X_p, X_{p-1}, \dots, X_1}(x_p, x_{p-1}, \dots, x_1; \vartheta)$$

Rappel La vraisemblance Gaussienne de $\mathbf{X}_n = (X_1, \dots, X_n)'$ s'écrit

$$\ell_n(\vartheta; X_1, \dots, X_n) = (2\pi)^{-n/2} (|\Gamma_n|)^{-1/2} \exp\left(-1/2 \mathbf{X}_n' \Gamma_n^{-1} \mathbf{X}_n\right)$$

- Pour $\mathbf{X}_p = (X_1, \dots, X_p)'$ on a

$$\Gamma_p = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{p-2} \\ \vdots & \vdots & \dots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \dots & \gamma_1 \end{pmatrix}$$

et donc la densité Gaussienne des p premières observations devient

$$f_{X_p, X_{p-1}, \dots, X_1}(x_p, x_{p-1}, \dots, x_1; \vartheta) = (2\pi)^{-p/2} (|\Gamma_p|)^{-1/2} \exp\left(-\frac{1}{2} \mathbf{X}_p' \Gamma_p^{-1} \mathbf{X}_p\right)$$

menant à la densité Gaussienne complète $f_{X_n, \dots, X_1}(x_n, \dots, x_1; \vartheta) =$

$$f_{X_p, X_{p-1}, \dots, X_1}(x_p, x_{p-1}, \dots, x_1; \vartheta) \prod_{t=p+1}^n f_{X_t | X_{t-1}, \dots, X_{t-p}}(x_t | x_{t-1}, \dots, x_{t-p}; \vartheta)$$

Maximum de Vraisemblance Gaussien conditionnel et AR(p)

- Il est possible de considérer les p observations comme déterministes
- En conditionnant la densité complète sur ces p observations on obtient

$$f_{X_n, \dots, X_{p+1} | X_1, \dots, X_p}(\mathbf{x}_n, \dots, \mathbf{x}_{p+1} | \mathbf{x}_1, \dots, \mathbf{x}_p; \vartheta) = \prod_{t=p+1}^n f_{X_t | X_{t-1}, \dots, X_{t-p}}(x_t | x_{t-1}, \dots, x_{t-p}; \vartheta)$$

- Dès lors, dans la construction de la vraisemblance, on sacrifiera

$$\mathbf{X}_p = (X_1, \dots, X_p)'$$

- Mais le problème d'optimisation s'en trouvera simplifié

$$\ell_n(\vartheta; \mathbf{x}_p, \dots, \mathbf{x}_n) = -\frac{1}{2} \ln \left(\frac{1}{(\sigma_\varepsilon \sqrt{2\pi})^n} \right) - \frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=p+1}^n \varepsilon_t^2 \right)$$

et il y aura équivalence avec les OLS

Maximum de Vraisemblance Gaussien conditionnel et MA(q)

- Partons de X_1 où

$$X_1 = \varepsilon_1 + \theta_1 \varepsilon_0 + \dots + \theta_q \varepsilon_{-q+1}$$

sous l'hypothèse que $\varepsilon_0 = (\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1})' = \mathbf{0}$, qui nous donne

$$(X_1 | \varepsilon_0 = \mathbf{0}) \sim \mathcal{N}(0, \sigma_\varepsilon^2) \Rightarrow f_{X_1 | \varepsilon_0}(x_1 | \varepsilon_0; \vartheta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{(X_1^2 = \varepsilon_1^2)}{2\sigma_\varepsilon^2}\right)$$

Pour X_2 , notons que X_1 et donc ε_1 est observable

$$X_2 = \varepsilon_2 + \theta_1 \varepsilon_1 + \dots + \theta_q \varepsilon_{-q+2}$$

ce qui nous donne

$$(X_2 | X_1 = x_1, \varepsilon_0 = \mathbf{0}) \sim \mathcal{N}(\theta_1 \varepsilon_1, \sigma_\varepsilon^2)$$

et donc $f_{X_2 | X_1, \varepsilon_0}(x_2 | x_1, \varepsilon_0; \vartheta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{(X_2 - \theta_1 \varepsilon_1)^2}{2\sigma_\varepsilon^2}\right)$

$$= \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{(\varepsilon_2)^2}{2\sigma_\varepsilon^2}\right)$$

En itérant on obtient, de la même manière que pour les AR, la vraisemblance conditionnelle

$$\ell_n(\vartheta; \mathbf{x}_1, \dots, \mathbf{x}_n) = -\frac{n}{2} \ln\left(\frac{1}{(\sigma_\varepsilon \sqrt{2\pi})}\right) - \frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=1}^n \varepsilon_t^2\right)$$

Vraisemblance Gaussienne et ARMA(p, q)

- Pour $p > 0$ et $q > 0$, en se basant sur l'AI (Ch. 1) et

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

une écriture générale de la vraisemblance Gaussienne est possible

$$L_n(\vartheta; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{1}{\sqrt{(2\pi\sigma_\varepsilon^2)r_0 \dots r_{n-1}}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=p+1}^n \frac{(x_j - \widehat{x}_j)^2}{r_{j-1}}\right)$$

- Les estimateurs du maximum de vraisemblance (Gaussien) sont alors

$$\widehat{\sigma}_\varepsilon^2 = n^{-1} S(\widehat{\vartheta}_{p,q}; \mathbf{x}_1, \dots, \mathbf{x}_n) \text{ où } S(\widehat{\vartheta}_{p,q}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{(x_j - \widehat{x}_j)^2}{r_{j-1}}$$

et

$$\widehat{\vartheta}_{p,q} = (\widehat{\phi}_1, \dots, \widehat{\phi}_p, \widehat{\theta}_1, \dots, \widehat{\theta}_q)'$$

sont les valeurs qui minimisent (l'opposée de) la log-vraisemblance Gaussienne concentrée

$$\ell_n(\vartheta_{p,q}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \ln(n^{-1} S(\vartheta_{p,q}; \mathbf{x}_1, \dots, \mathbf{x}_n)) + n^{-1} \sum_{j=p+1}^n \ln r_{j-1}$$

Note Concentrer : résoudre sur un paramètre puis le faire dépendre des autres

Inférence et vraisemblance Gaussienne d'un ARMA(p, q)

- Pour $n \rightarrow \infty$ on peut montrer que $r_n = 1$ et

$$\hat{\vartheta} \xrightarrow{a.a.d} \mathcal{N}(\vartheta, n^{-1}\mathbb{V}(\vartheta))$$

La covariance $n^{-1}\mathbb{V}(\vartheta)$ peut être approximée via

$$\hat{H}_n(\vartheta; \mathbf{x}_1, \dots, \mathbf{x}_n)^{-1}$$

la Hessienne numérique évaluée par l'optimiseur à l'optimum

⇒ Les écarts-type seront donc aussi évaluables

- Pour des ARMA simples, la variance asymptotique est disponible

⇒ Pour n grand, l'usage direct de $n^{-1}\mathbb{V}(\vartheta)$ sera donc possible

Note Si n est petit, vous verrez au S2 qu'il existe des techniques de Bootstrap

Variance asymptotique et vraisemblance Gaussienne

- Pour un AR(p) stationnaire, $\mathbb{V}(\vartheta)$ est identique à celle de Yule-Walker

$$\mathbb{V}(\vartheta_p) = \sigma_\varepsilon^2 \Gamma_p^{-1}$$

et plus particulièrement pour un AR(1) et un AR(2) on obtient

$$\mathbb{V}(\vartheta_1) = (1 - \phi_1^2) \text{ et } \mathbb{V}(\vartheta_2) = \begin{pmatrix} (1 - \phi_1^2) & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & (1 - \phi_2^2) \end{pmatrix}$$

- Pour un MA(q) inversible, il faut adopter une écriture en AR(q)

$$X_t + \theta_1 X_{t-1} + \dots + \theta_q X_{t-q} = \varepsilon_t$$

et il peut être montré que $\mathbb{V}(\vartheta_q) = \sigma_\varepsilon^2 \Gamma_q^{-1}$ et e.g. pour $q = 1$ et $q = 2$

$$\mathbb{V}(\vartheta_1) = (1 - \theta_1^2) \text{ et } \mathbb{V}(\vartheta_2) = \begin{pmatrix} (1 - \theta_1^2) & \theta_1(1 - \theta_2) \\ \theta_1(1 - \theta_2) & (1 - \theta_2^2) \end{pmatrix}$$

- Pour un ARMA(1, 1) stationnaire causal on peut montrer que

$$\mathbb{V}(\vartheta_{1,1}) = \frac{1 + \phi_1 \theta_1}{(\phi_1 + \theta_1)^2} \begin{pmatrix} (1 - \phi_1^2)(1 + \phi_1 \theta_1) & -(1 - \theta_1^2)(1 - \phi_1^2) \\ -(1 - \theta_1^2)(1 - \phi_1^2) & (1 - \theta_1^2)(1 + \phi_1 \theta_1) \end{pmatrix}$$

Estimation d'un AR(1) et inférence

- Considérons les rendements du Dow Jones journalier sur quelques mois

$$n = 77$$

- L'ACF et la PACF nous révèlent une faible dépendance au passé
- L'estimation par MLE d'un AR(1) nous donne

$$X_t = 0.4471X_{t-1} + \varepsilon_t$$

et le logiciel utilisé nous retourne également $\hat{\sigma}_{\phi_1} = 0.1050$

- Pour l'écart-type asymptotique, $\sigma_{\phi_1} = \sqrt{\mathbb{V}(\phi_1)n^{-1}}$, on trouve

$$\sqrt{(1 - 0.4471^2)/77} = 0.1019$$

⇒ σ_{ϕ_1} et $\hat{\sigma}_{\phi_1}$ sont relativement proches

- Pour le calcul de l'intervalle de confiance à 95% on obtient

$$0.4471 \pm 1.96 \times 0.1050 = (0.2413, 0.6529)$$

ou

$$0.4471 \pm 1.96 \times 0.1019 = (0.2473, 0.6468)$$

Comment sélectionner un modèle

- Dans l'exemple S107, on peut se demander si l'AR(1) est approprié
- ⇒ plusieurs réponses existent à la question de la sélection des retards
- 1 Dans un premier temps, les examens des ACF et PACF ont été utilisés
- ⇒ approche à la Box et Jenkins pour déterminer si MA, AR, ou ARMA
 - 2 Les critères d'information sont plus fiables pour le choix de p et q

⇒ on présentera AIC, BIC et HQ

 - 3 Des tests de validation suivent ces procédures d'identifications

⇒ divers tests sur les résidus

La méthodologie de Box et Jenkins

- L'approche Box et Jenkins s'inscrit dans la recherche de la parcimonie
- ⇒ comment prédire avec le minimum de paramètres
- On peut résumer la méthodologie de BJ en 4 étapes
 - 1 Transformer les données pour satisfaire l'hypothèse de stationnarité
 - 2 Sélectionner p, q petits pour décrire la série à l'aide d'un $\text{ARMA}(p, q)$
 - 3 Estimer les paramètres AR et MA du modèle
 - 4 Procéder à une validation de la spécification retenue
 - 5 Prévisions
 - L'étape 1 sera détaillée au C3 et l'étape 3 relève de la section précédente
 - L'étape 2 est appelée étape d'identification par BJ et repose sur
$$\hat{\rho}_j \text{ et } \hat{\phi}_{j,j}, \quad j = 1, 2, \dots$$
dont les comportements peuvent nous renseigner sur p et q
 - L'étape 4 est détaillée dans la section suivante

L'identification de p chez Box et Jenkins

Rappel Au C1 nous avons vu que pour un AR($p > 0$)

$$p = \inf \{j | \phi_{j,j} = 0\}$$

- Dans la pratique il faut donc estimer $\phi_{j,j}$ pour $j = 1, 2, \dots, h$ et tester

$$H_0 : \phi_{1,1} = 0 \text{ versus } H_1 : \phi_{1,1} \neq 0$$

- D'après le S97 on sait que si $|\hat{\phi}_{1,1}| > 1.96/\sqrt{n}$ on rejette H_0

⇒ si H_0 n'est pas rejeté on conclut que $p < 1$, mais si on rejette, on teste

$$H_0 : \phi_{2,2} = 0 \text{ versus } H_1 : \phi_{2,2} \neq 0$$

- De nouveau, si $|\hat{\phi}_{2,2}| > 1.96/\sqrt{n}$ on rejette H_0

⇒ si on ne rejette pas, $p = 1$, mais si on rejette, on teste

$$H_0 : \phi_{3,3} = 0 \text{ versus } H_1 : \phi_{3,3} \neq 0$$

- Cette procédure se prolonge ainsi jusqu'à arriver à

$$H_0 : \phi_{h,h} = 0 \text{ versus } H_1 : \phi_{h,h} \neq 0$$

avec $h = p + 1$

L'identification de q chez Box et Jenkins

Rappel Au C1 nous avons vu que pour un MA($q > 0$), $q = \inf\{j | \rho_j = 0\}$

- La même stratégie peut donc être mise en place en estimant ρ_j
- Pour un MA(q) Gaussien, la formule de Bartlett (cf. S89) nous donne

$$\mathbb{V}(\hat{\rho}_j) = \frac{1}{n} \left(1 + 2 \sum_{i=1}^q \rho_i^2 \right), \quad j = q + 1, q + 2, \dots$$

où dans la pratique on remplacera ρ_i par $\hat{\rho}_i$

- La procédure sera alors de commencer par estimer ρ_1 et de tester

$$H_0 : \rho_1 = 0 \text{ versus } H_1 : \rho_1 \neq 0$$

sachant que si $|\hat{\rho}_1| > 1.96/\sqrt{n}$ on rejette H_0

⇒ si H_0 n'est pas rejeté on conclut que $q < 1$, mais si on rejette, on teste

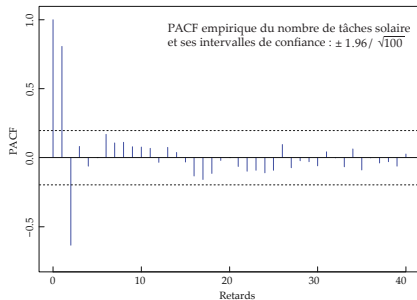
$$H_0 : \rho_2 = 0 \text{ versus } H_1 : \rho_2 \neq 0$$

sachant que si $|\hat{\rho}_2| > 1.96\sqrt{n^{-1}(1 + 2\hat{\rho}_1^2)} = 1.96\sqrt{\mathbb{V}(\hat{\rho}_2)}$ on rejette H_0

- Cette procédure se prolonge ainsi jusqu'à

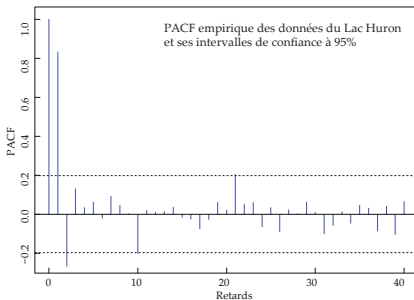
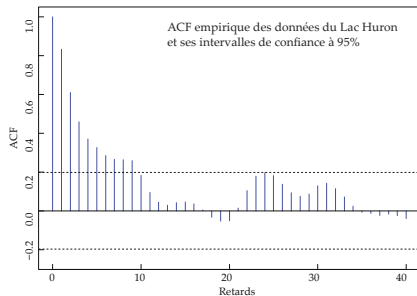
$$H_0 : \rho_h = 0 \text{ versus } H_1 : \rho_h \neq 0, \text{ avec } h = q + 1$$

Exemple d'analyse de la PACF chez Box et Jenkins



- Soit la séquence du nombre de Wolf (1770-1869) et sa PACF
 - Analysons les PACFs significatives (attention, $\phi_{0,0} = \rho_0 = 1$)
 - Au delà de 2 retards ($\phi_{1,1}$ et $\phi_{2,2}$), les PACFs sont dans les intervalles
- ⇒ Cela suggère qu'un modèle candidat est serait l'AR(2)

Exemple d'analyse graphique chez Box et Jenkins



- Soit les données du Lac Huron vues au S90
 - Analysons les ACFs et PACFs significatives
 - La décroissance de l'ACF est graduelle ne suggérant pas un MA
 - Inversement, une rupture dans la PACF s'observe au delà de 2 retards
- ⇒ Cela suggère de nouveau qu'un modèle candidat serait l'AR(2)

La philosophie du critère AIC (Akaike Information Criterion)

- L'idée de l'AIC est de minimiser une divergence distributionnelle

$$d(\tilde{\vartheta}|\vartheta) = \Delta(\tilde{\vartheta}|\vartheta) - \Delta(\vartheta|\vartheta)$$

où la distance de Kullback-Leibler

$$\Delta(\tilde{\vartheta}|\vartheta) = \mathbb{E}(-2 \ln f(\mathbf{X}; \tilde{\vartheta})) = \int_{\mathbb{R}^n} -2f(\mathbf{x}; \vartheta) \ln f(\mathbf{x}; \tilde{\vartheta}) d\mathbf{x}$$

mesure la dissimilarité entre la famille de fonctions de densités

$$\{f(\mathbf{x}; \tilde{\vartheta}), \tilde{\vartheta} \in \Theta\}$$

dont on suppose que $\mathbf{X} = (X_1, \dots, X_n)'$ est tiré, et $f(\mathbf{x}; \vartheta)$

- En effet on peut voir que par l'inégalité de Jensen,

$$\begin{aligned}d(\tilde{\vartheta}|\vartheta) &= \int_{\mathbb{R}^n} -2f(\mathbf{x}; \vartheta) \ln \left(\frac{f(\mathbf{x}; \tilde{\vartheta})}{f(\mathbf{x}; \vartheta)} \right) d\mathbf{x} \\ &\geq -2 \ln \left(\int_{\mathbb{R}^n} \frac{f(\mathbf{x}; \tilde{\vartheta})}{f(\mathbf{x}; \vartheta)} f(\mathbf{x}; \vartheta) d\mathbf{x} \right) \\ &= -2 \ln \left(\int_{\mathbb{R}^n} f(\mathbf{x}; \tilde{\vartheta}) d\mathbf{x} \right) = 0\end{aligned}$$

La construction du critère AIC corrigé

- Bien sur $d(\hat{\vartheta}|\vartheta)$ doit être estimé et pour cela on suppose la Normalité
- ⇒ Pour $\vartheta = (\vartheta_{p,q}, \sigma_\varepsilon^2)'$, $f(\cdot; \vartheta)$ est la fonction de densité de

$$\mathbf{Y} = (Y_1, \dots, Y_n)'$$

où Y_t est un ARMA(p, q) Gaussien de paramètre $\vartheta_{p,q}$ et σ_ε^2

- Supposons que $\mathbf{x} = (x_1, \dots, x_n)'$ provient d'un ARMA Gaussien avec p, q et ϑ ses vrais paramètres
- Notons $\hat{\vartheta}$ l'estimateur MLE basé sur \mathbf{X} , alors

$$-2 \ln L_n(\hat{\vartheta}; \mathbf{Y}) = -2 \ln L_n(\hat{\vartheta}; \mathbf{x}) + \hat{\sigma}_\varepsilon^{-2} \mathcal{S}(\hat{\vartheta}_{p,q}; \mathbf{Y}) - n$$

et donc on peut écrire $\mathbb{E}_\vartheta(\Delta(\hat{\vartheta}|\vartheta))$ comme

$$\mathbb{E}_\vartheta(-2 \ln L_n(\hat{\vartheta}; \mathbf{Y})) = \mathbb{E}_\vartheta(-2 \ln L_n(\hat{\vartheta}; \mathbf{x})) + \mathbb{E}_\vartheta(\hat{\sigma}_\varepsilon^{-2} \mathcal{S}(\hat{\vartheta}_{p,q}; \mathbf{Y})) - n$$

dont l'approximation asymptotique du terme

$$\mathbb{E}_\vartheta(\hat{\sigma}_\varepsilon^{-2} \mathcal{S}(\hat{\vartheta}_{p,q}; \mathbf{Y})) = \frac{2(p+q+1)n}{n-p-q-2}$$

nous fournissent un estimateur non-biaisé de la distance de KL espérée

$$AICc = -2 \ln L_n(\hat{\vartheta}; \mathbf{x}) + \frac{2(p+q+1)n}{n-p-q-2}$$

Critère AIC versus AICc

- A l'origine, le critère AIC fut proposé comme

$$AIC = -2 \ln L_n(\hat{\vartheta}; \mathbf{x}) + 2(p + q + 1)$$

⇒ Asymptotiquement AIC et AICc sont équivalents

Note 1 En échantillon fini, des simulations montrent que l'AIC surestime p

⇒ phénomène d'overfitting

Note 2 Si le vrai p est grand, l'AICc sous-estime souvent p

⇒ phénomène overfitting inverse

Note 3 Les deux critères sont basés sur deux éléments

⇒ "qualité d'ajustement" (vraisemblance) + "pénalité" (# de paramètres)

Critères BIC et HQ

- De nombreux autres critères existent dont le BIC et l'HQ

BIC Le Bayesian Information Criterion se distingue par sa pénalité

$$BIC = -2 \ln L_n(\hat{\vartheta}; \mathbf{x}) + (p + q + 1) \ln(n)$$

HQ Le critère d'Hannan-Quinn se distingue aussi par sa pénalité

$$HQ = -2 \ln L_n(\hat{\vartheta}; \mathbf{x}) + (p + q + 1) \ln(\ln(n))$$

- Comparaison avec AICc :
 - BIC et HQ sont consistants (pas AICc): $\tilde{\vartheta} \in \Theta$
 - ⇒ $\hat{p} \rightarrow p$ et $\hat{q} \rightarrow q$ avec une probabilité unitaire
 - AICc est efficace, i.e. minimise la MSE (pas BIC ni HQ) : $\tilde{\vartheta} \notin \Theta$
 - ⇒ il minimise le risque de choisir un très mauvais modèle

Comparaison par simulation

- Soit des simulations Monte Carlo : AR(1), AR(2), MA(2), ARMA(1, 1)
- ⇒ 10000 simulations de $n \in \{30, 60, 180, 500\}$ observations
- On fixe $p_{\max} = q_{\max} = 4$ et on applique chacun des 3 critères
- ⇒ les tableaux suivants reportent les % de sélection de p et q par critère

Note V désigne le nombre de procédures d'estimation ayant abouti

DGP : AR(1)		P				
		0	1	2	3	4
n = 30	AIC	1%	71%	14%	7%	7%
	BIC	2%	87%	7%	2%	1%
	HQ	1%	79%	11%	5%	4%
n = 60	AIC	0%	75%	13%	7%	6%
	BIC	0%	94%	5%	1%	0%
	HQ	0%	86%	9%	3%	2%
n = 180	AIC	0%	75%	13%	7%	6%
	BIC	0%	97%	2%	0%	0%
	HQ	0%	90%	7%	2%	1%
n = 500	AIC	0%	77%	12%	6%	5%
	BIC	0%	98%	2%	0%	0%
	HQ	0%	93%	5%	1%	0%

DGP : AR(2)		P				
		0	1	2	3	4
n = 30	AIC	11%	28%	42%	11%	8%
	BIC	21%	39%	34%	4%	2%
	HQ	14%	33%	40%	8%	5%
n = 60	AIC	1%	8%	70%	12%	9%
	BIC	3%	22%	71%	3%	1%
	HQ	1%	14%	74%	7%	4%
n = 180	AIC	0%	0%	78%	14%	8%
	BIC	0%	0%	97%	2%	0%
	HQ	0%	0%	91%	7%	2%
n = 500	AIC	0%	0%	78%	13%	9%
	BIC	0%	0%	99%	1%	0%
	HQ	0%	0%	93%	5%	2%

DGP : MA(2)		q				
		0	1	2	3	4
n = 30	AIC	17%	15%	42%	14%	12%
	BIC	36%	18%	35%	6%	5%
	HQ	23%	16%	41%	11%	9%
	V	10000	9977	9902	9245	8076
n = 60	AIC	2%	5%	67%	15%	11%
	BIC	12%	13%	69%	4%	2%
	HQ	5%	8%	72%	9%	6%
	V	10000	10000	9995	9961	9802
n = 180	AIC	0%	0%	77%	13%	9%
	BIC	0%	0%	97%	3%	0%
	HQ	0%	0%	90%	7%	3%
	V	10000	10000	10000	10000	10000
n = 500	AIC	0%	0%	78%	14%	8%
	BIC	0%	0%	99%	1%	0%
	HQ	0%	0%	93%	5%	2%
	V	10000	10000	10000	10000	10000

Comparaison par simulation : ARMA(1, 1)

DGP : ARMA(1,1)													DGP : ARMA(1,1)																						
n = 30						n = 180						n = 60						n = 500																	
AIC						HQ						BIC						AIC																	
q						q						q						q																	
0 1 2 3 4						0 1 2 3 4						0 1 2 3 4						0 1 2 3 4																	
p	0	0%	1%	4%	2%	9%	p	0	16%	0%	0%	0%	0%	p	0	0%	0%	0%	0%	0%	p	0	0%	0%	0%	0%	0%	p	0	0%	0%	0%	0%	0%	
	1	4%	28%	5%	2%	4%		1	41%	0%	0%	0%	0%		1	0%	46%	4%	2%	2%		1	0%	41%	3%	2%	1%		1	0%	92%	1%	0%	0%	
	2	9%	6%	1%	2%	1%		2	21%	0%	0%	0%	0%		2	0%	5%	9%	2%	2%		2	0%	2%	2%	0%	0%		2	0%	4%	9%	2%	1%	
	3	8%	1%	1%	0%	0%		3	10%	0%	0%	0%	0%		3	0%	2%	2%	11%	3%		3	0%	0%	0%	1%	0%		3	0%	2%	2%	17%	5%	
4	13%	3%	1%	0%	0%	4	13%	0%	0%	0%	0%	4	0%	2%	1%	3%	3%	4	0%	0%	0%	0%	0%	4	0%	0%	0%	0%	0%	4	0%	1%	1%	4%	6%
p	0	0%	1%	6%	2%	6%	p	0	10000	7676	7514	5159	5050	p	0	0%	0%	0%	0%	0%	p	0	10000	9711	9958	9663	9911								
	1	2%	39%	3%	1%	2%		1	10000	8066	7247	4955	3755		1	0%	92%	1%	0%	0%		1	10000	9996	9989	9948	9935								
	2	17%	4%	1%	1%	1%		2	10000	6726	2411	1937	993		2	0%	2%	2%	0%	0%		2	10000	9908	7304	7870	7451								
	3	6%	1%	0%	0%	0%		3	10000	6089	2432	608	248		3	0%	0%	0%	1%	0%		3	10000	9913	8500	6059	5577								
4	6%	0%	0%	0%	0%	4	10000	5489	1790	452	99	4	0%	0%	0%	0%	0%	4	10000	9796	7739	5674	3250												
p	0	0%	0%	1%	5%	5%	p	0	6%	0%	0%	0%	0%	p	0	0%	0%	0%	0%	0%	p	0	0%	0%	0%	0%	0%								
	1	0%	44%	5%	2%	3%		1	66%	0%	0%	0%	0%		1	0%	41%	3%	2%	1%		1	82%	0%	0%	0%	0%								
	2	1%	7%	4%	3%	3%		2	13%	0%	0%	0%	0%		2	0%	4%	9%	2%	1%		2	9%	0%	0%	0%	0%								
	3	3%	2%	1%	2%	1%		3	7%	0%	0%	0%	0%		3	0%	2%	2%	17%	5%		3	7%	0%	0%	0%	0%								
4	7%	3%	1%	1%	0%	4	8%	0%	0%	0%	0%	4	0%	1%	1%	4%	6%	4	2%	0%	0%	0%	0%												
p	0	0%	0%	1%	3%	4%	p	0	10000	8599	9271	7843	8215	p	0	0%	0%	0%	0%	0%	p	0	10000	9991	9997	9954	9994								
	1	0%	73%	3%	0%	1%		1	10000	9444	9228	8261	7667		1	0%	96%	1%	0%	0%		1	10000	10000	10000	10000	10000								
	2	3%	4%	1%	1%	1%		2	10000	8713	4472	4697	3861		2	0%	1%	1%	0%	0%		2	10000	9998	8650	9002	8717								
	3	3%	0%	0%	0%	0%		3	10000	8432	5193	2165	1520		3	0%	0%	0%	0%	0%		3	10000	9997	9380	8143	7778								
4	3%	0%	0%	0%	0%	4	10000	7867	4381	1902	698	4	0%	0%	0%	0%	0%	4	10000	9994	8983	7967	5808												

Limites de l'utilisation des critères

- On considère de nouveau les données du lac Huron (cf. S90)
- Si l'économètre balaye uniquement sur $p > 0$:

⇒ Les AICc et BIC minimum sont trouvés pour $p = 2$ et on obtient

$$X_t - 1.0441X_{t-1} + 0.2503X_{t-2} = \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 0.4789)$$

avec $AICc = 213.54$ et $BIC = 217.63$

- Si l'économètre balaye sur $p > 0$ et $q > 0$:

⇒ Les AICc et BIC minimum sont trouvés pour $p = 1$ et $q = 1$ et on obtient

$$X_t - 0.7446X_{t-1} = \varepsilon_t + 0.3213\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{N}(0, 0.4750)$$

avec $AICc = 212.77$ et $BIC = 217.86$

Note Les critères étant très proches, difficile de déterminer le meilleur modèle

Identification

- Il est possible d'intégrer des variables exogènes, X_t dans un ARMA
- ⇒ mais l'identification et l'estimation du processus se complexifient

- Pour comprendre cela considérons l'ARMAX suivant

$$Y_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \dots + \alpha_m X_{t-m} + \nu_t$$

où ν_t suit un ARMA(p, q) indépendant $\forall t$ de X_t

- Sous l'hypothèse que $X_t \sim WN(0, \sigma_X^2)$, on a

$$\mathbb{E}(Y_t X_t) = \alpha_0 \sigma_X^2 \Rightarrow \text{Corr}(Y_t, X_t) = \alpha_0 \sigma_X \sigma_Y^{-1}$$

$$\mathbb{E}(Y_t X_{t-1}) = \alpha_1 \sigma_X^2 \Rightarrow \text{Corr}(Y_t, X_{t-1}) = \alpha_1 \sigma_X \sigma_Y^{-1}$$

$$\mathbb{E}(Y_t X_{t-2}) = \alpha_2 \sigma_X^2 \Rightarrow \text{Corr}(Y_t, X_{t-2}) = \alpha_2 \sigma_X \sigma_Y^{-1}$$

ce qui implique que $\text{Corr}(Y_t, X_{t-j})$ est proportionnelle à $\partial Y_t / (\partial X_{t-j})$

- ⇒ on peut identifier les retards de X_t entrant dans le modèle de Y_t

Pre-whitening

- Mais cette identification n'est possible que si $X_t \sim WN(0, \sigma_X^2)$

⇒ Si ce n'est pas le cas, il faut blanchir X_t de toute dépendance ...

... sans affecter sa relation avec Y_t , c'est le **pre-whitening**

- Pour comprendre cette étape considérons à présent que $X_t \sim ARMA(p, q)$

$$Y_t = \alpha(L)X_t + \nu_t, \quad \Phi_X(L)X_t = \Theta_X(L)u_t$$

où $\nu_t \sim ARMA(p, q)$ indépendant $\forall t$ de X_t , u_t est un bruit blanc et

$$u_t = \Theta_X(L)^{-1}\Phi_X(L)X_t$$

Il suffit alors de multiplier Y_t par le filtre $\Theta_X(L)^{-1}\Phi_X(L)$ pour obtenir

$$\Theta_X(L)^{-1}\Phi_X(L)Y_t = \alpha(L)\Theta_X(L)^{-1}\Phi_X(L)X_t + \Theta_X(L)^{-1}\Phi_X(L)\nu_t$$

ce qui nous donne

$$\tilde{Y}_t = \alpha(L)u_t + \tilde{\nu}_t$$

où u_t est bien un bruit blanc et les $\alpha(L)$ sont donc identifiables

Estimation en étapes

- L'estimation des ARMAX nécessite donc plusieurs étapes
 - identification du processus ARMA afférent X_t pour déterminer
$$\Theta_X(L)^{-1}\Phi_X(L)$$
 - blanchiment de X_t par application du filtre $\Theta_X(L)^{-1}\Phi_X(L)$ à X_t et Y_t
 - calcul des corrélations croisées entre \tilde{Y}_t et u_t
 - ⇒ les corrélations non-nulles signalent les lags devant intégrer le modèle
 - identification du processus ARMA afférent à \tilde{Y}_t

Note 1 L'inférence sur les corrélations croisées est simplifiée puisque

$$u_t \sim WN(0, \sigma_u^2)$$

⇒ ce qui nous ramène à la formule de Bartlett simplifiée vue au S91

Fonction de transfert générale

- Soit $\mathbf{X}_t = X_{1,t}, \dots, X_{k,t}$ dont tous les éléments sont orthogonaux
- La fonction de transfert d'un ARMAX sur Y_t sera alors

$$Y_t = \frac{\alpha_1(L)}{\beta_1(L)} X_{1,t-d_1} + \dots + \frac{\alpha_k(L)}{\beta_k(L)} X_{k,t-d_k} + \frac{\Theta_Y(L)}{\Phi_Y(L)} \varepsilon_t$$

où $X_{i,t-d_i}$ indique que l'exogène rentre dans le modèle avec un retard qui lui est propre

Note 1 Les polynômes $\beta_k(L)$ ajoute de la généralité mais le plus souvent

$$\beta(L) = \Phi_Y(L)^{-1}, \quad \forall k$$

Note 2 En effet rappelons que si l'on part du modèle

$$\Phi_Y(L)Y_t = \alpha(L)\mathbf{X}_t + \Theta_Y(L)\varepsilon_t$$

et donc

$$Y_t = \alpha(L)\Phi_Y(L)^{-1}\mathbf{X}_t + \Theta_Y(L)\Phi_Y(L)^{-1}\varepsilon_t$$

Exemple de fonction de transfert

- Soit $Y_t = 0.6Y_{t-1} + 1.2X_{t-2} + \varepsilon_t$ et $X_t = 0.8X_{t-1} + u_t + 0.5u_{t-1}$

— En utilisant la représentation en fonction de transfert on obtient

$$Y_t = \frac{1.2}{1 - 0.6L} X_{t-2} + \frac{1}{1 - 0.6L} \varepsilon_t$$

— Les divisions polynômiales engendrent une décroissance infinie

$$\begin{aligned} Y_t &= 1.2(1 + 0.6L + 0.6^2L^2 + 0.6^3L^3 + \dots)X_{t-2} \\ &\quad + (1 + 0.6L + 0.6^2L^2 + 0.6^3L^3 + \dots)\varepsilon_t \\ &= 1.2X_{t-2} + 0.72X_{t-3} + 0.432X_{t-4} + 0.2592X_{t-5} + \dots \\ &\quad + \varepsilon_t + 0.6\varepsilon_{t-1} + 0.36\varepsilon_{t-2} + 0.216\varepsilon_{t-3} + 0.1296\varepsilon_{t-4} + \dots \end{aligned}$$

dans la dépendance entre Y_t et X_{t-i} , $i > 2$

⇒ On s'attend donc à voir une fonction d'autocorrélation croisée nulle en $i = 0, 1$ et non-nulle pour $i > 2$

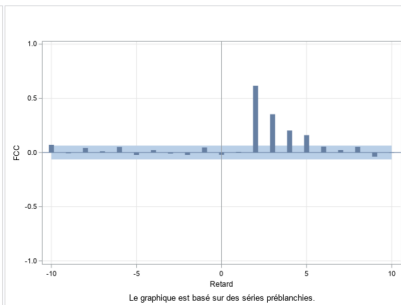
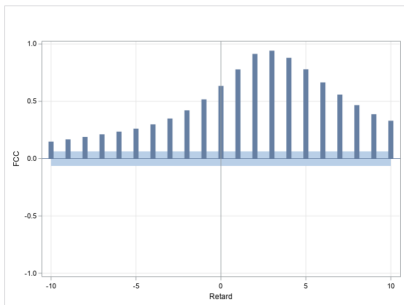
Importance du prewhitening

- Poursuivons avec l'exemple précédent et construisons \tilde{Y}_t et u_t

$$\tilde{Y}_t = (1 + 0.5L)^{-1}(1 - 0.8L)Y_t$$

$$u_t = (1 + 0.5L)^{-1}(1 - 0.8L)X_t$$

- Comparons à présent $\rho(\tilde{Y}_t, u_{t-i})$ et $\rho(Y_t, X_{t-i})$



Tests de validation

- Les procédures d'identification ne suffisent pas ...

... et doivent être couplées à des tests de validation sur les résidus :

- Test de significativité et nullité jointe (pour rappel)
- Coefficient de détermination
- Test de nullité de moyenne des résidus
- Tests de Ljung-Box et de Box-Pierce
- Test de Jarque-Bera

⇒ En effet, si le modèle est bien spécifié on devrait avoir $\hat{\varepsilon}_t \sim WN(0, \hat{\sigma}^2)$

Note Ces tests pourraient également s'appliquer sur les observées

Test de significativité et nullité jointe

- Sous certaines hypothèses de régularité on a vu que le MLE

$$\sqrt{n}(\hat{\vartheta} - \vartheta_0) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}(\hat{\vartheta}))$$

⇒ cela autorise la construction de tests inférenciels usuels car

$$\sqrt{n}((\hat{\vartheta} - \vartheta_0)/\sigma_{\hat{\vartheta}}) \sim \mathcal{N}(0, 1)$$

- Dans la pratique $\sigma_{\hat{\vartheta}}$ est inconnu et on considère $\hat{\sigma}_{\hat{\vartheta}}^2 \sim \chi^2(n-1)$

Rappel Si $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(\nu)$, alors $Z = X(\sqrt{Y/\nu})^{-1} \sim t(\nu)$

⇒ Le test de Student est alors ainsi formulé :

$$\frac{\hat{\vartheta} - \vartheta_{H_0}}{\hat{\sigma}_{\hat{\vartheta}}/n} \sim t(n-1)$$

- Le test de nullité jointe de Fisher s'applique de façon standard

⇒ si l'on souhaite tester

$$H_0 : \vartheta_1 = \dots = \vartheta_{p+q} = 0$$

contre $H_1 : \exists j$ tel que $\vartheta_j \neq 0$ la statistique de Fisher est

$$\frac{(SST - RSS)/(p+q)}{RSS/(n - (p+q) - 1)} \sim F(p+q, n - (p+q) - 1)$$

Coefficients de détermination

- Les coefficients de détermination usuels sont donnés par

$$R^2 = 1 - \frac{\sum_{t=1}^n \hat{\varepsilon}_t^2}{\sum_{t=1}^n X_t^2} \quad (4)$$

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-q} \frac{\sum_{t=1}^n \hat{\varepsilon}_t^2}{\sum_{t=1}^n X_t^2} \quad (5)$$

où l'on préférera \bar{R}^2 qui tient compte des retards AR et MA

Test de nullité de moyenne des résidus

- Si $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, on s'attend à ce que $\mathbb{E}(\hat{\varepsilon}_t) = 0$, et donc

$$\bar{\varepsilon} = n^{-1} \sum_{t=1}^n \hat{\varepsilon}_t \xrightarrow{p} 0$$

Par application du théorème central limite on a donc

$$n^{1/2} \frac{\bar{\varepsilon}}{\sigma_\varepsilon} \xrightarrow{d} \mathcal{N}(0, 1)$$

- On donc peut tester la nullité de $\bar{\varepsilon}$ en construisant

$$\bar{\varepsilon} \pm \Phi_{1-\alpha/2} n^{-1/2} \hat{\sigma}_\varepsilon$$

Tests de Ljung-Box et de Box-Pierce

- Si $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, on s'attend à ce que $\gamma(j) = \rho(j) = 0, \forall j > 0$
- Plutôt qu'étudier chaque IC autour de $\rho(j)$ comme au S90...
- ... on peut construire une seule statistique (de type portmanteau) :

$$Q = n \sum_{j=1}^h \hat{\rho}^2(j)$$

- La statistique de Box-Pierce étudie le comportement limite de Q

$$Q_{BP} = n \sum_{j=1}^h \hat{\rho}^2(j) \sim \chi^2(h - p - q)$$

si on test $H_0 : \rho_1 = \dots = \rho_h = 0$ contre $H_1 : \exists j$ tel que $\rho_j \neq 0$

- La statistique de Ljung-Box est un raffinement de Q où

$$Q_{LB} = n(n+2) \sum_{j=1}^h \hat{\rho}^2(j)/(n-j) \sim \chi^2(h - p - q)$$

si on test $H_0 : \rho_1 = \dots = \rho_h = 0$ contre $H_1 : \exists j$ tel que $\rho_j \neq 0$

Note Si ces tests sont appliqués sur les observés, les distributions sont $\chi^2(h)$

Test de Jarque-Bera

- Si on suppose que $\varepsilon_t \sim \text{i. i. d. } \mathcal{N}(0, \sigma_\varepsilon^2)$, on peut tester cette hypothèse
- Le test de Jarque-Bera permet cela à travers une unique statistique

$$S_{JB} = \frac{n}{6} S_k + \frac{n}{24} (K_u - 3)^2 \xrightarrow{d} \chi^2(2)$$

où S_k et K_u représentent les coefficients de Skewness et Kurtosis resp.

⇒ Si $S_{JB} \geq \chi_{1-\alpha}^2(2)$ on rejette H_0 de normalité des résidus au seuil de $\alpha\%$

Méthode du Maximum de Vraisemblance

- Partons d'un exemple: soit un échantillon $X_t = X_1, \dots, X_n \sim P(\vartheta)$

— $P(\vartheta)$ dénote la distribution de Poisson dont la fonction de masse est

$$\Pr(X_i = x) = \frac{\exp(-\vartheta)\vartheta^x}{x!}, \quad \vartheta > 0, \quad \forall x \in \mathbb{N}$$

- Soit une réalisation de l'échantillon $x_t = x_1, \dots, x_n$
- La probabilité d'observer cette réalisation est

$$\Pr((X_1 = x_1) \cap \dots \cap (X_n = x_n))$$

- L'indépendance des tirages donne l'équivalence avec le produit des probabilités marginales

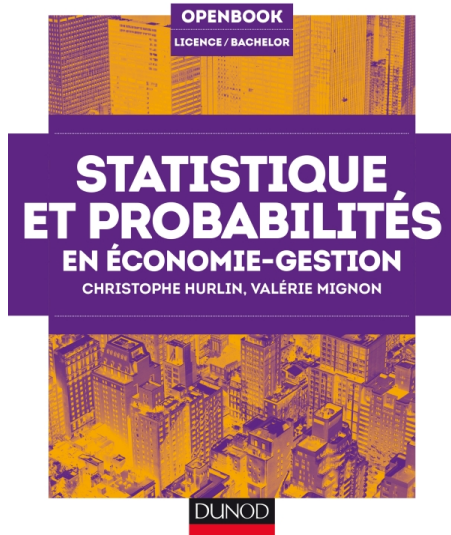
$$\Pr((X_1 = x_1) \cap \dots \cap (X_n = x_n)) = \prod_{i=1}^n \Pr(X_i = x_i)$$

Chapter 3

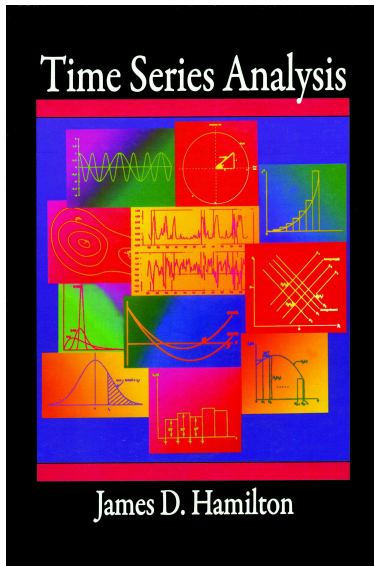
Les Processus Nonstationnaires



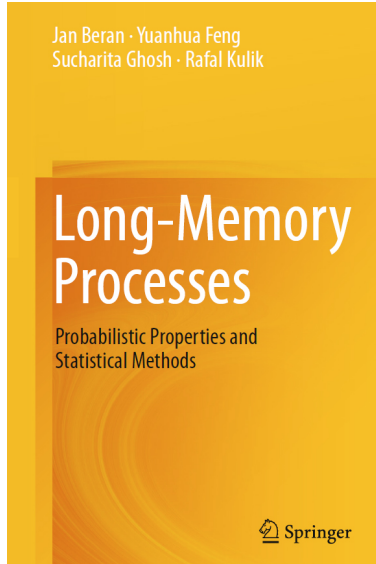
Références



Références



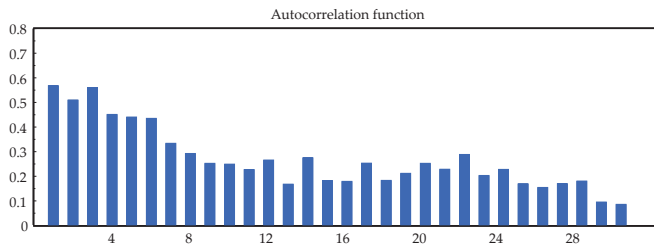
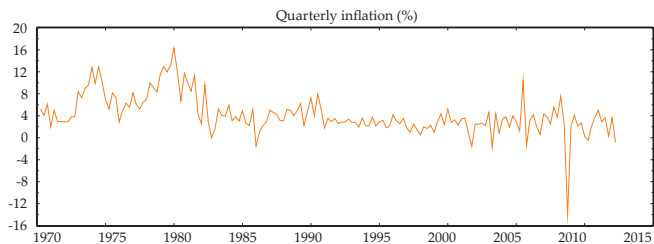
Références



Non-linéarité et cycles

- Question centrale: la non-stationnarité des séries temporelles
 - De nombreuses de séries économiques semblent non-stationnaires
 - La non-stationnarité est complexe à définir car protéiforme
 - La non-stationnarité engendre des complications importantes dans la théorie limite des estimateurs
- Quelques exemples:
 - Inflation: Non-stationnarité globale/ locale? Racine unitaire?
 - Volatilité financière: Stationnarité? Non-stationnarité?

Variables macroéconomiques

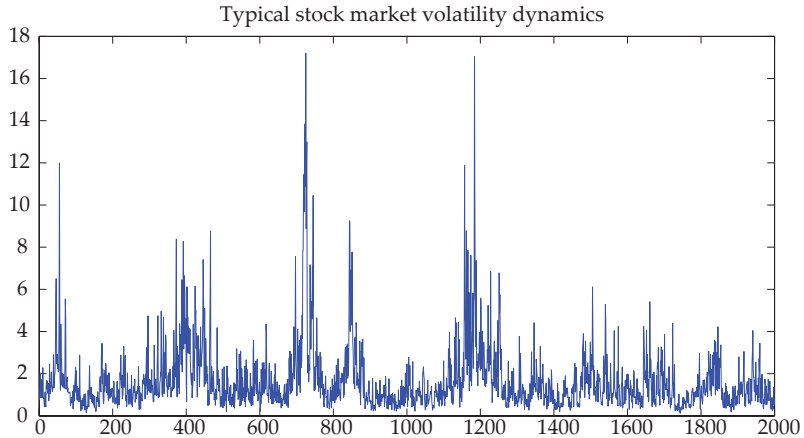


Quarterly US consumer price inflation 1970:1–2012:2 and its autocorrelation function

Dynamique persistante

- Difficile de déterminer visuellement si l'inflation est non-stationnaire
 - La fonction d'autocorrélation de l'inflation indique une très forte persistance
 - Pour autant, la série semble stable autour d'une moyenne de long terme

Variables financières



Dynamique extrême

- Difficile de déterminer visuellement si la volatilité est non-stationnaire
 - La série est animée par de nombreux événements extrêmes
 - Pour autant, la série semble stable autour d'une moyenne de long terme

Définir la non-stationnarité

- Plusieurs **phénomènes** peuvent mener à la **non-stationnarité**
 - ⇒ tendance déterministe
 - ⇒ saisonnalité
 - ⇒ tendance stochastique
 - ⇒ non-linéarités (dépasse le niveau M1)
 - ⇒ chocs de grandes valeurs et non-causalité (dépasse le niveau M1)

Tendance déterministe

- Il existe de nombreuses façons de violer les hypothèses d'ergodicité et de stationnarité
- Il existe donc tout autant de type de processus non-stationnaires
- Le plus simple d'entre eux: la **tendance déterministe linéaire**

- $X_t = \mu + \delta t + \varepsilon_t$ avec $\varepsilon_t \sim i.i.d. (0, \sigma_\varepsilon^2 < \infty)$

- on voit immédiatement que $\mathbb{V}(X_t) = \sigma_\varepsilon^2$ mais $\mathbb{E}(X_t) = \mu + \delta t$

- ⇒ l'espérance dépendant du temps, X_t est non-stationnaire

- Ce raisonnement tient pour des fonctions **nonlinéaires** du temps

$$X_t = \mu + \delta(t) + \varepsilon_t$$

avec e.g. $\delta(\cdot)$ une fonction polynomiale

Tendance déterministe et estimation

- Soit $X_t = \phi_0 + \delta t + \phi_1 X_{t-1} + \varepsilon_t$ un AR(1) avec tendance et $|\phi_1| < 1$
- Malgré sa non-stationnarité, ce modèle peut s'estimer par OLS car

$$Z_t = \left(\phi_0 - \delta \frac{\phi_1}{1 - \phi_1} \right) + \phi_1 Z_{t-1} + \varepsilon_t$$

est une réécriture, valide si $|\phi_1| < 1$, qui s'obtient en considérant

$$Y_t = X_t - \delta \left(\sum_{i=1}^{t-1} \phi_1^i \right) \times t$$

i.e. l'écart à la tendance, dont on prendra l'espérance de la limite

⇒ la théorie limite reste standard et les tests usuels s'appliquent

Note Cela peut être généralisé au cas d'un AR(p)

Tendance déterministe et prévision

- La tendance affecte trivialement l'espérance conditionnelle

$$\widehat{X}_{n+h} = \mathbb{E}(X_{n+h}|X_n, \dots, X_1) = \delta(n)_{n+h} + \mathbb{E}(\tilde{X}_{n+h}|X_n, \dots, X_1)$$

où $\delta(n)_{n+h}$ est la prévision de la tendance pour $n + h$ et

$$\tilde{X}_t = X_t - \delta(t)$$

- Par sa nature déterministe, la tendance n'affecte pas la variance

$$\mathbb{V}(X_{n+h}|X_n, \dots, X_1) = \mathbb{V}(\tilde{X}_{n+h}|X_n, \dots, X_1)$$

Tendance stochastique et racine unitaire

- Soit un AR(1): $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d. (0, \sigma_\varepsilon^2 < \infty)$
- Si $\rho = 1$ on a $X_t = X_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$ (supposons $X_0 = 0$) et donc

$$\text{Cov}(X_t, X_{t-j}) = (t-j)\sigma_\varepsilon^2 \text{ et } \mathbb{V}(X_t) = \mathbb{V}\left(\sum_{j=0}^{t-1} \varepsilon_{t-j}\right) = t\sigma_\varepsilon^2$$

⇒ La variance de X_t dépend de t ⇒ X_t est **non-stationnaire**

⇒ $\rho = 1$ place une solution du polynôme de retard sur le cercle unité donc X_t est un processus **racine unitaire**

⇒ X_t est une **marche aléatoire** d'espérance conditionnelle

$$\mathbb{E}(X_t | X_{t-1}, X_{t-2}, \dots, X_0) = X_{t-1} = \sum_{j=1}^{t-1} \varepsilon_{t-j} = \text{stochastic trend}$$

⇒ X_t est une **martingale** de **tendance stochastique** $\sum_j \varepsilon_{t-j}$

Tendance stochastique et différenciation

- Soit un AR(1): $X_t = \mu + \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d. (0, \sigma_\varepsilon^2 < \infty)$
- Si $\rho = 1$ et $X_0 = 0$, la représentation MA(∞) nous donne

$$X_t = \rho^t X_0 + \mu \sum_{j=0}^t \rho^j + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} = \mu t + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

- On constate alors que

$$\mathbb{E}(X_t) = \mu t$$

- La **variance** et l'**espérance** ne sont pas indépendantes de t

⇒ X_t est un processus **explosif** de type **marche aléatoire** et de **dérive** μ

- X_t est **stationnaire** en première différence car

$$\Delta X_t = (1 - L)X_t = X_t - X_{t-1} = \mu + \varepsilon_t$$

Non-stationnarité globale

- Les processus de type tendance **déterministe** ou **stochastique** sont **globalement non-stationnaires**
 - le processus viole les conditions de stationnarité
 - les paramètres du processus sont invariants
- ⇒ la non-stationnarité existe pour toute évolution du processus

Non-stationnarité locale

- Les processus dont les paramètres évoluent dans le temps sont possiblement **localement non-stationnaires**
- Par exemple, un modèle à changement de régime peut être
 - stationnaire dans un régime
 - non-stationnaire dans autre régime (localement non-stationnaire)

⇒ globalement stationnaire ou non-stationnaire
- Ce type de processus étant non-linéaire ⇒ cours d'économétrie non-linéaire

Stationnarité locale

- Les processus dont les paramètres évoluent dans le temps sont possiblement **localement stationnaires**
- Par exemple, un modèle à coefficients aléatoires dépendant du temps
 - est non-linéaire et **globalement non-stationnaire**
 - peut s'approximer localement par des processus stationnaires
 - ⇒ concept de **stationnarité locale** (dépasse le niveau M2)
- Références: Rao (2006), Dahlhaus et Rao (2006)

ARMA stationnaire

- Au Chapitre 1 nous avons introduit les ARMA leurs propriétés
- Supposons $X_t \sim ARMA(p, q)$

$$\begin{aligned} X_t &= \sum_{j=1}^p \phi_j X_{t-j} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2 < \infty) \\ &= \frac{1 - \theta_1 L + \dots + \theta_q L^q}{1 - \phi_1 L + \dots + \phi_p L^p} \varepsilon_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t = \psi(L) \varepsilon_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \end{aligned}$$

dont les racines du polynôme

$$\Phi(L) = (1 - \phi_1 L + \dots + \phi_p L^p)$$

ne sont pas situées sur le cercle unité et avec L l'opérateur retard

- On sait que sous cette condition de stationnarité

$$\sum_{i=0}^{\infty} |\psi_j| < \infty$$

Rappel sur l'ergodicité des ARMA

- Or, à partir de la forme $MA(\infty)$ de X_t on a

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

- En valeur absolue et par l'inégalité triangulaire on obtient

$$|\gamma(h)| = \sigma_\varepsilon^2 \left| \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \right| \leq \sigma_\varepsilon^2 \sum_{j=0}^{\infty} |\psi_j \psi_{j+h}|$$

- On constate alors que les ARMA sont **ergodiques** car

$$\sum_{h=0}^{\infty} |\gamma(h)| \leq \sigma_\varepsilon^2 \sum_{h=0}^{\infty} \sum_{j=0}^{\infty} |\psi_j \psi_{j+h}| \quad (6)$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} |\psi_j| \sum_{h=0}^{\infty} |\psi_{j+h}| < \infty, \quad (7)$$

et la convolution de deux séries absolument sommables est absolument sommable

Integration

Definition 18 (18)

Un processus stationnaire d'autocovariance $\gamma(h)$ absolument sommable est dit **faiblement dépendant**

Definition 19 (19)

Si un processus est **faiblement dépendant** après δ différenciations, il est dit **intégré d'ordre δ** ou $I(\delta)$

- Exemple: si $X_t = \mu + X_{t-1} + \varepsilon_t$ avec $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty)$

$\Rightarrow \Delta X_t = \mu + \varepsilon_t$ est **faiblement dépendant**

$\Rightarrow X_t$ est intégré d'ordre 1 également noté $X_t \sim I(1)$

$\Rightarrow \Delta X_t$ est intégré d'ordre nul également noté $\Delta X_t \sim I(0)$

ARIMA

Definition 20 (20)

Un processus X_t est un ARIMA(p, δ, q) si une fois différencié δ fois il peut s'exprimer comme un processus stationnaire et invertible de type ARMA(p, q)

$$\phi(L)\Delta^\delta X_t = \alpha + \theta(L)\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2 < \infty)$$

- Exemple: en cas de racine unitaire on a $\delta = 1$ et donc

$$\phi(L)\Delta X_t = \alpha + \theta(L)\varepsilon_t \Leftrightarrow \Delta X_t = \mu + \psi(L)\varepsilon_t$$

avec $\mu = \phi(L)^{-1}\alpha$ et $\psi(L) = \phi(L)^{-1}\theta(L)$

- Dans le même esprit que la marche aléatoire on observe que

$$X_t = X_0 + \mu t + \psi(L) \sum_{j=1}^{\infty} \varepsilon_{t-j}$$

⇒ Dans les cas simples $\delta \in \mathbb{Z}$, mais que se passe-t-il si $\delta \in \mathbb{R}$?

Intégration fractionnaire

- Soit un processus $(1 - L)^\delta X_t = \varepsilon_t$ avec $\delta > -1/2$
 - X_t est un **bruit blanc fractionnaire** si $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2 < \infty)$
- Une représentation $MA(\infty)$ nous donne

$$X_t = (1 - L)^{-\delta} \varepsilon_t = \sum_{j=0}^{\infty} \frac{\Gamma(j + \delta)}{\Gamma(j + 1)\Gamma(\delta)} L^j \varepsilon_t = \sum_{j=0}^{\infty} \alpha_j(-\delta) \varepsilon_{t-j}$$

- Le **polynôme de différenciation fractionnaire** $(1 - L)^{-\delta}$

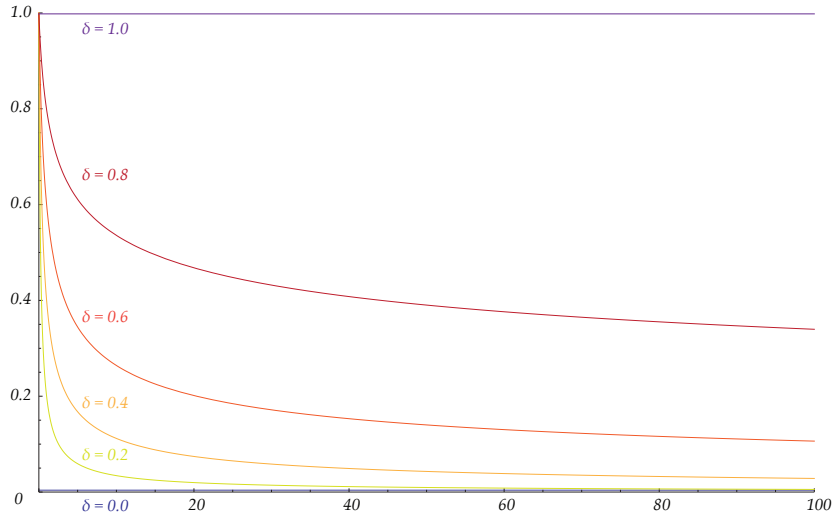
$$(1 - L)^{-\delta} = 1 + \delta L + \frac{\delta(1 + \delta)}{2!} L^2 + \frac{\delta(1 + \delta)(2 + \delta)}{3!} L^3 + \dots$$

- chaque numérateur est un **symbole de Pochhammer**

$$(\delta)_n = \delta(\delta + 1)(\delta + 2) \dots (\delta + n - 1) = \Gamma(n + \delta)\Gamma(\delta)^{-1}$$

- chaque dénominateur est une **factorielle**, or $n! = \Gamma(n + 1)$
- le j -ième terme de la suite est donc $(\delta)_j \times \Gamma(j + 1)^{-1} L^j$

Filtre fractionnaire



Analyse du filtre fractionnaire

- Une analyse à la limite nous apprend que

$$\alpha_j(-\delta) \sim \frac{1}{\Gamma(\delta)} j^{-1+\delta} \left(1 + O(j^{-1})\right), \quad j \rightarrow \infty$$

- Granger et Joyeux (1980) soulignent alors que

- si $\delta = 1/2$ on a $\sum_{j=1}^{\infty} \alpha_j(-\delta)^2 = \infty$

⇒ en effet $\alpha_j(-\delta)^2 \approx j^{-1}$ décrit une **série Harmonique** (divergente)

- si $\delta > 1/2$ on a $\sum_{j=1}^{\infty} \alpha_j(-\delta)^2 = \infty$

⇒ $j^{2(\delta-1)=p>-1}$ est une p -série (hyper-Harmonique) divergente

- si $\delta < 1/2$ on a $\sum_{j=1}^{\infty} \alpha_j(-\delta)^2 < \infty$

⇒ $j^{2(\delta-1)=p<-1}$ est une p -série (hyper-Harmonique) convergente

Bruit blanc fractionnaire et non-stationnarité

- Une représentation $MA(\infty)$ alternative de X_t est alors

$$X_t = \sum_{j=0}^{\infty} \alpha_j(-\delta) \varepsilon_{t-j} \approx \frac{1}{\Gamma(\delta)} \left(\sum_{j=1}^{\infty} j^{-1+\delta} \varepsilon_{t-j} + \varepsilon_t \right)$$

- La variance apparaît alors sous une forme simple

$$\mathbb{V}(X_t) = \sigma_\varepsilon^2 \left(\frac{1}{\Gamma^2(\delta)} + \sum_{j=1}^{\infty} \alpha_j(-\delta)^2 \right) = \frac{1}{\Gamma^2(\delta)} \sigma_\varepsilon^2 \left(1 + \sum_{j=1}^{\infty} j^{2(\delta-1)} \right)$$

- $\mathbb{V}(X_t)$ sera finie si $\delta < 1/2 \Rightarrow X_t$ **stationnaire**
 - $\mathbb{V}(X_t)$ sera infinie si $\delta \geq 1/2 \Rightarrow X_t$ **non-stationnaire**
- Il existe une formulation non-asymptotique de la variance

$$\mathbb{V}(X_t) = \sigma_\varepsilon^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)} = \sigma_\varepsilon^2 \nu_\delta$$

où $\nu_\delta = 1$ si $\delta = 0$ et $\nu_\delta = \infty$ si $\delta \geq 1/2$

Bruit blanc fractionnaire et autocovariance

- Les calculs d'autocovariance obtenus par Hosking (1981) donnent

$$\begin{aligned}\gamma(h) &= \mathbb{E}\left(X_t X_{t-h}\right) \\ &= \sigma_\varepsilon^2 \frac{\Gamma(h+\delta)\Gamma(1-2\delta)}{\Gamma(h+1-\delta)\Gamma(1-\delta)\Gamma(\delta)} \\ &= \frac{(\sigma_\varepsilon^2/2\pi) \sin(\pi\delta)\Gamma(h+\delta)\Gamma(1-2\delta)}{\Gamma(h+1-\delta)}\end{aligned}$$

avec $\delta \in (-1/2, 1/2)$

- Approximation asymptotique de Lieberman et Phillips (2008)

$$\gamma(h) \sim \frac{\sigma_\varepsilon^2}{\pi} \frac{\Gamma(1-2\delta) \sin(\pi\delta)}{h^{1-2\delta}} + O(h^{2\delta-3}), \quad h \rightarrow \infty$$

avec $\delta \in (-1/2, 0) \cup (0, 1/2)$

Bruit blanc fractionnaire et sommabilité

- Via Lieberman et Phillips (2008) on constate que

$$\gamma(h) \sim Ch^{2\delta-1}, \quad C > 0$$

- Or pour $\delta \in (0, 1/2)$ on a $2\delta - 1 \in (-1, 0)$ et donc

$$\sum_{j=0}^{\infty} |\gamma(j)| = \infty \text{ bien que } \mathbb{V}(X_t) < \infty$$

- En revanche, pour $\delta \in (-1/2, 0)$, $2\delta - 1 \in (-2, -1)$ et donc

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty$$

- Si $\delta \in (-1/2, 0)$, $\gamma(j)$ est absolument sommable
 $\Rightarrow X_t$ est dit **faiblement dépendant** et à **mémoire courte**
- Si $\delta \in (0, 1/2)$, X_t , $\gamma(j)$ n'est pas absolument sommable
 $\Rightarrow X_t$ est dit **fortement dépendant** et à **mémoire longue**

Bruit blanc fractionnaire et mean reversion

Definition 21 (21)

Un processus X_t est **mean-reverting** (revient vers sa moyenne) si sa fonction de réponse impulsionnelle cumulée (CIR) à l'infini tend vers zéro.

- Soit $(1 - L)^\delta X_t = \varepsilon_t$ un bruit blanc fractionnaire avec $\delta > 1/2$
- Considérons à présent la représentation $MA(\infty)$ de $(1 - L)X_t$

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

- On a alors l'impact d'un choc unitaire en t sur X en $t + h$

$$CIR_h = \sum_{j=0}^h \alpha_j (1 - \delta) \sim C \sum_{j=1}^h j^{\delta-2}, \quad C > 0$$

dont la convergence est vérifiée pour $\delta < 1$ quand $h \rightarrow \infty$

ARFIMA

- On peut généraliser le bruit blanc fractionnaire en ajoutant de la dynamique de court terme

Definition 22 (22)

Un processus X_t est un ARFIMA(p, δ, q) si une fois différencié δ fois il peut s'exprimer comme un processus stationnaire et invertible de type ARMA(p, q)

$$\phi(L)\Delta^\delta X_t = \alpha + \theta(L)\varepsilon_t$$

Note 1 Les démonstrations précédentes tiennent pour un ARFIMA bien que les formules se complexifient

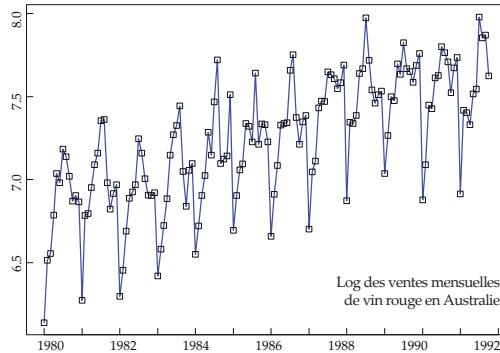
Note 2 Pour $\delta = 1$, une racine unitaire est présente et on parle d'ARIMA

ARFIMA

- A la différence des processus ARIMA(p, δ, q)
 - $\Rightarrow X_t$ est stationnaire à mémoire longue si $\delta < 1/2$
 - $\Rightarrow X_t$ est non-stationnaire à mémoire longue si $\delta > 1/2$
 - $\Rightarrow X_t$ est non-stationnaire mean-reverting si $1/2 \leq \delta < 1$
 - $\Rightarrow X_t$ est stationnaire à mémoire courte si $\delta \leq 0$
 - $\Rightarrow X_t$ est dit anti-persistent si $\delta < 0$
 - $\Rightarrow X_t$ possède une représentation MA(∞) si $\delta > -1/2$
 - $\Rightarrow X_t$ possède une représentation AR(∞) si $\delta < 1/2$
- Les deux derniers points sont discutés par Hosking (1981)

Saisonnalité

- La saisonnalité stochastique n'implique pas la non-stationnarité
⇒ mais s'analyse à travers un modèle général : Seasonal ARIMA
- e.g. les ventes mensuelles de vin montrent une forte saisonnalité
⇒ pic en juillet et creux en janvier en plus d'une tendance haussière



Dépendance saisonnière

- La dépendance au passé sera donc importante à 12 mois

$$\rho_{12}, \rho_{24}, \rho_{36}, \dots$$

entre juillet (janvier) de l'année en cours et l'année passée

⇒ Un modèle $\text{ARMA}(P, Q)_s$ troué de **span** $s = 12$ sera approprié

e.g. 1 dans le cas d'un $\text{MA}(1)_{12}$ on aura

$$X_t = (1 + \vartheta_1 L^{12})\varepsilon_t \Leftrightarrow X_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-12}$$

e.g. 2 dans le cas d'un $\text{AR}(1)_{12}$ on aura

$$(1 - \varphi_1 L^{12})X_t = \varepsilon_t \Leftrightarrow X_t = \varphi_1 X_{t-12} + \varepsilon_t$$

Note Le span dépend de la fréquence :

$$s = 2 \text{ (semestriel)}, s = 4 \text{ (trimestriel)}, s = 52 \text{ (hebdomadaire)}$$

Filtre saisonnier et SARMA

- Après application d'un $\text{ARMA}(P, Q)_s$, une dépendance peut rester
- L' $\text{ARMA}(P, Q)_s$ agit comme un filtre qui retourne Y_t

$$\Phi(L)_s X_t = \Theta(L)_s Y_t$$

- Un second filtre de type $\text{ARMA}(p, q)$ est donc nécessaire

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

afin que la dépendance disparaisse et de récupérer $\varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$

⇒ Le produit des deux filtres donne le modèle $\text{SARMA}(p, q) \times (P, Q)_s$

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

Note Les conditions de stationnarité et inversibilité usuelles s'appliquent

SARMA(0, 0) \times (0, 1)₄

- Il s'agit d'un SMA du type

$$X_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-4}$$

approprié pour des données trimestrielles par exemple

Type MA(q) \Rightarrow inversible si ses 4 racines

$$-i\sqrt[4]{\frac{1}{\vartheta_1}}, \quad i\sqrt[4]{\frac{1}{\vartheta_1}}, \quad \sqrt[4]{\frac{1}{\vartheta_1}}, \quad -\sqrt[4]{\frac{1}{\vartheta_1}}$$

sont en dehors du cercle unitaire et donc $|\vartheta_1| < 1$

- Si $|\vartheta_1| < 1$, X_t possèdent une forme AR(∞)

$$X_t = - \sum_{j=1}^{\infty} \vartheta_1^j X_{t-4j} + \varepsilon_t$$

- On peut aisément vérifier que $\gamma_0 = (1 + \vartheta_1^2)\sigma_\varepsilon^2$,

$$\gamma_4 = \vartheta_1 \sigma_\varepsilon^2$$

et $\gamma_j = 0$ pour $j \neq 0, 4$

SARMA(0, 1) \times (0, 1)₅

- Il s'agit d'un SARMA du type

$$X_t = (1 + \theta_1 L)(1 + \vartheta_1 L^5)\varepsilon_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \vartheta_1 \varepsilon_{t-5} + \theta_1 \vartheta_1 \varepsilon_{t-6}$$

approprié pour des données journalières (jours ouvrés)

Type MA(6) \Rightarrow inversible si ses 6 racines

... sont en dehors du cercle unitaire

- On peut aisément vérifier que

$$\gamma_0 = (1 + \theta_1^2 + \vartheta_1^2 + \theta_1^2 \vartheta_1^2) \sigma_\varepsilon^2 = (1 + \theta_1^2)(1 + \vartheta_1^2) \sigma_\varepsilon^2$$

$$\gamma_1 = (\theta_1 + \theta_1 \vartheta_1^2) \sigma_\varepsilon^2 = \theta_1 (1 + \vartheta_1^2) \sigma_\varepsilon^2$$

$$\gamma_4 = (\theta_1 \vartheta_1) \sigma_\varepsilon^2$$

$$\gamma_5 = (\vartheta_1 + \theta_1^2 \vartheta_1) \sigma_\varepsilon^2 = \vartheta_1 (1 + \theta_1^2) \sigma_\varepsilon^2$$

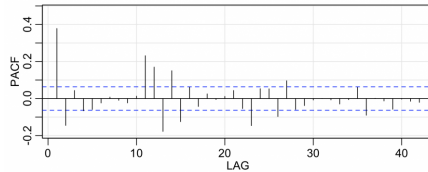
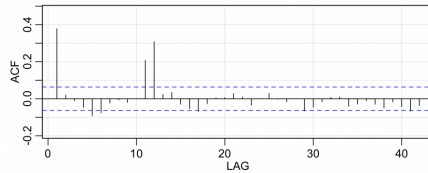
$$\gamma_6 = (\theta_1 \vartheta_1) \sigma_\varepsilon^2$$

et $\gamma_j = 0$ pour $j \neq 0, 1, 4, 5, 6$ et $j > 6$

SARMA(0, 1) \times (0, 1)₁₂

- Etudions le comportement (simulé) des ACFs et PACFs si

$$\theta_1 = 0.7 \text{ et } \vartheta_1 = 0.6 \text{ et } \sigma_\varepsilon^2 = 1$$



SARMA(0, q) × (0, Q)_s

- Quelques généralités se dégagent de l'exemple précédent
 - L'ACovF et l'ACF sont atypiques autour de s
- ⇒ Autour de s , les q autocovariances sont symétriques
- Comme pour tout MA(q), la PACF ne s'annule pas
- ⇒ Elle décroît vers 0 quand $h \rightarrow \infty$
- ⇒ Des résurgences apparaîtront autour $s, 2s, 3s, \dots$

SARMA(0, 1) \times (1, 0)₇

- Il s'agit d'un SARMA du type

$$(1 - \varphi_1 L^7)X_t = (1 + \theta_1 L)\varepsilon_t$$

approprié pour des données journalières (semaine pleine)

Type AR(7) + MA(1) \Rightarrow stationnaire et inversible si les 7 racines

... de $(1 - \varphi_1 L^7)$ sont en dehors du cercle unitaire et si $|\theta_1| < 1$

- On peut vérifier que

$$\gamma_0 = (1 + \theta_1^2)(1 - \varphi_1^2)^{-1}\sigma_\varepsilon^2$$

$$\gamma_1 = \varphi_1 \mathbb{E}(X_{t-7}X_{t-1}) + \mathbb{E}(\varepsilon_t X_{t-1}) - \theta_1 \mathbb{E}(\varepsilon_{t-1} X_{t-1}) = \varphi_1 \gamma_6 - \theta_1 \sigma_\varepsilon^2$$

$$\gamma_j = \varphi_1 \gamma_{j-7}, \quad \forall j > 2$$

$$\gamma_6 = \varphi_1 \gamma_1 \Rightarrow \gamma_1 = -\theta_1 \sigma_\varepsilon^2 (1 - \varphi_1^2)^{-1}$$

$$\gamma_k = \varphi_1^j \gamma_1, \quad \forall k = 7j \pm 1, \quad j = 1, 2, \dots \text{ et}$$

$$\gamma_k = \varphi_1^j \gamma_0, \quad \forall k = 7j, \quad j = 0, 1, 2, \dots$$

et $\gamma_k = 0$ pour les k non considérés ci-dessus

SARMA(1, 1) × (2, 1)₁₂

- Ici les filtres sont multiplicatifs mais il existe des filtres additifs

⇒ Cela va impacter l'écriture des polynômes

e.g. L'écriture d'un SARMA(1, 1) × (2, 1)₁₂ sera

$$(1 - \varphi_1 L)(1 - \phi_1 L^{12} - \phi_2 L^{24})X_t = (1 + \theta_1 L)(1 + \vartheta_1 L^{12})\varepsilon_t$$

e.g. L'écriture d'un SARMA(1, 1) + (2, 1)₁₂ sera

$$(1 - \varphi_1 L - \phi_1 L^{12} - \phi_2 L^{24})X_t = (1 + \theta_1 L + \vartheta_1 L^{12})\varepsilon_t$$

SARIMA(p, δ, q) \times (P, D, Q) $_s$

- On a vu que Δ^δ était important pour stationnariser une série
- On peut également construire un opérateur

$$\Delta_s^D = (1 - L^s)^D$$

de différenciation saisonnière

- En effet, certaines séries sont cyclo-stationnaires et avec $D = 1$, si

$$\Delta_s X_t = X_t - X_{t-s}$$

est stationnaire, c'est que X_t était une marche aléatoire saisonnière

e.g. $X_{01/2021} = X_{01/2020} + \varepsilon_{01/2021}$ suit une marche aléatoire saisonnière

- Plus généralement, $X_t \sim \text{SARIMA}(p, \delta, q) \times (P, D, Q)_s$ si

$$Y_t = \Delta^\delta \Delta_s^D X_t \sim \text{SARMA}(p, q) \times (P, Q)_s$$

X13-ARIMA-SEATS

- La filtration de la saisonnalité est une tâche fastidieuse
- Des procédures de traitement automatique existent depuis 1970

⇒ X11-ARIMA, TRAMO-SEATS, X12-ARIMA (dispo sous SAS)

Réf. Une revue de ses méthodes est faite par Darné (2004)

- La plus récente et populaire : la procédure X13-ARIMA-SEATS

⇒ basée sur des régressions de SARIMA, elle peut aussi traiter

... des effets très spécifiques comme

- valeurs aberrantes
- jours ouvrables
- effets de Pâques
- traitement spectral de la saisonnalité résiduelle
- ...

Rappels sur les OLS

- La non-stationnarité impact la **théorie limite** des estimateurs
- Pour comprendre cela repartons des OLS dans le modèle linéaire

$$Y_t = X_t\beta + \varepsilon_t \quad \varepsilon_t \sim \text{i. i. d. } (0, \sigma_\varepsilon^2)$$

- Supposons $X_t \perp \varepsilon_t$
- L'estimateur OLS est alors donné par

$$\begin{aligned}\hat{\beta} &= \frac{\text{Cov}(Y, X)}{\mathbb{V}(X)} = \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right)^{-1} \frac{1}{n} \sum_{t=1}^n X_t Y_t \\ &= \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right)^{-1} \frac{1}{n} \sum_{t=1}^n (\beta X_t + \varepsilon_t) X_t = \beta + \frac{n^{-1} \sum_{t=1}^n X_t \varepsilon_t}{n^{-1} \sum_{t=1}^n X_t^2}\end{aligned}$$

Consistance des OLS

- Analysons la **consistance** de $\hat{\beta}$ si $X_t \sim$ i. ni. d. ($0, \sigma_X^2 < \infty$)

$$\text{plim } \hat{\beta} = \beta + \frac{\text{plim } n^{-1} \sum_{t=1}^n X_t \varepsilon_t}{\text{plim } n^{-1} \sum_{t=1}^n X_t^2}$$

- On observe que $\mathbb{E}(X_t \varepsilon_t) = \mathbb{E}(X_t) \mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{V}(X_t \varepsilon_t) = \sigma_X^2 \sigma_\varepsilon^2$

- $\mathbb{V}(X_t \varepsilon_t) = \mathbb{E}(X_t)^2 \mathbb{E}(\varepsilon_t)^2 - \mathbb{E}(X_t)^2 \mathbb{V}(\varepsilon_t) - \mathbb{E}(\varepsilon_t)^2 \mathbb{V}(X_t) + \mathbb{V}(X_t) \mathbb{V}(\varepsilon_t)$

- dans la LLN de Markov (10), $\sum \sigma_X^2 \sigma_\varepsilon^2 / t^2 < \infty$ pour $m = 1$

$$\Rightarrow n^{-1} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{p} \mathbb{E}(X_t \varepsilon_t) = 0$$

- On suppose que $\mathbb{E}(X_t^4) < \infty$ existe et par la LLN de Markov,

$$n^{-1} \sum_{t=1}^n X_t^2 \xrightarrow{p} \sigma_X^2$$

- Par application du théorème de **Mann-Wald** (13) on a

$$\text{plim } \hat{\beta} = \beta + \frac{0}{\text{plim } n^{-1} \sum_{t=1}^n X_t^2} = \beta$$

Distribution limite des OLS

- La consistance donne une distribution dégénérée: $\widehat{\beta} \xrightarrow{p} \beta$

- Avec un transformation $\mathbb{T}(\widehat{\beta}) \xrightarrow{d} \mathcal{L}$ on peut montrer que

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2)$$

- Commençons par analyser $n^{-1} \sum_{t=1}^n X_t \varepsilon_t$ sachant $\mathbb{E}(X_t \varepsilon_t) = 0$ et

$$\mathbb{V}(X_t \varepsilon_t) = \sigma_X^2 \sigma_\varepsilon^2$$

⇒ L'application du TCL de **Lyapunov** (12) donne

$$\sqrt{n}(n^{-1} \sum_{t=1}^n X_t \varepsilon_t - 0) = n^{-1/2} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma_X^2 \sigma_\varepsilon^2)$$

- La consistance a montré que $n^{-1} \sum_{t=1}^n X_t^2 \xrightarrow{p} \sigma_X^2$

⇒ D'après le théorème de **Slutsky** (14) on a alors

$$\sqrt{n}(\widehat{\beta} - \beta) = \sqrt{n} \frac{n^{-1} \sum_{t=1}^n X_t \varepsilon_t}{n^{-1} \sum_{t=1}^n X_t^2} \xrightarrow{d} \frac{\mathcal{N}(0, \sigma_X^2 \sigma_\varepsilon^2)}{\sigma_X^2} \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 \sigma_X^{-2})$$

Théorie limite dans le cas d'un AR(1)

- Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2)$ et $X_0 = 0$
- L'estimateur OLS est donné par

$$\hat{\rho} = \rho + \left(\sum_{t=1}^n X_t^2 \right)^{-1} \left(\sum_{t=1}^n X_{t-1} \varepsilon_t \right)$$

- Sous l'hypothèse que $|\rho| < 1$, X_t est stationnaire et

$$\frac{1}{n} \sum_{t=1}^n X_t^2 \xrightarrow{p} \sigma_X^2 = \mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right) = \frac{\sigma_\varepsilon^2}{1 - \rho^2} < \infty$$

⇒ D'après le théorème de **Slutsky** (14) on a alors

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} \frac{\mathcal{N}(0, \sigma_\varepsilon^2 \sigma_X^2)}{\sigma_X^2} \stackrel{d}{=} \mathcal{N}(0, 1 - \rho^2) \quad (8)$$

- Qu'observez-vous si $\rho = 1$?

Limite de la théorie asymptotique standard

- Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$, $X_0 = 0$ et $\rho = 1$
- X_t est donc une **marche aléatoire de variance** $\mathbb{V}(X_t) = t\sigma_\varepsilon^2$
- Impossible alors d'appliquer un TCL comme dans (8) puisque

$$\frac{1}{n} \sum_{t=1}^n X_t^2 \xrightarrow{p} \mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n X_t^2 \right) = \frac{1}{n} \sum_{t=1}^n \mathbb{E}(X_t^2) = \sigma_\varepsilon^2 \frac{\sum_{t=1}^n t}{n} \rightarrow \infty$$

- Pour autant une **théorie limite non-standard** est possible si

$$\mathbb{T}(\hat{\rho}) = n(\hat{\rho} - \rho) \neq \sqrt{n}(\hat{\rho} - \rho)$$

⇒ Théorie asymptotique applicable dans le cas **non-stationnaire**

- Ce type de théorie fait intervenir les **processus de Wiener**

Marche aléatoire et TCL

- Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i. i. d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$
- On a vu que si $\rho = 1$ alors $X_n = \sum_{t=1}^n \varepsilon_t$
- On considère à présent

$$\mathbb{T}(X_n) = \sqrt{n} \frac{1}{n} X_n \sigma_{\mathbb{T}(X)}^{-1} = \sqrt{n}(\bar{\varepsilon}) \sigma_{\mathbb{T}(X)}^{-1} = n^{-1/2} X_n \sigma_{\mathbb{T}(X)}^{-1}$$

avec

$$\sigma_{\mathbb{T}(X)}^2 = \mathbb{V}(\mathbb{T}(X_n)) = \mathbb{E} \left(n^{-1} \left(\sum_{t=1}^n \varepsilon_t \right)^2 \right) = n^{-1} n \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \quad (9)$$

- D'après le TCL de Lindeberg-Lévy (11), on constate alors que

$$n^{-1/2} X_n \sigma_{\mathbb{T}(X)}^{-1} \xrightarrow{d} \mathcal{N}(0, 1)$$

Les sommes partielles

- Poursuivons ce raisonnement sur une somme partielle de X_n

$$X_n(r) = \sum_{s=1}^{[nr]} \varepsilon_s$$

avec $0 \leq r < 1$ et $[nr]$ le plus grand entier $\leq nr$

- Après transformation de $X_n(r)$ on a donc

$$W_n(r) = n^{-1/2} X_n(r) \sigma_{\mathbb{T}(X)}^{-1} = n^{-1/2} \sum_{s=1}^{[nr]} \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

... que l'on peut réécrire

$$W_n(r) = \underbrace{\left(n^{-1/2} [nr]^{1/2} \right)}_{\rightarrow r^{1/2}} \times \underbrace{\left([nr]^{-1/2} \sum_{s=1}^{[nr]} \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1} \right)}_{\xrightarrow{d} \mathcal{N}(0,1)}$$

$$\Rightarrow W_n(r) \xrightarrow{d} r^{1/2} \mathcal{N}(0, 1) \xrightarrow{d} \mathcal{N}(0, r)$$

Vers les processus de Wiener

- A présent définissons $r_t = t/n$ de telle sorte que $[nr_t] = t$ et

$$W_n(r) = n^{-1/2} \sum_{s=1}^t \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

- Si $t/n < r_i < (t+1)/n$ on a également $[nr_i] = t$ et donc

$$W_n(r) = n^{-1/2} \sum_{s=1}^t \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

- Mais si $r_i = (t+1)/n$ on a

$$W_n(r) = n^{-1/2} \sum_{s=1}^{t+1} \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1}$$

⇒ par définition $r \in [0, 1]$ et

$$W_n(1) = n^{-1/2} \sum_{s=1}^n \varepsilon_s \sigma_{\mathbb{T}(X)}^{-1} \tag{10}$$

Les processus de Wiener

- $W_n(1) \xrightarrow{d} W(1)$ avec $W(1)$ un processus de Wiener

Definition 23 (23)

Soit un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. $W : \Omega \times [0, 1] \rightarrow \mathbb{R}^1$ est un processus de Wiener standard si pour chaque $r \in [0, 1]$, $W(r)$ est \mathcal{F} -mesurable et si

le processus $W(r)$ débute en 0: $\mathbb{P}(W(0) = 0) = 1$

le processus $W(r)$ possède des incréments indépendants: si $0 \leq t_0 \leq t_1 \leq \dots \leq t_k \leq 1$, $W(t_i) - W(t_{i-1})$ est indépendant de $W(t_j) - W(t_{j-1})$ pour $j = 1, \dots, k, j \neq i, \forall i = 1, \dots, k$

le processus $W(r)$ possède des incréments normalement distribués: pour $0 \leq a \leq b \leq 1$, l'incrément $W(b) - W(a) \sim \mathcal{N}(0, b - a)$

- Mais vers quoi converge $W_n(r)$?

Théorème central limite fonctionnel

- $W_n(r)$ est une **fonction aléatoire** (via ε_t) de r

Theorem 4 (4: Théorème de Donsker)

Soit ε_t une séquence de variables aléatoires telle que $\varepsilon_t \sim$ i. i. d. $(0, \sigma_\varepsilon^2 < \infty)$, alors

$$W_n \xrightarrow{d} W$$

- Le TCL fonctionnel requiert (pas suffisant) la convergence point par point de $W_n(r)$

$$W_n(r) \xrightarrow{d} W(r)$$

Théorème de Mann-Wald fonctionnel

Theorem 5 (5: Functional Continuous Mapping Theorem)

Soit $X_n(\cdot) \xrightarrow{d} X(\cdot)$ une fonction aléatoire convergente et $g(\cdot)$ une fonction continue à valeur dans \mathbb{R} en $X(\cdot)$.
Alors,

$$X_n(\cdot) \xrightarrow{d} X(\cdot) \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$X_n(\cdot) \xrightarrow{p} X(\cdot) \Rightarrow g(X_n) \xrightarrow{p} g(X)$$

$$X_n(\cdot) \xrightarrow{a.s.} X(\cdot) \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

Théorie limite non standard des OLS

Theorem 6 (6)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, la distribution limite de l'estimateur OLS est donnée par

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{1/2\sigma_\varepsilon^2(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr}$$

- Il s'agit d'une distribution non-standard complexe à manipuler
- En présence de non-stationnarité, ce type de distribution survient souvent

Théorie limite non standard des OLS: démonstration

- L'estimateur OLS est donnée par

$$\hat{\rho} = \rho + \frac{n^{-1} \sum_{t=1}^n X_{t-1} \varepsilon_t}{n^{-1} \sum_{t=1}^n X_t^2}$$

- Si $\rho = 1$ et via une normalisation en n on a

$$n(\hat{\rho} - 1) = \frac{n^{-1} \sum_{t=1}^n X_{t-1} \varepsilon_t}{n^{-2} \sum_{t=1}^n X_t^2}$$

- A partir des lemmes (2) et (3) on obtient alors

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{1/2(W(1))^2 - 1}{\int_0^1 W(r)^2 dr}$$

- On peut également démontrer (un peu plus fastidieux) que

$$\hat{t}_{\rho} = \frac{(\hat{\rho} - 1)}{\hat{\sigma}_{\rho}} \xrightarrow{d} \frac{1/2(W(1))^2 - 1}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}} \quad (11)$$

Théorie limite non standard des OLS: lemmes

Lemma 1 (1)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i. i. d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, alors

$$n^{-1/2} \sum_{t=1}^n \varepsilon_t \xrightarrow{d} \sigma_\varepsilon W(1)$$

- Voir le CLT de l'équation (10) pour la démonstration

Théorie limite non standard des OLS: lemmes

Lemma 2 (2)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i. i. d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, alors

$$n^{-2} \sum_{t=1}^n X_{t-1}^2 \xrightarrow{d} \sigma_\varepsilon^2 \int_0^1 W(r)^2 dr$$

- Voir (12) pour la démonstration

Théorie limite non standard des OLS: lemmes

Lemma 3 (3)

Soit $X_t = \rho X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{i. i. d. } (0, \sigma_\varepsilon^2)$ et $X_0 = 0$. Si $\rho = 1$, alors

$$n^{-1} \sum_{t=1}^n X_{t-1} \varepsilon_t \xrightarrow{d} \frac{1}{2} \sigma_\varepsilon^2 (W(1)^2 - 1)$$

- Voir (13) pour la démonstration

Théorie limite non standard: Lemme 2

- Si $(t-1)/n \leq r_{t-1} < t/n$ on sait que

$$W_n(r) = n^{-1/2} X_{t-1} \sigma_{\mathbb{T}(X)}^{-1} \Rightarrow \sum_{t=1}^n W_n(r) = n^{-1/2} \sigma_{\mathbb{T}(X)}^{-1} \sum_{t=1}^n X_{t-1}$$

- On en déduit alors que $n^{-2} \sum_t X_{t-1}^2 = n^{-1} \sigma_{\mathbb{T}(X)}^2 \sum_{t=1}^n W_n(r)^2$
- On sait que $W_n(r)$ est constant si $(t-1)/n \leq r_{t-1} < t/n$ et donc

$$n^{-1} \sum_{t=1}^n W_n(r)^2 = \sum_{t=1}^n \int_{(t-1)/n}^{t/n} W_n(r)^2 dr = \int_0^1 W_n(r)^2 dr$$

- Le théorème de Mann-Wald fonctionnel nous assure alors que

$$n^{-2} \sum_t X_{t-1}^2 \xrightarrow{d} \sigma_{\mathbb{T}(X)}^2 \int_0^1 W(r)^2 dr \quad (12)$$

avec $\sigma_{\mathbb{T}(X)} = \sigma_\varepsilon$ (voir équation 9)

Théorie limite non standard: Lemme 3

- Si $\rho = 1$ dans X_t alors $X_t^2 = (X_{t-1} + \varepsilon_t)^2 = X_{t-1}^2 + 2X_{t-1}\varepsilon_t + \varepsilon_t^2$
- Il vient immédiatement que

$$X_{t-1}\varepsilon_t = \frac{1}{2}(X_t^2 - X_{t-1}^2 - \varepsilon_t^2) \Rightarrow \sum_{t=1}^n X_{t-1}\varepsilon_t = \frac{1}{2} \left(X_n^2 - X_0^2 - \sum_{t=1}^n \varepsilon_t^2 \right)$$

- Sous l'hypothèse que $X_0 = 0$ et en multipliant tout par n^{-1} on a

$$n^{-1} \sum_{t=1}^n X_{t-1}\varepsilon_t = \frac{1}{2} \left(n^{-1} X_n^2 - n^{-1} \sum_{t=1}^n \varepsilon_t^2 \right)$$

- Via le lemme (1), le premier terme nous donne

$$n^{-1} X_n^2 = (n^{-1/2} \sum_{t=1}^n \varepsilon_t)^2 \xrightarrow{d} \sigma_\varepsilon^2 W(1)^2$$

- Via le théorème de Kolmogorov (9), $n^{-1} \sum_{t=1}^n \varepsilon_t^2 \xrightarrow{a.s.} \sigma_\varepsilon^2$ et

$$n^{-1} \sum_{t=1}^n X_{t-1}\varepsilon_t \xrightarrow{d} \frac{1}{2} \sigma_\varepsilon^2 (W(1)^2 - 1) \quad (13)$$

Généralité sur les tests de racine unitaire

- Soit un AR(1) dont on souhaite tester si $\rho = 1$
- Les difficultés résident dans la formulation d'un test car
 - la distribution limite ne sera pas standard
- Dickey-Fuller ont proposé deux tests : *DF* et *DF* augmenté (*ADF*)
- Phillips-Perron ont proposé une alternative non-paramétrique d'ADF

Un premier test de racine unitaire

- Soit x_t le taux de change journalier EUR/USD sur 2000-2014
- On sait que si $x_t = \rho x_{t-1} + \varepsilon_t$ avec $\rho = 1$, d'après Th.6

$$z_{\hat{\rho}} = n(\hat{\rho} - 1) \xrightarrow{d} \frac{1/2(W(1))^2 - 1}{\int_0^1 W(r)^2 dr} = \mathcal{L}(\hat{\rho})$$

- $\mathcal{L}(\hat{\rho})$ est tabulée au S196 et peut servir pour tester $H_0 : \rho = 1$
- Calculons la statistique $z_{\hat{\rho}}$ pour $n = 3760$ sachant que

$$\hat{\rho} = 0.9999919311 \Rightarrow z_{\hat{\rho}} = n(\hat{\rho} - 1) = -0.03033893$$

- Dans la table, pour $n > 500 \equiv \infty$, 95% des fois lorsque le processus est vraiment une marche aléatoire, $s_{\hat{\rho}}$ est supérieur à -8.1

\Rightarrow puisque $-0.03 > -8.1$, l'hypothèse nulle $\rho = 1$ est acceptée au seuil de 5% et $x_t \sim I(1)$

Un premier test de racine unitaire : annexe

TABLE B.5
Critical Values for the Phillips-Perron Z_{ρ} Test and for the Dickey-Fuller Test
Based on Estimated OLS Autoregressive Coefficient

Sample size T	Probability that $T(\hat{\rho} - 1)$ is less than entry							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
<i>Case 1</i>								
25	-11.9	-9.3	-7.3	-5.3	1.01	1.40	1.79	2.28
50	-12.9	-9.9	-7.7	-5.5	0.97	1.35	1.70	2.16
100	-13.3	-10.2	-7.9	-5.6	0.95	1.31	1.65	2.09
250	-13.6	-10.3	-8.0	-5.7	0.93	1.28	1.62	2.04
500	-13.7	-10.4	-8.0	-5.7	0.93	1.28	1.61	2.04
∞	-13.8	-10.5	-8.1	-5.7	0.93	1.28	1.60	2.03
<i>Case 2</i>								
25	-17.2	-14.6	-12.5	-10.2	-0.76	0.01	0.65	1.40
50	-18.9	-15.7	-13.3	-10.7	-0.81	-0.07	0.53	1.22
100	-19.8	-16.3	-13.7	-11.0	-0.83	-0.10	0.47	1.14
250	-20.3	-16.6	-14.0	-11.2	-0.84	-0.12	0.43	1.09
500	-20.5	-16.8	-14.0	-11.2	-0.84	-0.13	0.42	1.06
∞	-20.7	-16.9	-14.1	-11.3	-0.85	-0.13	0.41	1.04
<i>Case 4</i>								
25	-22.5	-19.9	-17.9	-15.6	-3.66	-2.51	-1.53	-0.43
50	-25.7	-22.4	-19.8	-16.8	-3.71	-2.60	-1.66	-0.65
100	-27.4	-23.6	-20.7	-17.5	-3.74	-2.62	-1.73	-0.75
250	-28.4	-24.4	-21.3	-18.0	-3.75	-2.64	-1.78	-0.82
500	-28.9	-24.8	-21.5	-18.1	-3.76	-2.65	-1.78	-0.84
∞	-29.5	-25.1	-21.8	-18.3	-3.77	-2.66	-1.79	-0.87

The probability shown at the head of the column is the area in the left-hand tail.

Source: Wayne A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 1976, p. 371.

Le test de Dickey-Fuller : intuition

- Dans un premier temps, considérons un AR(1) simple

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

- Dickey-Fuller (DF) proposent la reformulation suivante

$$\Delta X_t = (\rho - 1)X_{t-1} + \varepsilon_t = \phi X_{t-1} + \varepsilon_t$$

- La formulation du test de racine unitaire devient alors

$$H_0 : \phi = 0 \text{ versus } H_1 : \phi < 0$$

- Grâce au Th.6, on connaît la distribution limite des OLS pour

$$\hat{\phi} = (\rho - 1)$$

⇒ la distribution de la statistique de “type” Student

$$\hat{\tau}_3 = \frac{\hat{\phi}}{\hat{\sigma}_\phi}$$

ne sera pas standard non plus (voir Eq. 11)

Le test de Dickey-Fuller : cas général

- Dans la pratique il convient d'envisager la présence de

$$\mu \neq 0 \text{ et/ou } \delta \neq 0$$

⇒ 3 modèles distinct sont alors envisageables

M1 Le modèle le plus général, avec $\mu \neq 0$ et/ou $\delta \neq 0$

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

M2 Le modèle sans tendance mais avec $\mu \neq 0$

$$\Delta X_t = \mu + \phi_2 X_{t-1} + \varepsilon_t$$

M3 Le modèle le plus spécifique décrivant une simple marche aléatoire

$$\Delta X_t = \phi_3 X_{t-1} + \varepsilon_t$$

- Comme sous H_0 on a une racine unitaire, les distributions de

$$\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$$

sont non-standards, tout comme les distributions de

$$\hat{\tau}_1 = \hat{\phi}_1 / \hat{\sigma}_\phi, \quad \hat{\tau}_2 = \hat{\phi}_2 / \hat{\sigma}_\phi, \quad \hat{\tau}_3 = \hat{\phi}_3 / \hat{\sigma}_\phi$$

Le test de Dickey-Fuller : les alternatives

- L'existence de plusieurs alternatives complique l'implémentation

e.g. Supposons que l'observation des données révèle une tendance

⇒ la question est alors de savoir si cette tendance vient de

$$\Delta X_t = \mu + \phi X_{t-1} + \varepsilon_t \Leftrightarrow X_t = \mu + X_{t-1} + \varepsilon_t \text{ sous } H_0$$

i.e. une marche aléatoire avec dérive, ou si elle vient de

$$\Delta X_t = \mu + \delta t + \phi X_{t-1} + \varepsilon_t \Leftrightarrow X_t = \mu + \delta t + \rho X_{t-1} + \varepsilon_t \text{ sous } H_1$$

i.e. un AR(1) stationnaire ($\phi < 0 \Leftrightarrow \rho < 1$) autour d'un trend

- Dans cette situation, il est préférable de tester $\phi = 0$ via

$$\hat{\tau}_1 = \hat{\phi}_1 / \hat{\sigma}_\phi$$

et de compléter cela par un test de nullité jointe (type Fisher)

$$F_1 := \hat{\phi}_1 = \hat{\delta} = 0$$

ou

$$F_2 := \hat{\phi}_1 = \hat{\delta} = \hat{\mu} = 0$$

Le test de Dickey-Fuller : méthodologie

- L'étape 1 de la procédure à suivre en absence d'a priori sera
 - 1.a Estimer M1 et tester $H_0 : \phi_1 = 0$ avec $\hat{\tau}_1$
 - 1.b Si rejet de H_0 , $(X_t - \delta t) \sim I(0)$
 - ⇒ on peut tester $\delta = 0$ et $\mu = 0$ avec des valeurs critiques usuelles
 - 1.c Si non-rejet de H_0 , X_t possède une racine unitaire
 - ⇒ il faut tester $\hat{\phi}_1 = \hat{\delta} = 0$ avec F_1 (valeurs critiques tabulées)
 - 1.d Si rejet, on garde M1 et on recommence la procédure sur ΔX_t
 - 1.e Si non-rejet, on peut soit considérer F_2 , soit passer à l'étape 2
 - ⇒ Si non-rejet avec F_2 , on retient M3

Le test de Dickey-Fuller : méthodologie

- L'étape 2 de la procédure à suivre en absence d'a priori sera

2.a Estimer M2 et tester $H_0 : \phi_2 = 0$ avec $\hat{\tau}_2$

2.b Si rejet de $H_0, X_t \sim I(0)$

⇒ on peut tester $\mu = 0$ avec des valeurs critiques usuelles

2.c Si non-rejet de H_0, X_t possède une racine unitaire

⇒ il faut tester $\hat{\phi}_2 = \hat{\mu} = 0$ avec F_3 (valeurs critiques tabulées)

2.d Si rejet, on garde M2 et on recommence la procédure sur ΔX_t

2.e Si non-rejet, on peut passer à l'étape 3

3.a Estimer M3 et tester $H_0 : \phi_3 = 0$ avec $\hat{\tau}_3$

3.b Si rejet de $H_0, X_t \sim I(0)$

3.c Si non-rejet, on garde M3 on recommence la procédure sur ΔX_t

Les valeurs critiques de Dickey-Fuller

- La forme non-standard des distributions limites de

$$\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{F}_1, \hat{F}_2, \hat{F}_3$$

empêchent le calcul analytique de valeurs critiques

- Dickey-Fuller proposent des valeurs critiques tabulées pour

$$F_i = \frac{(SSR_c - SSR_{nc})/r}{SSR_{nc}/(n - k)}$$

avec r et k les # de restrictions et paramètres (modèle non-contraint)

	Étape (1)						Étape (2)		
	$\hat{F}_2 := \hat{\phi}_1 = \hat{\mu} = \hat{\delta} = 0$			$\hat{F}_1 := \hat{\phi}_1 = \hat{\mu} = 0$			$\hat{F}_3 := \hat{\phi}_2 = \hat{\mu} = 0$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
T									
50	9.31	6.73	5.61	7.02	5.13	4.31	7.06	4.86	3.94
100	8.73	6.49	5.47	6.50	4.88	4.16	6.70	4.71	3.86
250	8.43	6.34	5.39	6.22	4.75	4.07	6.52	4.63	3.81
500	8.34	6.30	5.36	6.15	4.71	4.05	6.47	4.61	3.79
∞	8.27	6.25	5.34	6.09	4.68	4.03	6.43	4.59	3.78

Les valeurs critiques de Dickey-Fuller

- Dickey-Fuller proposent des valeurs critiques tabulées pour

$$\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3$$

Note 1 Les distributions limites ressemblent toutes à l'Eq. 11

Note 2 Elles diffèrent légèrement selon la présence de $\mu \neq 0$ et/ou $\delta \neq 0$,

T	M1 : $\hat{\tau}_1$			M2 : $\hat{\tau}_2$			M3 : $\hat{\tau}_3$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	-4.15	-3.50	-3.18	-3.58	-2.93	-2.60	-2.62	-1.95	-1.61
100	-4.04	-3.45	-3.15	-3.51	-2.89	-2.58	-2.6	-1.95	-1.61
250	-3.99	-3.43	-3.13	-3.46	-2.88	-2.57	-2.58	-1.95	-1.62
500	-3.98	-3.42	-3.13	-3.44	-2.87	-2.57	-2.58	-1.95	-1.62
∞	-3.96	-3.41	-3.12	-3.43	-2.86	-2.57	-2.58	-1.95	-1.62

Le test de Dickey-Fuller Augmenté : le modèle AR(p)

- Le test précédent se limite au modèle AR(1) mais qu'en est-il si

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

car déterminer et tester les p racines est un tâche complexe ?

- Pour simplifier le problème introduisons $\rho \equiv \phi_1 + \dots + \phi_p$ et

$$\xi_j \equiv -(\phi_{j+1} + \phi_{j+2} + \dots + \phi_p)$$

pour $j = 1, 2, \dots, p-1$ de sorte qu'on puisse écrire

$$\begin{aligned} \varepsilon_t &= \left((1 - \rho L) - (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1}) (1 - L) \right) X_t \\ &= X_t - \rho X_{t-1} - \xi_1 \Delta X_{t-1} - \xi_2 \Delta X_{t-2} - \dots - \xi_{p-1} \Delta X_{t-p+1} \end{aligned}$$

et $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$ de manière équivalente car

$$(1 - \rho L) - (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1}) (1 - L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

Le test de Dickey-Fuller Augmenté : intuition

- La reformulation nous ramène vers un modèle proche du test DF

$$X_t = \rho X_{t-1} + \sum_{j=1}^p \xi_j \Delta X_{t-j} + \varepsilon_t$$

- Si une des racine de $1 - \phi_1 z - \dots - \phi_p z^p$ est $z = 1$, $\rho = 1$ ou $\phi = 0$

$$\Delta X_t = (\rho - 1)X_{t-1} + \sum_{j=1}^p \xi_j \Delta X_{t-j} + \varepsilon_t = \phi X_{t-1} + \sum_{j=1}^p \xi_j \Delta X_{t-j} + \varepsilon_t$$

- La formulation du test de racine unitaire reste alors

$$H_0 : \phi = 0 \text{ versus } H_1 : \phi < 0$$

- La distribution limite des OLS pour ϕ reste celle du Th.6

⇒ la distribution de la statistique de "type" Student

$$\hat{\tau}_3 = \frac{\hat{\phi}_3}{\hat{\sigma}_\phi}$$

reste celle de l'Eq. 11 mais les $\hat{\tau}_{\xi_j}$ ont une distribution standard

La procédure de Dickey-Fuller Augmenté

- La procédure du test ADF est identique à celle du test DF
- Il est néanmoins crucial ici de déterminer p auparavant

0.a. Estimer le modèle

$$\Delta X_t = \mu + \delta t + \phi X_{t-1} + \sum_{j=1}^{pmax} \xi_j \Delta X_{t-j} + \varepsilon_t$$

0.b. Contrôler la présence d'auto-corrélation sérielle des résidus (test de Ljung-Box)

0.c. Si les résidus sont auto-corrélés on augmente $pmax$

0.d. Si les résidus sont non auto-corrélés on teste avec $p < pmax$

0.e. On garde le plus petit p qui satisfait 0.d. et on passe à l'étape 1.a.

Note L'utilisation des critères de sélection est aussi possible

La procédure ADF en présence d'ARMA(p, q)

- La procédure du test ADF reste identique
- ⇒ Elle repose la représentation AR(∞) des ARMA(p, q)
- Il faut chercher une **l'approximation** par un AR($p < \infty$)
 - Dans ce context les critères AIC ou BIC sont à privilégier

Note La procédure de sélection de p est importante pour que

$$\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3$$

possèdent les distributions (non-standard) attendues

Le test de Phillips-Perron (PP) : intuition

- L'éradication de l'autocorrélation sérielle de ε_t peut être délicate
- ⇒ PP proposent une correction non-paramétrique basée sur

$$\lambda_n^2,$$

la variance de long terme de ε_t

- L'estimateur proposé repose sur les autocovariances empiriques

$$\hat{\gamma}_n(j) = n^{-1} \sum_{t=j+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}$$

et s'exprime comme (l'estimateur de Newey-West)

$$\hat{\lambda}_n^2 = \hat{\gamma}_n(0) + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1}\right) \hat{\gamma}_n(j)$$

- Une statistique corrigée de type Student en découle

$$\hat{\tau}_j^* = \sqrt{\frac{\hat{\gamma}_n(0)}{\hat{\lambda}_n^2} \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}}} - \frac{1}{2} (\hat{\lambda}_n^2 - \hat{\gamma}_n(0)) \frac{1}{\hat{\lambda}_n} \frac{n \hat{\sigma}_{\hat{\rho}}}{\hat{\sigma}_{\hat{\varepsilon}}}, \quad j = 1, 2, 3$$

Le test de Phillips-Perron (PP) : implémentation

- Le test de PP s'applique de manière identique au test de DF
- La correction dépend en revanche d'un paramètre de tuning q

⇒ ce paramètre de troncature ne doit pas croître plus vite que

$$n^{1/4}$$

sous peine de non-convergence de $\widehat{\lambda}_n^2$ vers λ_n^2

⇒ Sous cette condition, la correction est valide, et

$$\widehat{\tau}_j^* \stackrel{d}{=} \widehat{\tau}_j$$

Note Les distributions des statistiques de type Student sont identiques à DF et ADF pour chacun des 3 modèles

Le concept de régression factice

- La non-stationnarité n'a pas pour seule conséquence l'émergence de distributions limites non-standards
- Supposons que l'on cherche à estimer le modèle

$$Y_t = \beta X_t + \varepsilon_t$$

avec $\beta = 0$

- Si Y_t et X_t sont des marches aléatoires, l'estimateur **OLS n'est pas consistant**
 - ⇒ $\hat{\beta}$ converge vers une **variable aléatoire non-dégénérée**
- $\hat{\beta}$ ne pouvant révéler l'absence de relation entre Y_t et X_t on parle de **régression factice**

Théorie limite et régression factice

Theorem 7 (7)

Soit Y_t et X_t des marches aléatoires indépendantes, $Y_t = Y_{t-1} + \eta_t$ et $X_t = X_{t-1} + \nu_t$ avec $\eta_t \perp \nu_t$. On considère la régression

$$Y_t = \beta X_t + \varepsilon_t$$

avec $\beta = 0$. Alors la théorie limite de l'estimateur OLS de β

$$\hat{\beta} = 0 + \left(n^{-1} \sum_t X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t X_{t-1} Y_{t-1} \right)$$

nous donne après normalisation par n

$$\hat{\beta} - 0 \xrightarrow{d} \left(\sigma_\nu \int_0^1 W_X(r)^2 dr \right)^{-1} \left(\sigma_\eta \int_0^1 W_X(r) W_Y(r) dr \right)$$

et révèle donc l'inconsistance de $\hat{\beta}$ si $\beta = 0$

Théorie limite et régression factice : normalisation

- En multipliant $\hat{\beta}$ par n de chaque côté on obtient

$$n(\hat{\beta} - \beta) = \frac{1}{n^{-1}} \left(n^{-1} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t^n X_{t-1} Y_{t-1} \right)$$

- Ce qui nous donne

$$n(\hat{\beta} - 0) = \left(n^{-2} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-1} \sum_t^n X_{t-1} Y_{t-1} \right)$$

⇒ En faisant passer le facteur n du terme de gauche, à droite

$$\hat{\beta} = \left(n^{-2} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-2} \sum_t^n X_{t-1} Y_{t-1} \right)$$

- La normalisation par n avait fonctionné pour la marche aléatoire
- Ici, n va disparaître et annihiler la vitesse de convergence

Théorie limite et régression factice : démonstration

- En posant $r = r_{t-1} = (t - 1)/n$, on sait que

$$n^{-1/2}X_{t-1}\sigma_\nu^{-1} = W_{X_n}(r) \text{ et } n^{-1/2}Y_{t-1}\sigma_\eta^{-1} = W_{Y_n}(r)$$

- Par le Lemme (2) on obtient

$$n^{-2} \sum_t^n X_{t-1}^2 \xrightarrow{d} \sigma_\nu^2 \int_0^1 W(r)^2 dr$$

- Puisque $\eta_t \perp \nu_t$, une application multivariée du TCLF donne

$$\begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\nu^2 \end{pmatrix}^{-1/2} n^{-1/2} \sum_{t=1}^{[nr]} \begin{pmatrix} \eta_t \\ \nu_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_Y(r) \\ W_X(r) \end{pmatrix} \quad (14)$$

- Nous pouvons alors à analyser le dénominateur $n^{-2} \sum_t^n X_{t-1}Y_{t-1}$

Théorie limite et régression factice : démonstration

- Sachant

- le Lemme (1) qui nécessite un $n^{-1} = n^{-1/2}n^{-1/2}$ car X_{t-1} ET Y_{t-1}
- que $\eta_t \perp \nu_t$ et l'équation (14)
- que $W_n(r)$ est constant si $(t-1)/n \leq r_{t-1} < t/n$
- le théorème de Mann-Wald fonctionnel
- que l'astuce du Lemme (3) n'est pas possible ici (pas les bons chocs)

$$\begin{aligned}n^{-2} \sum_t^n X_{t-1} Y_{t-1} &= n^{-1} \sum_{t=1}^n \sigma_\eta W_{Y_n}(r) \sigma_\nu W_{X_n}(r) \\ &= \sigma_\eta \sigma_\nu \sum_{t=1}^n \int_{(t-1)/n}^{t/n} W_{Y_n}(r) W_{X_n}(r) dr \\ &= \sigma_\eta \sigma_\nu \int_0^1 W_{Y_n}(r) W_{X_n}(r) dr \\ &\xrightarrow{d} \sigma_\eta \sigma_\nu \int_0^1 W_Y(r) W_X(r) dr\end{aligned}$$

- n^{-1} disparaît pour faire apparaître l'intégrale

Théorie limite et régression factice : démonstration

- On constate alors que

$$\begin{aligned}\hat{\beta} &= \left(n^{-2} \sum_t^n X_{t-1}^2 \right)^{-1} \left(n^{-2} \sum_t^n X_{t-1} Y_{t-1} \right) \\ &= \left(\sigma_\nu^2 \int_0^1 W_X(r)^2 dr \right)^{-1} \left(\sigma_\eta \sigma_\nu \int_0^1 W_Y(r) W_X(r) dr \right) \\ &\xrightarrow{d} \left(\sigma_\nu \int_0^1 W_X(r)^2 dr \right)^{-1} \left(\sigma_\eta \int_0^1 W_X(r) W_Y(r) dr \right)\end{aligned}$$

- Phillips (1986) démontre un résultat similaire pour $n^{-1/2} \hat{t}_\beta$
 \Rightarrow Si $n \rightarrow \infty$ la probabilité de trouver $\hat{\beta}$ significatif approche 1 car la distribution de \hat{t}_β diverge à une vitesse $n^{1/2}$
- Phillips (1986) montre également qu'en présence d'une constante α dans la régression, la distribution de $\hat{\alpha}$ diverge

Rappels : test de Durbin-Watson et R^2

- Le test DW est un test d'autocorrélation à l'ordre 1 pour

$$y_t = \beta x_t + u_t, \quad u_t = \varphi u_{t-1} + \varepsilon_t$$

avec $|\varphi| < 1$ et $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

- Le test requiert également que $x_t \neq y_{t-1}$ et sous

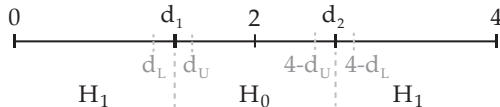
$$H_0 : \varphi = 0 \text{ versus } H_1 : \varphi \neq 0$$

voit sa statistique de test

$$DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^n \hat{u}_t^2} \approx 2(1 - \hat{\varphi}) \in [0, 4]$$

suivre une distribution non standard dont les valeurs critiques d_U et d_L sont tabulées et encadrent des seuils inconnus d_1 et d_2

⇒ Une lecture rapide pointerait vers un non rejet de H_0 si $DW \approx 2$



Rappels : R^2

- Le coefficient de détermination est donné par

$$\widehat{R}^2 = 1 - \left(\sum_{t=1}^n y_t^2 \right)^{-1} \left(\sum_{t=1}^n \widehat{\varepsilon}_t^2 \right)$$

- La statistique \widehat{R}^2 un estimateur de la quantité théorique

$$R^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}$$

- Dans le cadre des ARMA(p, q), on ne connaît sa distribution limite

$$\sqrt{n}(\widehat{R}^2 - R^2) \sim \mathcal{N}\left(0, 4\sigma_\varepsilon^2(\sigma_y^2)^{-1} \sum_{i=1}^{\infty} \rho_i^2\right)$$

que pour $q = 0$ et dans un cadre non-stationnaire, elle devient non-standard

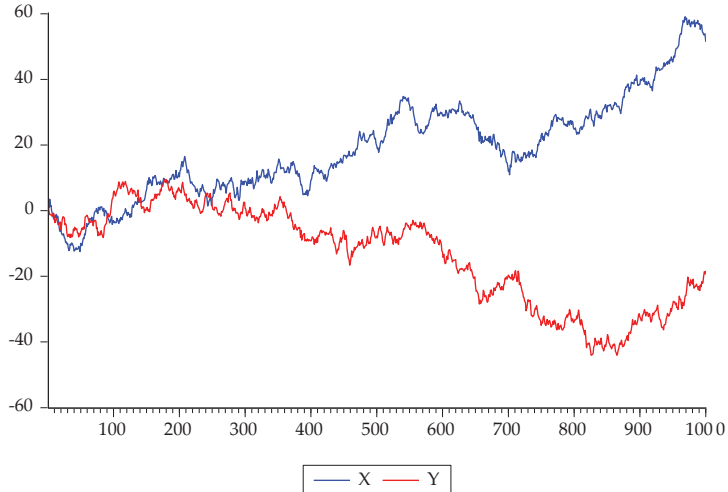
- En présence de régression fallacieuse, elle est non-dégénérée

⇒ \widehat{R}^2 pointe vers des valeurs > 0 et donc erronées

⇒ DW pointe vers des valeurs anormalement basses

Exemple : régression factice

- Soit X_t et Y_t deux marches aléatoires indépendantes



Exemple : régression factice

- Estimons le modèle

$$X_t = \beta_0 + \beta_1 Y_t + \varepsilon_t$$

pour lequel on s'attend à trouver $\hat{\beta}_1 = 0$

⇒ pourtant les résultats indiquent un relation significative

Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.557269	0.482904	19.79125	0.0000
Y	-0.700810	0.024364	-28.76363	0.0000
R-squared	0.453254	Mean dependent var		18.87462
Adjusted R-squared	0.452707	S.D. dependent var		15.30898
S.E. of regression	11.32547	Akaike info criterion		7.693984
Sum squared resid	128009.8	Schwarz criterion		7.703799
Log likelihood	-3844.992	Hannan-Quinn criter.		7.697714
F-statistic	827.3464	Du rbin-Watson stat		0.011903
Prob(F-statistic)	0.000000			

Exemple : régression factice

- Estimons à présent le modèle

$$\Delta X_t = \beta_1 \Delta Y_t + u_t$$

pour lequel on s'attend à trouver $\hat{\beta}_1 = 0$

⇒ à présent on constate qu'il n'y a aucune relation entre X_t et Y_t

Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.051278	0.031154	1.645935	0.1001
DY	0.035033	0.030722	1.140296	0.2544
R-squared	0.001302	Mean dependent var		0.050595
Adjusted R-squared	0.000301	S.D. dependent var		0.984653
S.E. of regression	0.984505	Akaike info criterion		2.808644
Sum squared resid	966.3420	Schwarz criterion		2.818467
Log likelihood	-1400.918	Hannan-Quinn criter.		2.812378
F-statistic	1.300276	Durbin-Watson stat		2.072721
Prob(F-statistic)	0.254437			

Régression factice et intégration fractionnaire

- Tsay et Chung (2000) étendent les résultats de Phillips (1986) au cas fractionnaire
- Soit deux processus à mémoire longue

$$y_t \sim I(\delta_y) \text{ et } x_t \sim I(\delta_x) \text{ avec } \delta_y, \delta_x \in (0, 1/2)$$

- Tsay et Chung (2000) montrent que
 - le risque de régression factice existe dès lors que $\delta_y + \delta_x > 1/2$
 - ⇒ même si les deux processus sont stationnaires !
- La démonstration de ce résultat est bien plus complexe et dépasse de loin le niveau M2

Ce qu'il faut retenir

- La notion de non-stationnarité est protéiforme
- Dans un cadre linéaire, la stationnarité disparaît si $\gamma(\mathbf{h})^2 = \infty$
- Il existe des processus stationnaires dont $\gamma(\mathbf{h})^2 < \infty$ mais $|\gamma(\mathbf{h})| = \infty$, on parle de processus mémoire longue
- Plus généralement les processus sont $I(\delta \in \mathbb{R})$ et $\delta = 0$ ou $\delta = 1$ sont des cas particuliers
- Si $\delta \geq 1/2$, la non-stationnarité survient car $\gamma(\mathbf{h})^2 = \infty$
- Si $\delta = 0$ on connaît la théorie limite des OLS
- Si $\delta = 1$ la théorie limite des OLS devient non standard et le risque de régression factice émerge
- Ce risque émerge en réalité pour $\delta_x + \delta_y > 1/2$

Les concepts de convergences

- Soit X_i une fonction de n variables aléatoires $X_n = f(Y_1, \dots, Y_n)$
 - L'étude du comportement de X_n quand $n \rightarrow \infty$ est cruciale
 - $f(\cdot)$ sera souvent un **estimateur**
 - L'étude de ce comportement **limite** repose sur différentes notions de convergence
 - convergence **presque sûre**
 - convergence en **probabilité**
 - convergence en **moyenne quadratique**
 - convergence en **loi**

Convergence presque sûre

- **Implications:** quand $n \rightarrow \infty$, X_n tend de façon certaine vers une constante (i.e. une variable aléatoire dégénérée)

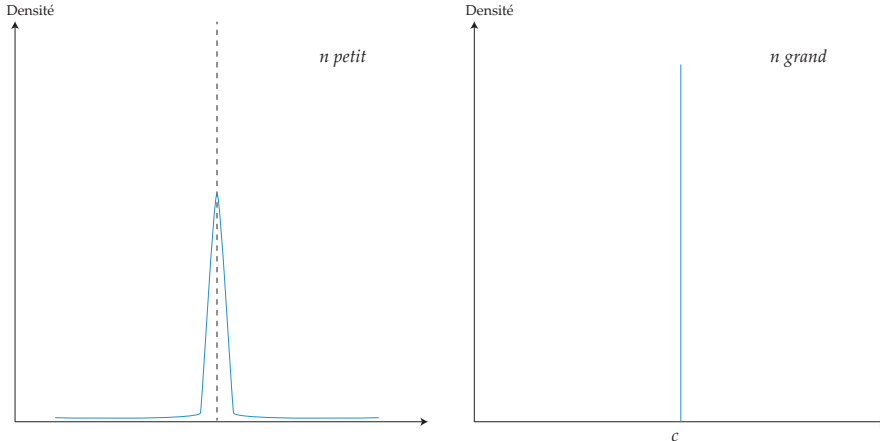
Definition 24 (24)

x_n converge presque sûrement vers une constante c si,

$$\Pr \left(\lim_{n \rightarrow \infty} X_n = c \right) = 1$$

- **Notation mathématique:** $X_n \xrightarrow{a.s.} c$
- **Explications:** Comme X_n tend vers une valeur **constante** de manière **certaine**, sa **distribution asymptotique** est une **masse ponctuelle**

Convergence presque sûre: Illustration



Convergence en probabilité

- **Implications:** quand $n \rightarrow \infty$, X_n tend vers une constante

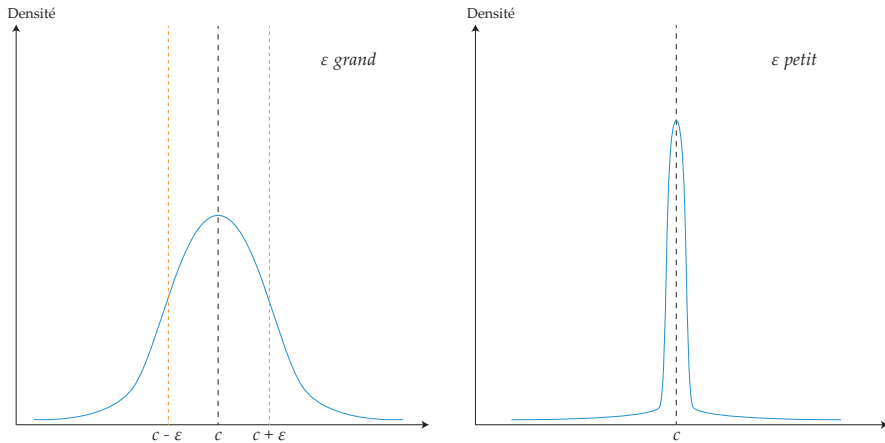
Definition 25 (25)

X_n converge en probabilité vers une constante c , si pour toute valeur de $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - c| > \epsilon) = 0$$

- **Notation mathématique:** $X_n \xrightarrow{p} c$ ou $\text{plim } X_n = c$
- **Explications:** La convergence en probabilité n'est pas stricte comme la convergence presque sûre et on parle également de convergence au sens faible. Par conséquent, X_n converge asymptotiquement vers une quantité aléatoire dont la densité est très concentrée autour de c

Convergence en probabilité: Illustration



Convergence en moyenne quadratique

- **Implications:** quand $n \rightarrow \infty$, X_n tend vers une constante

Definition 26 (26)

X_n converge en moyenne quadratique vers une constante c , si $\mathbb{E}(|X_n|^2) < \infty$ et si pour toute valeur de $\gamma > 0$,

$$\mathbb{E}(|X_n - c|^2) < \gamma$$

- **Notation mathématique:** $X_n \xrightarrow{m.s.} c$
- **Explications:** X_n converge en moyenne quadratique si sa distribution est centrée sur c , i.e. $\mathbb{E}(X_n) = c$ et si sa variance tend vers 0 asymptotiquement, impliquant une densité très concentrée autour de c

Convergence en loi

- **Implications:** quand $n \rightarrow \infty$, X_n tend vers une autre variable aléatoire et dont la distribution est asymptotiquement équivalente

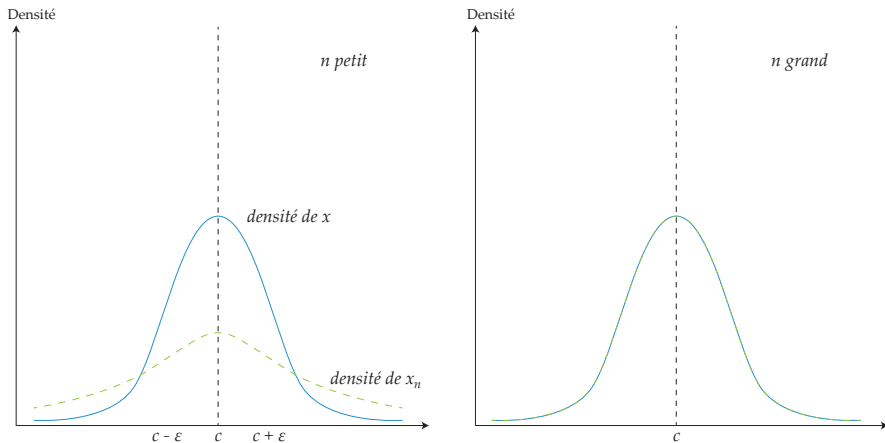
Definition 27 (27)

Soit $F_n(\cdot)$ la fonction de répartition de X_n . X_n converge en loi vers une variable aléatoire X définie sur un support $X(\Omega)$ et ayant pour fonction de répartition $F(\cdot)$ si,

$$\lim_{n \rightarrow \infty} F_n(z) = F(z), \quad \forall z \in X(\Omega)$$

- **Notation mathématique:** $X_n \xrightarrow{d} X$ ou $X_n \xrightarrow{d} \mathcal{L}$
- **Explications:** Asymptotiquement la distribution de X_n est donc identique à celle de X , ce qui implique des fonctions de densité et de répartition identiques: X_n et X sont identiquement distribuées

Convergence en loi: Illustration



Loi faible des grands nombres

- **Implications:** il s'agit d'un théorème portant sur une séquence de variables aléatoires i. i. d.

Theorem 8 (8: Weak Law of Large Numbers)

Pour une séquence de variables aléatoires i. i. d. , $X_t = X_1, \dots, X_n$, la moyenne empirique de ces variables converge en probabilité vers l'espérance de X_t

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \xrightarrow{p} \mathbb{E}(X_t)$$

- Par la suite on notera **LLN** la loi des grands nombres

Loi forte des grands nombres

- **Notes:** Il existe des versions fortes de cette loi

$$\text{plim } \bar{X} = \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \mathbb{E}(X_t) \text{ si } X_t \text{ indépendant } \forall t \quad (15)$$

$$= \mu \text{ si } X_t \sim \text{i. i. d.} \quad (16)$$

Theorem 9 (9: LLN Kolmogorov)

Pour une séquence de variables aléatoires i. i. d. , $X_t = X_1, \dots, X_n$, si $\mathbb{E}(|X_t|) < \infty$, la moyenne empirique de ces variables converge presque sûrement vers l'espérance de X_t

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \xrightarrow{\text{a.s.}} \mathbb{E}(X_t)$$

Loi forte des grands nombres

- Le théorème LLN de Markov relâche l'hypothèse de distribution identique au coût de conditions sur les moments plus élevés

Theorem 10 (10: LLN Markov)

Pour une séquence de variables aléatoires indépendamment mais non-identiquement distribuées, $X_t = X_1, \dots, X_n$, avec $\mathbb{E}(X_t) = \mu_t$ et $\mathbb{V}(X_t) = \sigma_t^2$, si

$$\sum_{t=1}^{\infty} \left(\mathbb{E}(|X_t - \mu_t|^{1+m}) / t^{1+m} \right) < \infty$$

pour $m > 0$, alors

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \xrightarrow{a.s.} \frac{1}{n} \sum_{t=1}^n \mathbb{E}(X_t)$$

Vers un théorème de la limite centrée

- Soit une séquence i. i. d. , $X_t = X_1, \dots, X_n$ de moyenne

$$\bar{X} = n^{-1} \sum_{t=1}^n X_t$$

- D'après la LLN de Kolmogorov, $\bar{X} \xrightarrow{a.s.} \mathbb{E}(X_t)$ si $n \rightarrow \infty$

— la distribution asymptotique de \bar{X} est dégénérée

⇒ Comment construire une statistique inférentielle?

- La logique consiste à opérer une transformation de \bar{X} tel que

$$\mathbb{T}(\bar{X}) \xrightarrow{d} \mathcal{L}$$

avec \mathcal{L} une distribution non dégénérée

- Cette transformation sera généralement de la forme

$$\mathbb{T}(\bar{X}) = \sqrt{n} (\bar{X} - \mathbb{E}(X_t))$$

Pourquoi une telle transformation?

- Pour simplifier supposons $\mathbb{E}(X_t) = 0$ et $\mathbb{T}(\bar{X}) = n^\alpha \bar{X}$
 - Comme $X_t \sim$ i. i. d. , $\mathbb{E}(X_t) = 0$, $\mathbb{V}(X_t) = \sigma_X^2$ et $\mathbb{Cov}(X_i, X_j) = 0$, $i \neq j$

- On en déduit alors

$$\mathbb{E}(n^\alpha \bar{X}) = n^\alpha \mathbb{E}(\bar{X}) = 0 \quad , \quad \mathbb{V}(n^\alpha \bar{X}) = n^{2\alpha-1} \sigma_X^2$$

- Car:

$$\begin{aligned} \mathbb{V}(\bar{X}) &= \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n \mathbb{V}(X_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^{n-1} \mathbb{Cov}(X_i, X_j) \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma_X^2 = \frac{n\sigma_X^2}{n^2} = \frac{\sigma_X^2}{n} \end{aligned}$$

La vitesse de convergence

- Pour $\alpha \geq 0$ considérons trois cas

- $\alpha > 1/2$ et donc $2\alpha - 1 > 0$

$$\lim_{n \rightarrow \infty} \mathbb{V}(n^\alpha \bar{X}) = \sigma^2 \lim_{n \rightarrow \infty} n^{2\alpha-1} = \infty$$

- $\alpha < 1/2$ et donc $2\alpha - 1 < 0$

$$\lim_{n \rightarrow \infty} \mathbb{V}(n^\alpha \bar{X}) = \sigma^2 \lim_{n \rightarrow \infty} n^{2\alpha-1} = 0$$

- $\alpha = 1/2$ et donc $2\alpha - 1 = 0$

$$\lim_{n \rightarrow \infty} \mathbb{V}(n^\alpha \bar{X}) = \sigma^2 \lim_{n \rightarrow \infty} n^{2\alpha-1} = \sigma^2$$

⇒ La **normalisation** \sqrt{n} préserve la variance de X_t

⇒ On dit que $\bar{X} - \mathbb{E}(X_t)$ converge à la vitesse \sqrt{n}

Le théorème central limite

Theorem 11 (11: TCL de Lindeberg-Levy)

Soit une séquence i. i. d. , $X_t = X_1, \dots, X_n$ d'espérance $\mathbb{E}(X_t) = m$ et de variance finies $\mathbb{V}(X_t) = \sigma^2$. D'après le théorème central limite de Lindeberg-Levy,

$$\tilde{Z}_n = \sqrt{n}(\bar{X} - m) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \equiv Z_n = \frac{\tilde{Z}_n}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

avec $n \rightarrow \infty$ et \tilde{Z}_n s'exprimant également comme

$$\sqrt{n} \frac{1}{n} \sum_{t=1}^n (X_t - m) = n^{-1/2} \sum_{t=1}^n (X_t - m) = \sqrt{n} \left(\underbrace{\left(\frac{1}{n} \sum_{t=1}^n X_t \right)}_{\bar{X}} - \underbrace{\frac{1}{n} nm}_m \right)$$

- D'autres théorèmes centraux limites existent
 - TCL de Lyapunov (développé par la suite)
 - TCL pour martingales et processus mélangeants (non présentés)

Le théorème central limite

Theorem 12 (12: TCL de Lyapunov)

Soit une séquence i. ni. d. , $X_t = X_1, \dots, X_n$ d'espérance $\mathbb{E}(X_t) = \mu_t$ et de variance finies $\mathbb{V}(X_t) = \sigma_t^2$. D'après le théorème central limite de Lyapunov, si

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+m}} \left(\sum_{t=1}^n \mathbb{E}(|X_t - \mu_t|^{2+m}) \right) < \infty, \quad s_n^2 = \sum_{t=1}^n \sigma_t^2$$

pour $m > 0$,

$$Z_n = \frac{\sum_{t=1}^n (X_t - \mu_t)}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

La distribution asymptotique

- Soit une séquence i. i. d. , $X_t = X_1, \dots, X_n$ convergeant en loi vers X , une variable ayant pour fonction de répartition $F(\cdot)$

$$X_n \xrightarrow{d} \mathcal{L}_X$$

— $F(\cdot)$ est donc la fonction de répartition de la **distribution asymptotique** de X_n

- Dans le cadre du TCL, supposons

$$\sqrt{n}(X_n - m) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

- Peut-on en conclure que $X_n \xrightarrow{d} \mathcal{N}(m, \sigma^2/n)$?

⇒ En fait, on peut uniquement dire que

$$X_n \xrightarrow{a.a.d} \mathcal{N}(m, \sigma^2/n)$$

et que $\mathbb{V}_{asy}(X_n) = \sigma^2/n$ et $\mathbb{E}_{asy}(X_n) = m$

Théorème de Mann–Wald

Theorem 13 (13: Continuous Mapping Theorem)

Soit $X_t = X_1, \dots, X_n$ une séquence de variables aléatoires et $g(\cdot)$ une fonction continue à valeur dans \mathbb{R} en X . Alors,

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$

$$X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

- Exemples

- $(X_n, Y_n) \xrightarrow{p} (X, Y)$ implique $(X_n Y_n) \xrightarrow{p} XY$

- $(X_n, Y_n) \xrightarrow{p} (X, Y)$ implique $(X_n / Y_n) \xrightarrow{p} X/Y$ si $Y \neq 0$

Théorème de Slutsky

Theorem 14 (14: Théorème de Slutsky)









Soit $X_t = X_1, \dots, X_n$ et $Y_t = Y_1, \dots, Y_n$ deux séquences de variables aléatoires telles que $X_t \xrightarrow{d} X$ et $Y_t \xrightarrow{p} c \neq 0$. Alors,

$$X_t + Y_t \xrightarrow{d} X + c, \quad X_t Y_t \xrightarrow{d} Xc, \quad \frac{X_t}{Y_t} \xrightarrow{d} \frac{X}{c}$$

- En supposant que $X_t \xrightarrow{d} \mathcal{N}(m, \sigma^2)$ et que $Y_t \xrightarrow{p} 2$, le théorème nous apprend que

$$\frac{X_t}{Y_t} \xrightarrow{d} \mathcal{N}(m/2, \sigma^2/4)$$

References

-  Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, 66(1), 59-65.
-  Dahlhaus, R., Rao, S. S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics*, 34(3), 1075-1114.
-  Granger, C. W., Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1), 15-29.
-  Hosking, J. R. (1981). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics*, 73(1), 261-284.
-  Lieberman, O., Phillips, P.C.B. (2008). A complete asymptotic series for the autocovariance function of a long memory process. *Journal of Econometrics* 147, 99-103.
-  Phillips, P.C.B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311-340.
-  Rao, S. S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Advances in Applied Probability*, 38(4), 1155-1172.
-  Tsay, W.-J., Chung, C.-F. (2000). The spurious regression of fractionally integrated processes. *Journal of Econometrics*, 96(1), 155-182.

Thank you!