

Feuille d'exercices

Exercice 1 (Autocorrélation et tendance linéaire).

On s'intéresse à l'impact que peut avoir l'estimation de l'autocorrélation à partir d'une série temporelle ayant une tendance linéaire marquée.

- a) Rappelez l'expression de l'autocorrélation empirique.
 b) On considère la série temporelle (purement déterministe) suivante

$$X_t = at + b, \quad t \in \mathbb{Z},$$

que l'on observe en $t = 0, \dots, n$.

Des calculs simples (mais un peu pénible pour les faignants comme moi !) conduisent à montrer que pour $h \in \mathbb{N}$ fixé, on a $\hat{\rho}(h) \rightarrow 1$ lorsque $n \rightarrow \infty$. Comment interprétez vous ce résultat ?

Conseil pour les courageux : $\sum_{j=1}^n j^2 = n(n+1)(2n+1)/6$.

- c) Lisez la documentation de la fonction `acf` de R. Illustrez numériquement la convergence de la première question.
 d) Reprenez la question précédente mais en rajoutant à la série purement déterministe un bruit blanc gaussien de variance que vous ferez varier. Que constatez vous ?
 e) On considère désormais la série

$$X_t = \sin\left(\frac{2\pi t}{\omega}\right), \quad t \in \mathbb{Z}.$$

A l'aide de R, générez cette série temporelle sur une grande période et tracez sa fonction d'autocorrélation. Vous choisirez une fréquence ω de votre choix. Que constatez vous ? Est en accord avec l'ACF des données 'aaaa...hhhh' vues en cours ?

Solution 1. a) Bon bah c'est du cours

$$\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0), \quad \hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^{n-h} (X_{t+h} - \bar{X})(X_t - \bar{X}).$$

- b) Prenons donc $X_t = at + b$. On a d'une part

$$\bar{X} = n^{-1} \left(a \sum_{t=1}^n t + nb \right) = a(n+1)/2 + b,$$

et d'autre part

$$\hat{\gamma}(0) = \sum_{t=1}^n (at - a(n+1)/2)^2 = a^2 \sum_{t=1}^n t^2 - 2a^2(n+1)/2 \sum_{t=1}^n t - a^2 n^2(n+1)^2/4 = a^2 n(n+1)(2n+1)/6 - 2a^2(n+1)/2 \sum_{t=1}^n t$$

On a également

$$(X_{t+h} - \bar{X})(X_t - \bar{X}) = \{a(t+h - (n+1)/2)\} \{a(t - (n+1)/2)\} = a(t - (n+1)/2) \{(t - (n+1)/2) + h\},$$

de sorte que



Exercice 2 (Causalité et inversibilité). On s'intéresse à la série temporelle suivante

$$X_t = 0.4X_{t-1} + 0.45X_{t-2} + \varepsilon_t + \varepsilon_{t-1} + 0.25\varepsilon_{t-2}.$$

- Est ce un processus *ARMA* ? Si oui, déterminez ses polynômes $\theta(B)$ et $\phi(B)$.
- Ce processus est-il causal ?
- Ce processus est-il inversible ?
- Donnez la représentation causale de cette série temporelle. Comparez vos résultats à ceux obtenus via **R** à l'aide de la fonction **ARMAtoMA**
- Donnez la représentation inversible de cette série temporelle. Comparez vos résultats à ceux obtenus via **R** à l'aide de la fonction **ARMAtoMA**. *Astuce : il suffit juste d'invertir les polynômes dans l'utilisation de la fonction ;-)*

Solution 2. a) Un simple jeu de ré-écriture et l'on trouve

$$(1 - 0.4B - 0.45B^2)X_t = (1 + B + 0.25B^2)\varepsilon_t.$$

C'est donc apparament un *ARMA*(2, 2) avec

$$\begin{aligned}\phi(z) &= 1 - 0.4z - 0.45z^2 \\ \theta(z) &= 1 + z + 0.25z^2.\end{aligned}$$

Cependant l'étude des racines de ces 2 polynômes montre qu'ils ont une racine commune -2 et l'on peut donc simplifier ce modèle de série temporelle. En effet on a

$$\begin{aligned}\phi(z) &= (1 + z/2)(1 - 9z/10) \\ \theta(z) &= (1 + z/2)^2\end{aligned}$$

Ainsi la série temporelle précédente est équivalente à

$$(1 - 0.9B)X_t = (1 + 0.5B)\varepsilon_t.$$

C'est donc un *ARMA*(1, 1).

- Pour la causalité il faut donc calculer les racines de $\phi(z)$. Cette dernière vaut $1/0.9 = 10/9$ et est plus grande que 1. La série est donc causale.
- Pour l'inversibilité il faut donc calculer les racines de $\theta(z)$. La racine vaut -2 qui est en module supérieure à 1. La série est également inversible.
- Pour donner sa représentation causale on se sert de l'égalité

$$\psi(z)\theta(z) = \phi(z),$$

soit ici

$$\begin{aligned}(1 - 0.9z) \sum_{k \geq 0} \psi_k z^k &= 1 + 0.5z, & \psi_0 &= 1 \\ \iff \sum_{k \geq 0} \psi_k z^k - \sum_{k \geq 0} 0.9\psi_k z^{k+1} &= 1 + 0.5z \\ \iff \psi_0 + \sum_{k \geq 1} (\psi_k - 0.9\psi_{k-1})z^k &= 1 + 0.5z.\end{aligned}$$

Ainsi par identification on trouve

$$\psi_0 = 1, \quad \psi_1 - 0.9\psi_0 = 0.5, \quad \psi_{k+1} - 0.9\psi_k = 0, k \geq 1.$$

On trouve donc que $\psi_0 = 1, \psi_1 = 1.4, \psi_{k+1} = 0.9\psi_k = 0.9^k\psi_1 = 1.4 \times 0.9^k, k \geq 1$.

La série temporelle admet donc la représentation causale

$$X_t = \varepsilon_t + 1.4 \sum_{k \geq 1} 0.9^k \varepsilon_{t-k}.$$

e) On fera de même pour la représentation inversible.



Exercice 3 (Vraisemblance d'un $AR(1)$). On s'intéresse ici à la série temporelle

$$X_{t+1} = \mu + \phi(X_t - \mu) + \varepsilon_{t+1}, \quad t \in \mathbb{Z}, \quad |\phi| < 1,$$

où $\varepsilon_1, \varepsilon_2, \dots$ sont des v.a.i.i.d. $N(0, \sigma^2)$. On suppose de plus que l'on dispose de n observations $x_1, \dots, x_n$ de cette série temporelle.

a) Quels sont les paramètres du modèle statistique considéré ici ?

b) Montrez que pour tout $t \in \mathbb{Z}$ on a la propriété Markovienne

$$\Pr(X_{t+1} \leq a_{t+1} \mid X_t = a_t, \dots, X_0 = a_0) = \Pr(X_{t+1} \leq a_{t+1} \mid X_t = a_t).$$

c) En déduire que la vraisemblance issue des observations $x_1, \dots, x_n$ est donnée par

$$f(x_1)f(x_2 \mid x_1)f(x_3 \mid x_2) \cdots f(x_n \mid x_{n-1}).$$

d) Donnez la loi de $X_{t+1} \mid X_t = x_t$.

e) En utilisant la représentation causale de cet $AR(1)$, montrez que x_1 est une réalisation d'une $N\{\mu, \sigma^2/(1 - \phi^2)\}$. Faites le lien avec le slide du cours portant sur la modélisation de la température corporelle d'un castor.

f) Ecrivez un code R permettant d'ajuster cet $AR(1)$ par maximum de vraisemblance (via optimisation numérique) et comparez vos résultats à ceux de la fonction `(s)arima`. Vous ajusterez ce modèle sur le jeu de données `1h` inclus dans R.

Solution 3. a) Les paramètres du modèles sont μ, ϕ, σ^2 .

b) On a

$$\begin{aligned} \Pr(X_{t+1} \leq a_{t+1} \mid X_t = a_t, \dots, X_0 = a_0) &= \Pr(\mu + \phi(X_t - \mu) + \varepsilon_{t+1} \leq a_{t+1} \mid X_t = a_t, \dots, X_0 = a_0) \\ &= \Pr(\varepsilon_{t+1} \leq a_{t+1} - \mu - \phi(X_t - \mu) \mid X_t = a_t, \dots, X_0 = a_0) \\ &= \Pr(\varepsilon_{t+1} \leq a_{t+1} - \mu - \phi(X_t - \mu) \mid X_t = a_t, \dots, X_0 = a_0), \\ &= \Pr(X_{t+1} \leq a_{t+1} \mid X_t = a_t). \end{aligned}$$

c) La vraisemblance vaut, pour $\theta = (\mu, \phi, \sigma^2)^\top$,

$$\begin{aligned} L(\theta) &= f(x_1, \dots, x_n) \\ &= f(x_n \mid x_1, \dots, x_{n-1})f(x_1, \dots, x_{n-1}) \\ &= f(x_n \mid x_{n-1})f(x_1, \dots, x_{n-1}), \quad \text{prop. Markovienne} \\ &= \dots \\ &= f(x_1) \prod_{t=1}^{n-1} f(x_{t+1} \mid x_t). \end{aligned}$$

d) Puisque ε_t est un bruit blanc Gaussien, on a, conditionnellement à $X_t = x_t$,

$$X_{t+1} = \mu + \phi(x_t - \mu) + \varepsilon_{t+1} \sim N(\mu + \phi(x_t - \mu), \sigma^2).$$

e) Travaillons sur la série centrée $Y_t = X_t - \mu$. Pour cette série on a

$$\phi(B)Y_t = \theta(B)\varepsilon_t, \quad \phi(B) = 1 - \phi B, \quad \theta(B) = 1.$$

La racine de $\phi(z)$ est $1/\phi$ qui est en module plus grande que 1 donc la série admet une représentation causale

$$Y_t = \sum_{j \geq 0} \psi_j \varepsilon_{t-j}.$$

Trouvons le polynôme $\psi(z)$. On a pour tout $z \in \{z \in \mathbb{C} : |z| < 1\}$,

$$\begin{aligned} \psi(z) &= \frac{\theta(z)}{\phi(z)} = \frac{1}{\phi(z)} \\ &= \frac{1}{1 - \phi z} \\ &= \sum_{j \geq 0} \phi^j z^j. \end{aligned}$$

On en déduit par identification que $\psi_j = \phi^j$, $j \geq 0$.

En revenant à notre série initiale, on trouve alors

$$X_t = \mu + \sum_{j \geq 0} \phi^j \varepsilon_{t-j}.$$

f) D'après la question précédente on a

$$X_1 = \mu + \sum_{j \geq 0} \phi^j \varepsilon_{1-j}.$$

Identifions la loi de X_1 via sa transformée de Laplace. Posons $S_n = \sum_{j=0}^n \phi^j \varepsilon_{1-j}$. On a

$$\begin{aligned} \mathbb{E}(e^{tS_n}) &= \mathbb{E}\left(e^{t \sum_{j=0}^n \phi^j \varepsilon_{1-j}}\right) \\ &= \mathbb{E}\left(\prod_{j=0}^n e^{t\phi^j \varepsilon_{1-j}}\right) \\ &= \prod_{j=0}^n \mathbb{E}\left(e^{t\phi^j \varepsilon_{1-j}}\right), \quad \text{par ind.} \\ &= \prod_{j=0}^n \exp\left(\frac{t^2 \phi^{2j} \sigma^2}{2}\right), \quad \text{trans. Laplace } N(0, \sigma^2) \\ &= \exp\left(\sum_{j=0}^n \frac{t^2 \phi^{2j} \sigma^2}{2}\right) \\ &= \exp\left(\frac{t^2(1 - \phi^{2(n+1)})\sigma^2}{2(1 - \phi^2)}\right) \\ &\longrightarrow \exp\left(\frac{t^2 \sigma^2}{2(1 - \phi^2)}\right), \quad n \rightarrow \infty. \end{aligned}$$

Cette dernière expression est la transformée de Laplace d'une $N(0, \sigma^2/(1 - \phi^2))$. Ainsi la loi de X_1 est une $N(\mu, \sigma^2/(1 - \phi^2))$.

C'est la fameuse loi stationnaire indiquée en cours.

TABLE 1 – AIC pour différents modèles *ARIMA*, d fixé à ???, sur les données de température globale.

$p \backslash q$	0	1	2	3	4
0					
1					
2					
3					
4					

g) Rien de plus simple.



Exercice 4 (Un peu de lecture). Allez sur la page web <https://www.stat.pitt.edu/stoffer/tsa4/Rissues.htm> et lisez là. C'est très instructif. Forcez vous à bien comprendre les incohérences que l'auteur pointe.



Exercice 5 (Modélisation de la température globale). Dans cet exercice, on va tenter de modéliser la température globale terrestre.

- Installez la librairie **astsa** de R fournissant le jeu de données **gtemp** que nous allons utiliser. Renseignez vous sur ces données en lisant sa documentation.
- Commencez par tenter d'identifier l'ordre de votre modèle *ARIMA*.
- Ajustez votre modèle mais considérez également d'autres modèles *ARIMA* concurrents.
- Faites un tableau (synthétique) reportant l'AIC pour chacun des modèles ajustés considérés comme par exemple le Tableau 1.
- Analysez les résidus de votre meilleur modèle et en déduire si ce dernier est satisfaisant.
- Faites une prédiction pour les 50 prochaines années pour votre meilleur modèle.

