Partie I. Introduction aux Processus stochastiques stationnaires

N. Laïb

EISTI

10 décembre 2018

Plan

Introdcution

- Dans le cours de séries temporelles, nous avons traité les données et fait des prévisions sans supposé aucun modèle au préalable.
- Cette façon se révèle insuffisante pour prévoir des séries financières et économiques.
- Le mieux est de proposer un modèle capable de reproduire le "comportement" des données.
- Dans les travaux de Box et Jenkins dans les années 1970, on considère qu'à un instant t, la valeur x_t est une réalisation d'une v.a.
- Cela conduit à définir une famille de variables aléatoires indexées par le temps,
  - porte le nom de processus aléatoire
  - la chronique x_t porte le non d'échantillon ou trajectoire du processus.

▷ Formellement, cela suppose que les :

- données observées $x_1, x_2, \ldots, x_n$ forment un extrait d'une trajectoire d'un processus stochastique $(X_t(\omega))_{t \in \mathbb{Z}}$,
- C'est-à-dire qu'il existe ω tel que

$$(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega)).$$

chaque donnée de la série est la réalisation de l'une des variables aléatoires qui composent le processus stochastique.

▷ **Objectif** : proposer un modèle pour le processus $\{X_t\}$. Ce modèle sera utile seulement si :

- on peut vérifier sa stationnarité, i.e une certaine uniformité de structure par rapport au temps.
- Dans ce cas on peut associer deux caractéristiques essentielles au processus :

  - la fonction d'autocorrélation ,
  - et la fonction d'autocorrélation partielle.

Processus strictement (fortement) stationnaire

Définition 1

Un processus $X_t, t \in T$ est strictement (fortement) stationnaire si $\forall$

$$t_1 < t_2 < \ldots < t_n, \quad t_i \in T \quad \text{et } \forall h \in T \quad \text{avec} \quad t_i + h \in T,$$

$$\forall \ i = 1, \ldots, n \quad \mathscr{L}(X_{t_1+h}, \ldots, x_{t_n+h}) = \mathscr{L}(X_{t_1}, \ldots, X_{t_n})$$

Remarque.

▷ La stationnarité stricte est très restrictive, puisqu'elle exige que la série est totalement invariante dans le temps. On préfère alors la stationnarité au second ordre (ou faiblement stationnaire) :

Processus faiblement stationnaire ou Stationnarité au second ordre

Définition 2

▷ Un processus $X_t, t \in T$ est dit faiblement stationnaire (ou du second ordre) si :

- $\mathbb{E}(X_t) = \mu < \infty$, constante indépendante de t, $\forall t \in T$
- $V(X_t) < \infty$, indépendante du temps,
- $\text{Cov}(X_t, X_{t+h}) = \gamma(h) \quad \forall t \in T, \forall h \in T$ (indépendant du temps et dépend que de h)
- ou encore

$$\text{Cov}(X_1, X_{1+h}) = \text{Cov}(X_2, X_{2+h}) = \ldots = \text{Cov}(X_t, X_{t-h})$$

▷ La covariance dépend de la différence de temps seule et non du temps.

▷ On cherche à résumer la stabilité en loi du processus X_t uniquement à travers ses deux premiers moments

Considérations pratique pour apprécier la stationnarité

▷ On dispose d'un d'une série $\{x_t\}$ et on veut se faire une idée de sa stationnarité. Si elle est stationnaire, alors :

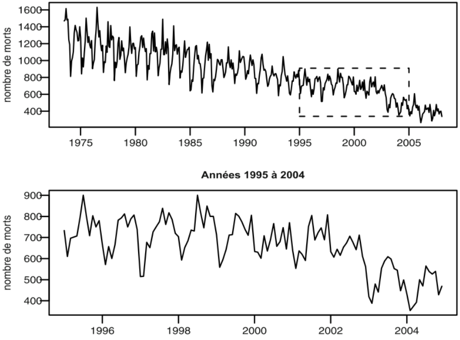
▷ La moyenne et la variance de la série soient constantes. Cela implique que :

- le graphique en fonction du temps montre un niveau moyen à peu près constant
- des fluctuations à peu près de même ampleur autour de la moyenne, quelle que soit la date autour de laquelle on examine la série.

Considérations pratique pour apprécier la stationnarité : Exemples

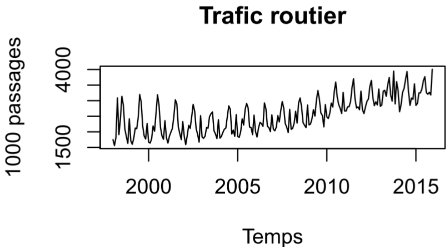
▷ Exemple : Morts par accident. Le nombre mensuel de morts et blessés graves par accident de la route en France (Fig. haut) montre d'importantes fluctuations saisonnières.

Figure: Accidents de la route - morts à 30 jours.
Figure en bas : zoom de la série sur quelques années



▷ Décroissance du niveau moyen de la série, qui s'accélère à partir de 2000, cette décroissance est accompagnée d'une diminution de la variabilité.

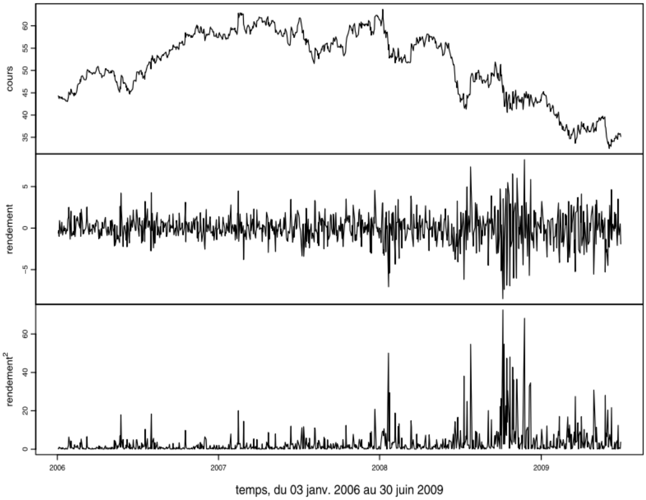
Considérations pratique pour apprécier la stationnarité : Exemples (suite)



- Moyenne et variance ne sont pas constantes
- Évolution générale : le nombre de passages augmente au cours du temps
- Évolution au cours de l'année : le nombre de passages est plus important l'été et pendant les vacances de Noël
- On constate des irrégularités

Exemple Danone

Figure: Danone - cours quotidien de l'action Danone , rendement, rendement au carré.



Exemple Danone (suite)

- (1) le cours est assez imprévisible : peut atteindre des valeurs très élevées aussi bien que très basses.

- (2) le rendement est presque de Moyenne Nulle. Élevé en valeur absolue pendant plusieurs périodes, puis rester faible un certain temps.

- (3) le rendement est presque nul, la moyenne de son carré dans un intervalle de quelques jours consécutifs représente la variance quotidienne dans cet intervalle.

▷ En général : On observe de faibles valeurs sur de longues périodes, puis de fortes valeurs sur d'autres périodes $\Rightarrow$ la variance à une date dépend des valeurs passées.

$$\Downarrow$$

▷ C'est un exemple de série à hétéroscédasticité conditionnelle :

- Le rendement d'une action est souvent de moyenne nulle
- les rendements à 2 dates consécutives souvent non corrélés
- Mais la variabilité du rendement qui, elle, montre une corrélation entre dates consécutives peut être prédite.
- Les modèles de séries à hétéroscédasticité conditionnelle sont susceptibles d'accomplir cette prévision.

Exemples des processus stationnaires : Bruit blanc

▷ Un processus stationnaire particulièrement simple est le processus purement aléatoire, appelé bruit blnac, constitué d'une suite de v.a. indépendante

Définition 3

$(\epsilon_t, t \in \mathbb{Z})$ est dit bruit blanc
(i) [faible] si

- $\mathbb{E}(\epsilon_t) = 0$ (moyenne nulle)
- $\mathbb{E}(\epsilon_t \epsilon_{t-h}) = 0$ pour tout t et $h \neq 0$ (non corrélées et pas nécessairement indépendantes) $\Leftrightarrow \quad \gamma(h) = 0, h \neq 0$
- $Var(\epsilon_t) = \sigma^2 = \gamma(0)$ (de même variance)
- On note $\epsilon_t \hookrightarrow BB(0, \sigma^2)$. C'est une série faiblement stationnaire.

(ii) [fort] lorsque les variables $(\epsilon_t)_{t \in \mathbb{Z}}$ sont indépendantes.

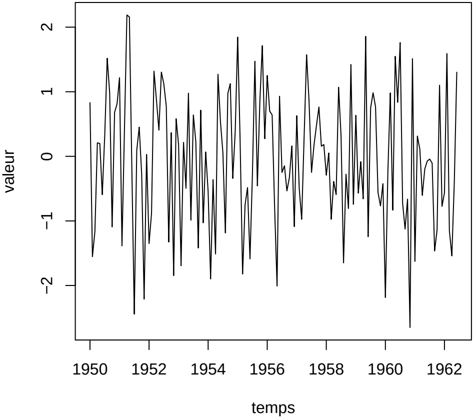
(iii) [très fort] lorsque les variables $(\epsilon_t)_{t \in \mathbb{Z}}$ sont i.i.d.

(iv) [gaussien], si $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$, on note $BBN(0, \sigma^2)$.
C'est une série strictement stationnaire.

Les trajectoires des bruits blancs peuvent avoir des formes très variées.

Création d'une série mensuelle, à partir de janvier 1950 bbg = ts(rnorm(150,0,1),start=c(1950,1),frequency=12)
plot.ts(bbg,xlab='temps',ylab='valeur')

Figure: Réalisation d'un bruit blanc



Ces notions, plus spécifiques à l'étude de série temporelle, renseignent sur la dépendance entre les données x_t.

Définition 4 (Fonctions d'autocovariances)

La fonction

$$h \mapsto \gamma(h) = \mathrm{Cov}(X_t, X_{t+h}) \quad \forall t \in T,\ h \in \mathbb{Z},$$

est appelée fonction d'auto-covariance d'ordre (ou de décalage) h (lag-h autocovariance). Elle apporte de l'information sur la variabilité de la série et sur les liaisons temporelles. Elle vérifie

- $\gamma(0) = \mathrm{Var}(X_t)$: un processus stationnaire a une variance constante
- $|\gamma(h)| \leq \gamma(0),\ \forall h.$
- $\gamma(h) = \gamma(-h),\ \forall h$. On ne la représente que pour $h = 0, 1, \ldots$.

Indice de dépendance : Coefficient d'autocorrélation et Fonction d'autocorrélation

Définition 5

i) Le coefficient d'autocorrélation (théorique) d'ordre h est

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

C'est le coefficient linéare simple calculé entre la série et cette même série décalée de h périodes de temps.

ii) La fonction d'autocorrélation (théorique) de $\{X_t\}$ est :

$$h \mapsto \rho(h) = \mathrm{Corr}(X_t, X_{t-h}) = \frac{\mathrm{Cov}(X_t, X_{t-h})}{\sqrt{V(X_t)}\sqrt{V(X_{t-h})}} = \frac{\gamma(h)}{\gamma(0)}$$

$\rho(\cdot)$ est à valeurs dans $[-1, +1]$, et $\rho(0) = 1$.

Fonction d'autocorrélation (suite)

▷ $\rho(h)$ mesure la corrélation entre X_t et X_{t-h}, puisque par stationnarité

$$\frac{Cov(X_t, X_{t-h})}{\sqrt{X_t}\sqrt{X_{t-h}}} = \frac{\gamma(h)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(h)}{\gamma(0)}$$

- $\rho(0) = 1$
- ρ mesure le degré de dépendance à la date t de ce qui s'est passé à la date $t-h$.
- La représentation graphique de $\rho(h)$ est appelée corrélogramme (théorique), on note FAC ou ACF (sortie de R).

  C'est une représentation graphique de la "mémoire" du processus, qui montre dans quelle mesure ses réalisations courantes sont influencées par ses réalisations passées.
- Le concept de corrélogramme n'a de sens que pour un processus stationnaire.

▷ Exemple 1. Soit X_t un bruite blanc (suite de v.a. iid)

$$\rho_h = \begin{cases} 1 & ,h=0 \\ 0 & ,\text{sinon} \end{cases}$$

Donc le processus est stationnaire.

▷ Exemple 2. Soit Y une v.a. telle que $var(Y) = \sigma^2$, et (Y_t) une suite de v.a. telle que $Y_1 = Y_2 = \ldots = Y_t = \ldots = Y$. Alors

$$\rho(h) = 1 \quad \forall h$$

Donc le processus est stationnaire. Cependant ce processus est complètement différent du premier.

- Pour $\{X_t\}$, connaître sa valeur à un moment donné t n'a rien à voir avec les autres valeurs.

- Pour $\{Y_t\}$, connaître Y_1 donne les valeurs de tous les autres Y_t.

En outre, on a par la loi du grand nombres

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E(X_1) = \mu, \quad n \to \infty$$

Mais

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = Y$$

Il y a autant de variabilité aléatoire dans le nième échantillon moyen que dans la première observation pour le processus $\{Y_t\}$. Pour éviter de telles situations, nous introduisons la définition suivante.

Définition 6

Si la moyenne d'échantillon formée à partir d'un trajectoire d'un processus converge vers le paramètre sous-jacent du processus, le processus est appelé ergodique.

Exemples de corrélogrammes

Les exemples suivants permettent de vérifier que la FAC joue le rôle de décomposition temporelle de la série.

$$\begin{aligned}
X_{1t} &= \epsilon_t \qquad \epsilon_t : \text{v.a. gaussienne} \\
X_{2t} &= a\ t + b + \epsilon_t \quad t = 1, \ldots, n \quad (t = tandance) \\
X_{3t} &= S_t + \epsilon_t \quad (S_t = \text{saisonnalité de quatre période}) \\
X_{3t} &= a\ t + b + S_t + \epsilon_t
\end{aligned}$$

- Les termes du corrélogramme de X_{1t} sont tous faibles à l'exception d'un seul, ceci est d'une v.a. de type bruit blanc
- Le corrélogramme de X_{2t} avec une décroissance lente de ses termes est spécifique d'une série affectée d'une tendance.
- Les termes très élevés (4, 8, 12 et 16) du corrélogramme de X_{3t} révèlent la saisonnalité de quatre périodes
- Le corrélogramme de X_{4t} est une combinaison des trois précédents.

Exemples de corrélogrammes

Figure: Représentation des processus analysés

Les corrélogrammes respectifs de chacun des processus sont présenté[s]
figure 1.4.

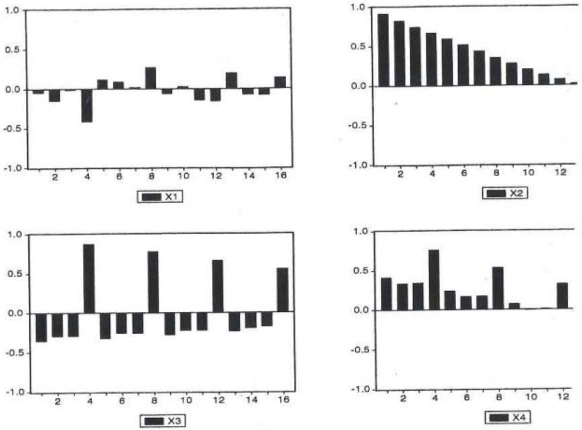


Figure 1.4 – Corrélogrammes des quatre processus

▷ **Exemple 1**.

Soit $(\epsilon_t)_t$ un processus de bruit blanc de variance σ^2.

On considère le processus stationnaire X définie par :

$$X_t = \epsilon_t - \epsilon_{t-12}, \quad t \in \mathbb{Z}.$$

▷ La fonction d'autocovariance est

$$\gamma(h) = \begin{cases} 2\sigma^2 & si\ \ h = 0 \\ -\sigma^2 & si\ \ h = \pm 12 \\ 0 & sinon \end{cases}$$

▷ La fonction d'autocorrélation est

$$\rho(h) = \begin{cases} 1 & si\ \ h = 0 \\ -\frac{1}{2} & si\ \ h = \pm 12 \\ 0 & sinon \end{cases}$$

Fonction d'autocorrélation partielle (rappel espérance conditionnelle)

▷ **Espérance conditionnelle si (X_t) gaussien**

Définition 7

Rappel : espérance conditionnelle

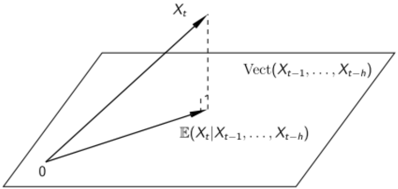
$$X^*_{t,h} = \mathbb{E}(X_t | X_{t-1}, \ldots, X_{t-h}) = a_1(h)X_{t-1} + \ldots + a_h(h)X_{t-h}$$

- $X^*_{t,h}$ est la a "meilleure" approximation de X_t par une combinaison linéaire de $(X_{t-1}, \ldots, X_{t-h})$ dans l'espace L^2 muni du produit scalaire $< X, Y > = \mathbb{E}(XY)$.
- $X^*_{t,h}$ est la régression affine de X_t sur $(X_{t-1}, \ldots, X_{t-h})$

▷ On a donc

$$X_t = X^*_{t,h} + \epsilon_{t,h}$$

$\epsilon_{t,h}$ est une variable aléatoire non corrélée avec $X_{t-1}, \ldots, X_{t-h}$.



$$\mathbb{E}(X_t|X_{t-1},\ldots,X_{t-h}) = a_1(h)X_{t-1} + \ldots + a_h(h)X_{t-h}$$

Définition 8 (PACF)

La fonction d'autocorrélation partielle r de (X_t) est définie par

$$\forall h \in \mathbb{Z}, \qquad r(h) = a_h(h) =: a_{hh}$$

Le graphe de cette fonction est appelé corrélogramme partiel.

Fonction d'autocorrélation partielle- Propriété

Propriété 1 (Propriété)

$$\begin{aligned} X_t &= a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_{h-1} X_{t-h+1} + U \\ &= X^*_{t,h-1} + U \end{aligned}$$

$$\begin{aligned} X_{t-h} &= b_1 X_{t-1} + b_2 X_{t-2} + \ldots + b_{h-1} X_{t-h+1} + V \\ &= X^*_{t-h,h-1} + V \end{aligned}$$

$$\forall h \geq 1, \quad r(h) = \operatorname{corr}(U, V) = \operatorname{corr}(X_t - X^*_{t,h-1}, \;\; X_{t-h} - X^*_{t-h,h-1})$$

- $r(h)$ représente le coefficient de corrélation linéaire entre
  - le résidu $U = X_t - X^*_{t,h-1}$ de la régression de X_t sur $X_{t-h+1}, \ldots, X_{t-1}$

  et
  - le résidu $V = X_{t-h} - X^*_{t-h,h-1}$ de la régression de X_{t-h} sur $X_{t-h+1}, \ldots, X_{t-1}$.

La fonction d'autocorrélation partielle : Remarques

Interprétation. Corrélation entre : X_t et X_{t-h} lorsque l'influence des variables $X_{t-1},\ldots,X_{t-h+1}$ a été retirée.

Elle vérifiée

- $r(h) = r(-h), \quad \forall h \neq 0, \quad \text{et} \quad r(1) = \rho(1).$
- Si (X_t) est stationnaire, $r(h)$ ne dépend que de la distance h.

X=rnorm(100)   as.vector(pacf(X))

- Les a_{hh} ont même interprétation que les coefficients d'une régression linéaire
- Si à partir d'une valeur h donné le corrélogramme de $r(h)$ est nul, c'est qu'il inutile de rechercher plus loin dans le passé une explication linéaire du présent.
- Un corrélogramme est nul lorsque toutes les corrélations sont nulles ($r(1)$ mis à part, qui vaut toujours 1). Ceci se produit notamment pour le processus bruit blanc. Ainsi, un processus stationnaire centré est un bruit blanc si et seulement si son corrélogramme est nul.

1. Calculer $r(1)$.
2. En faisant la régression de X_{t-1} sur X_{t-2} et X_t, calculer $r(2)$.

Rappel : Dans une régression

$$Y_i = \alpha X_i + \epsilon_i,$$

le coefficient α peut être calculé comme

$$\alpha = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \text{Corr}(X,Y)\frac{\sigma_Y}{\sigma_X} = \text{Corr}(X,Y) \quad \text{lorsque} \quad \sigma_Y = \sigma_X$$

$\triangleright$ En déduire que

$$r(h) = 0 \quad \forall\ h > 1$$

Lien entre autocorrélation et autocorrélation partielle

On a

$$r(1) = a_{1,1} = \frac{\mathrm{Cov}(X_t, X_{t-1})}{\sigma_X^2} = \rho(1)$$

▷ Déterminons maintenant une expression explicite de $r(2)$. Pour cela, écrivons

$$X_t = a_{0,2} + a_{1,2}X_{t-1} + a_{2,2}X_{t-2} + \epsilon_{t,2}$$

En multipliant par X_{t-1} et X_{t-2}, puis en prenant l'espérance, on obtient

$$\begin{cases} \rho(1) & = a_{1,2} + a_{2,2}\rho(1) \\ \rho(2) & = a_{1,2}\rho(1) + a_{2,2} \end{cases}$$

On en déduit

$$r(2) = a_{2,2} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}}$$

$$\begin{aligned} a_{11} &= \rho_1 \\ a_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ a_{33} &= \frac{\rho_1^3 - \rho_1\rho_2(2-\rho_2) + \rho_3(1-\rho_1^2)}{1 - \rho_2^2 - 2\rho_1^2(1-\rho_2)} \end{aligned}$$

$$a_{ii} = \begin{cases} \rho_1 & si \quad i = 1 \\ \frac{\rho_i - \sum_{j=1}^{i-1} \rho_{i-1,j} \ \rho_{i-j}}{1 - \sum_{j=1}^{i-1} \rho_{i-1,j} \ \rho_{ij}} & i = 2, \dots, k. \end{cases}$$

où

$$a_{ij} = \rho_{i-1,j} - \rho_{ii}\rho_{i-1,i-j} \quad j = 1, 2, \dots, i-1 \quad \rho_{i,i} = \rho_{ii}$$

En pratique le calcul des coefficients utilise les équations de Yule-Walker.

Fonction d'autocorrélation empirique

Définition 9

i) X_t une série observée, $t = 1, \ldots, n$. L'autocovariance empirique d'ordre h :

$$\widehat{\gamma}_h = \frac{\sum_{t=h+1}^{n}(X_t - \overline{X})(X_{t-h} - \overline{X})}{n-h} \quad \text{où} \quad \overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t, \quad 0 \leq h \leq n-1. \tag{1}$$

ii) Le coefficient d'autocorrélation empirique d'ordre h est

$$\widehat{\rho}_h = \frac{n}{n-h} \frac{\sum_{t=h+1}^{n}(X_t - \overline{X})(X_{t-h} - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}, \quad 0 \leq h \leq n-1. \tag{2}$$

iii) La fonction $h \mapsto \widehat{\rho}_h$, $h = 0, 1, \ldots$ est la fonction d'autocorrélation empirique, son graphique, noté ACF, est le corrélogramme empirique.

Propriété 2

$\widehat{\rho}_h$ est un estimateur convergent de ρ_h quand $n \to \infty$.

Remarque. En pratique on utilise ces quantités que lorsque $n > 50$ et $h \leq n/4$.

▷ Loi approximative de $\widehat{\rho}_h$

Propriété 3

Si $X_t, t = 1, \ldots, n$ est une observation d'une suite de v.a. i.i.d. avec $\mathbb{E}(X_t^2) < \infty$, les $\widehat{\rho}_h$ sont approximativement indépendants et

$$\widehat{\rho}_h \hookrightarrow \mathcal{N}(0, \frac{1}{n}).$$

Propriétés de la fonction d'autocorrélation : Remarques

1. Si $\{X_t\}$ est une suite de v.a. i.i.d., alors les $\widehat{\rho}_h$ doivent se trouver dans des intervalles contenant 0.
2. Une série stationnaire présente souvent une corrélation à court terme, une valeur élevée de $\rho(1)$ suivie de quelques corrélations plus petites qui deviennent ensuite de plus en plus petites.
3. Dans une série alternée, $\widehat{\rho}_k$ alterne des valeurs positives et négatives. Un exemple typique d'une telle série est le modèle AR (1) avec un coefficient négatif, $Y_t = -\phi Y_{t-1} + Z_t$, où ϕ est un paramètre positif et Z_t sont des variables aléatoires i.i.d.
4. Si le caractère saisonnier existe dans la série, cela sera reflété dans ACF. En particulier, si $X_t = a \cos t\omega$, on peut montrer que $\widehat{\rho}_k \simeq \cos k\omega$
5. Dans le cas des séries non stationnaires, $\widehat{\rho}_k$ ne diminue pas pour les grandes valeurs de k. Ceci est une indication de non-stationnarité et peut être causé par de nombreux facteurs.

Analyse des fonctions d'autocorrélations

▷ L'objectif de l'étude précédente était d'obtenir une série stationnaire

▷ L'étape suivante est de modéliser cette série

▷ La première chose à faire est de tester s'il y a dépendance entre ces termes

▷ Si ce n'est pas le cas, on dit que la série est un bruit blanc, il n'est pas donc utile d'aller plus loin dans la modélisation si ce n'est d'estimer la moyenne et variance du bruit blanc

▷ Quelles sont les termes $\widehat{\rho}_h$ qui sont significativement différents de 0 ?

▷ Si aucun terme n'est significativement différent de 0, on en conclure que le processus étudié est sans mémoire et donc qu'à ce titre il n'est affecté ni de tendance ni de saisonnalité.

Il n'est pas toujours facile de vérifier si une série est formée de v.a. i.i.d. Main on peut tester la nullité de coefficients d'autocorrélations, ρ_h (objectif du test du portemanteau).

On aimerait savoir si on est en présence d'un bruit blanc.

Proposition 1

Si $(\epsilon_k)_{k\geq 1}$ est un bruit blanc, alors ses fonctions d'auto-corrélation empiriques vérifient

$$\sqrt{n}\widehat{\rho}(h) \hookrightarrow \mathcal{N}(0,1), \quad n \to \infty, \quad \forall h \in \mathbb{N}^*$$

De plus, les $(\widehat{\rho}(h))_{h\geq 1}$ sont asymptotiquement indépendant (quand $n \to \infty$).

A. Test d'hypothèses simple.

On veut tester

$(H0)$ Les (ϵ_k) forment un bruit blanc.

contre

$(H1)$ Les (ϵ_k) ne forment pas un bruit blanc.

▷ La statistique de test utilisée est $\widehat{\rho}(1)$

▷ . Si $\alpha = 0,05$. La Proposition précédente nous dit que, si $(H0)$ est vraie,

$$\mathbb{P}(\sqrt{n}\widehat{\rho}(1) \in [-1.96,\ 1.96]) \longrightarrow \mathbb{P}(Z \in [-1.96,\ 1.96]) = 0.05$$

avec $Z \hookrightarrow \mathcal{N}(0,1)$.

▷ **Règle de décision :**

- si $\widehat{\rho}(1) \in [-1.96/\sqrt{n},\ 1.96//\sqrt{n}]$ alors on considère que (H0) est vraie,

- dans le cas contraire, on considère que (H0) n'est pas vraie.

▷ Si (H0) est vraie, la probabilité de rejeter (H0) à tort est

$$\mathbb{P}(\text{rejeter } (H0)|(H0)) \simeq \mathbb{P}(Z \notin [-1.96,\ 1.96]) = 0.05$$

▷ Pour avoir un bruit blanc, il est nécessaire qu'aucune valeur de l'autocorrélogramme ne soit significativement non-nulle.

Test de Box-Pierce, ou test de "portmanteau"

▷ Dans la pratique, on utilise des statistiques plus compliquées

Le test de **Box-Pierce** permet d'identifier les processus de bruit blanc. Elle permet de tester

$$cov(\epsilon_t,\ \epsilon_{t-h}) = 0 \quad \text{pour tout } h, \quad \text{soit} \quad \rho(h) = 0 \quad \text{pour tout } h.$$

▷ **Ce qui est équivaut à tester** :

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_H = 0 \quad \text{contre} \quad H_1 :\ \exists\ \rho_i \neq 0.$$

▷ **Statistique de test de Box-Pierce (ou du portmanteau)** :

$$Q(h) = n \sum_{k=1}^{h} \widehat{\rho}_k^2 = \sum_{k=1}^{n} \left(\frac{\widehat{\rho}_k - 0}{1\sqrt{n}} \right)^2$$

où h=nombre de retards (ou décalages)

$\widehat{\rho}_k$= autocorrélation empirique d'ordre k.

$$\text{Sous} \quad H_0 \qquad Q \hookrightarrow \quad \chi_h^2$$

▷ On rejette l'hypothèse de bruit blanc, au seuil α, si $Q > q_\alpha$ quantile de χ_h^2 au seuil $1-\alpha$

Test de Box-Pierce, ou test de "portmanteau" (suite)

▷ Pour des petits échantillons, on utilise la statistique de Lung-Box :

▷ On peut aussi utiliser la statistique de Ljung zt Box :

$$Q^*(h) = n(n+2)\sum_{k=1}^{h}\frac{\widehat{\rho}_k^2}{n-k}$$

▷ Elle a une distribution de parobabilité mieux approchée par un χ^2 que la statistique de Box-Pierce.

▷ Quand le test est appliqué non sur des v.a. indépendantes, mais sur les résidus d'un ajustement estimant q paramètres, la loi sous H_0 est un χ^2_{h-q}.

Test d'un ensemble de coefficients d'autocorrélation : Exemple

▷ Simuler 100 observation de X_t vérifiant :

$$\begin{aligned} X_t &= -0.7X_{t-1} + \epsilon_t, \qquad \text{série} \quad X1 \\ X_t &= -0.7X_{t-12} + \epsilon_t, \qquad \text{série} \quad X2 \end{aligned}$$

▷ Code :

```
require(caschrono)
set.seed(123)
x1=arima.sim(n=100, list(ar=-.7), sd=sqrt(4))
x2=arima.sim(n=100, list(ar=c(rep(0, 11),-.7)), sd=sqrt(4))
ret=c(3,6, 9, 12)
a1=Box.test.2(x1, nlag=ret, type="Ljung-Box", decim=2)
a2=Box.test.2(x2, nlag=ret, type="Ljung-Box", decim=2)
a12=cbind(a1, a2[,2])
colnames(a12)=c("Retard", "p-val. x1", "p-val.x2")
a12
```

Test d'un ensemble de coefficients d'autocorrélation : Exemple

	Retard	$p-val.x1$	$p-val.x2$
1,	3	0	0.66
2,	6	0	0.49
3,	9	0	0.43
4,	12	0	0.00

a) On rejette la blancheur $X1$ quel que soit le retard.

b) Pour $X2$, si on arrête à un ordre inférieur à 12, on peut conculre que la série est un bruit blanc.

- On s'arrêt habituellement à $k \simeq n/4$.
- Si la série est saisonnière période p (cas de $X2$), on doit choisir $h > p$.
- L'examen du corrélogramme empirique d'une série permet de faire une idée de sa stationnarité. Ainsi :
- Si $\{X_t\}$ est stationnaire, le corrélogramme $k \mapsto \widehat{\rho}(k)$, décroit exponentiellement vers 0.
- In versement, si $k \mapsto \widehat{\rho}(k)$, ne montre pas de décroissance rapide, cela est une indice d'une non-stationnarité.