Examen du 17/03/2023

MAP-STA2 : Séries chronologiques

Yannig Goude [yannig.goude@edf.fr]
Bianca Marin-Moreno [bianca.marin-moreno@edf.fr]

Exercice 1 /2

1. Donner une définition de la stationnarité forte et faible. /0.5

cf cours

2. Donner un exemple de processus stationnaire fort. /0.5

X_t Gaussien de moyenne, variance et covariance constantes.

3. Une somme de processus stationnaire est elle stationnaire? /1

Non, il faut de plus voir l'indépendance. On peu sinon construire un processus non stationnaire. Par exemple, si X_t stationnaire faible t.q $E(X_t) = 0$, $Y_t = (-1)^t X_t$ est stationnaire mais $X_t + Y_t$ ne l'est pas.

Exercice 2 /3 voir prenat

Soit Z_t une suite de gaussiennes iid centrée de variance 1. a, b e c des constantes réelles. Les processus suivants sont ils stationnaires faibles? Si oui donner leur moyenne et leur fonction d'autocovariance.

1. $X_t = a + bZ_t + cZ_{t-1}$ /0.5

$E(X_t) = a$, $V(X_t) = b^2 + c^2$,

$cov(X_, X_{t+h}) = bc$ si $h = 1$ ou $h = -1$, $cov(X_, X_{t+h}) = 0$ pour $|h| > 1$

donc X_t est stationnaire.

2. $X_t = a + bZ_0$ /0.5

$E(X_t) = a$, $cov(X_, X_{t+h}) = b^2$ donc X_t est stationnaire.

3. $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$ /0.5

$E(X_t) = 0$, $cov(X_, X_{t+h}) = \cos(ct)\cos(c(t+h))$ donc X_t n'est pas stationnaire (sauf si c est multiple de π).

4. $X_t = Z_0 \cos(ct)$ /0.5

$E(X_t) = 0$, pour la fonction d'autocovariance on distingue les cas:

- $h = 0$, $cov(X_, X_{t+h}) = 1$
- $h = 1$, $cov(X_, X_{t+h}) = cos(ct)\sin(c(t+1))$
- $h = -1$, $cov(X_, X_{t+h}) = \cos(c(t-1))\sin(ct)$
- 0 sinon

donc X_t n'est pas stationnaire (sauf si c est multiple de π).

5. $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$ /0.5

$E(X_t) = 0$, $cov(X_t, X_{t+h}) = 1$ si $h = 0$, 0 sinon.

donc X_t est stationnaire.

6. $X_t = Z_t Z_{t-1}$ /0.5

$E(X_t) = E(Z_t)E(Z_{t-1}) = 0$, $V(X_t) = \sigma^4$, $cov(X_t, X_{t+h}) = 0$ pour h différent de 0.

Exercice 3 /3

Soit les processus stationnaires du second ordre Y_t et X_t définis par:

- $Y_t = \phi_1 Y_{t-1} + aX_t + u_t$
- $X_t = \phi_2 X_{t-1} + v_t$

avec u et v deux bruits blancs décorrélés de variances respectives σ_u^2 et σ_v^2, $0 < |\phi_1| < 1$ et $0 < |\phi_2| < 1$.

1. soit $W = (1 - \phi_1 L)(1 - \phi_2 L)Y_t$, montrer que W est stationnaire du second ordre. Calculer sa fonction d'autocovariance γ_W et en déduire que W est un processus MA à déterminer (on supposera que $a^2\sigma_v^2 + (1 + \phi_2^2)\sigma_u^2 > 2\phi_2\sigma_u^2$). /2

   On remarque que $W_t = (1 - \phi_1 L)(1 - \phi_2 L)Y_t = av_t + u_t - \phi_2 u_{t-1}$. Il s'en suite que $E(W) = 0$ et $\gamma(0) = a^2\sigma_v^2 + (1 + \phi_2^2)\sigma_u^2$, $\gamma(1) = \gamma(-1) = -\phi_2\sigma_u^2$, $\gamma(h) = 0, \forall |h| > 1$.

   On peut donc constater qu'il s'agit d'une fonction d'autocovariance d'un MA(1). On cherche donc le polynôme de degré 1 Θ et l'innovation ε de variance σ_ε^2 tel que $W_t = \Theta(L)\varepsilon_t$. Soit λ tel que $|\lambda| < 1$ et le processus $Z = (1 - \lambda)\varepsilon$. On cherche ensuite à déterminer λ et ε tel que $Z = W$. Pour $h = 0, 1$, $\gamma_Z(h) = \gamma_W(h)$ implique:

$$\begin{array}{lcl} \lambda^2 - \frac{a^2\sigma_v^2 + (1+\phi_2^2)\sigma_u^2}{\phi_2\sigma_u^2}\lambda + 1 & = & 0 \\ \sigma_\varepsilon^2 & = & \frac{\phi_2\sigma_u^2}{\lambda} \end{array}$$

   $a^2\sigma_v^2 + (1 + \phi_2^2)\sigma_u^2 > 2\phi_2\sigma_u^2$, l'équation 1 on a donc deux racines réelles. Le produit des racines vaut 1, leur somme est positive, elles sont donc toutes les deux positives et la plus petite des deux est de module < 1. On choisit donc

$$\lambda = \frac{a^2\sigma_v^2 + (1 + \phi_2^2)\sigma_u^2 - \sqrt{(a^2\sigma_v^2 + (1 + \phi_2^2)\sigma_u^2)^2 - 4\phi_2\sigma_u^2}}{2\phi_2\sigma_u^2}$$

   et $\varepsilon = (1 - \lambda L)^{-1}W$ est par construction un bruit blanc de variance $\sigma_\varepsilon^2 = \frac{\phi_2\sigma_u^2}{\lambda}$.

2. montrer que Y est un processus ARMA à déterminer. /0.5

   On a $(1 - \phi_1 L)(1 - \phi_2 L)Y = (1 - \lambda L)\varepsilon$, avce $\lambda \neq \phi_1$ et $\lambda \neq \phi_2$ c'est donc un ARMA(2,1).

3. déterminer la prévision optimale à l'instant t notée Y_t^* et la variance de l'erreur de prévision $Var(Y_t - Y_t^*)$. /0.5

   Les racines de $\Phi(L) = (1 - \phi_1 L)(1 - \phi_2 L)$ sont de modules strictement supérieur à 1, on a donc un développement en moyenne mobile infinie $Y_t = \sum_{k>0} a_k \varepsilon_{t-k}$ et ε_t est l'innovation. D'ou $V(Y_t - Y_t^*) = \sigma_\varepsilon^2$ et $Y^* = Y - \varepsilon = (1 - \Phi(L)(1 - \lambda L)^{-1})Y$

Exercice 4 /2

Soit ε_t un bruit blanc faible de variance σ^2. Considérons le processus moyennes mobiles infinie:

$X_t = \sum_{i\in\mathbf{Z}} \theta_i \varepsilon_{t-i}, \ \ \sum_{i\in\mathbf{Z}} |\theta_i| < +\infty$

1. Déterminer la fonction d'autocovariance de X_t. /1

$$\sum_{i\in\mathbf{Z}} ||\theta_i \varepsilon_{t-i}|| = \sum_{i\in\mathbf{Z}} |\theta_i| ||\varepsilon_{t-i}|| = \sigma \sum_{i\in\mathbf{Z}} |\theta_i| < \infty$$

ou cf la propriété du cours: soit $(X_t)_{t\in\mathbf{Z}}$ un processus stationnaire d'espérance μ et de fonction d'autocovariance $\gamma(h)$, $(a_i)_{i\in\mathbf{Z}}$ une suite de nombres réels absolument sommable $\sum_{i\in\mathbf{Z}} |a_i| < \infty$, alors $Y_t = \sum_{i\in\mathbf{Z}} a_i X_{t-i}$ est un processus stationnaire.

ensuite:

$$\gamma(h) = \sigma^2 \sum_{j\in\mathbf{Z}} \theta_j \theta_{j+h}$$

2. Déterminer sa densité spectrale, montrer qu'il est équivalent de connaitre la fonction d'auto-covariance et la densité spectrale de ce processus. /1

la densité spectrale du processus est la fonction définie par:

$$f(\omega) = \frac{1}{2\pi} \sum_{h\in\mathbf{Z}} \gamma(h) \mathrm{e}^{i\omega h}, \quad \forall \omega \in \mathbf{R}$$

cette fonction existe car la série $(\gamma(h)\mathrm{e}^{i\omega h})_{h\in\mathbf{Z}}$ est absolument sommable.

D'autre part c'est une fonction réelle car γ est paire:

$$f(\omega) = \frac{1}{2\pi}[\gamma(0) + \sum_{h=1}^{+\infty} \gamma(h)(\mathrm{e}^{i\omega h} + \mathrm{e}^{-i\omega h})]$$

$$f(\omega) = \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{h=1}^{+\infty} \gamma(h) \cos(\omega h) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) \cos(\omega h)$$

On a défini en 2) la densité spectrale à partir de la fonction d'autocovariance. D'autre part on peut montrer que :

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) \cos(\omega h) d\omega = \int_{-\pi}^{\pi} f(\omega) \mathrm{e}^{i\omega h} d\omega$$

Tout d'abord notons que $\sum_{h=1}^{+\infty} |\gamma(h)| < +\infty$ car $\sum_{h=1}^{+\infty} |\gamma(h)| = \sigma^2 \sum_{h=1}^{+\infty} |\sum_{i=1}^{+\infty} \theta_i a_{i-h}| \leq \sigma^2 \sum_{h=1}^{+\infty} \sum_{i=1}^{+\infty} |\theta_i||\theta_{i-h}| = \sigma^2 (\sum_{i=-\infty}^{+\infty} |\theta_i)|)^2 < \infty$.

On peut donc intervertir $\int$ et $\sum$ et dérouler le calcul ci-dessous:

$$\int_{-\pi}^{\pi} f(\omega) \cos(\omega h) d\omega = \int_{-\pi}^{\pi} f(\omega) \mathrm{e}^{-i\omega h} d\omega$$

$$= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} \gamma(j) \mathrm{e}^{i\omega j} \mathrm{e}^{-i\omega h} d\omega$$

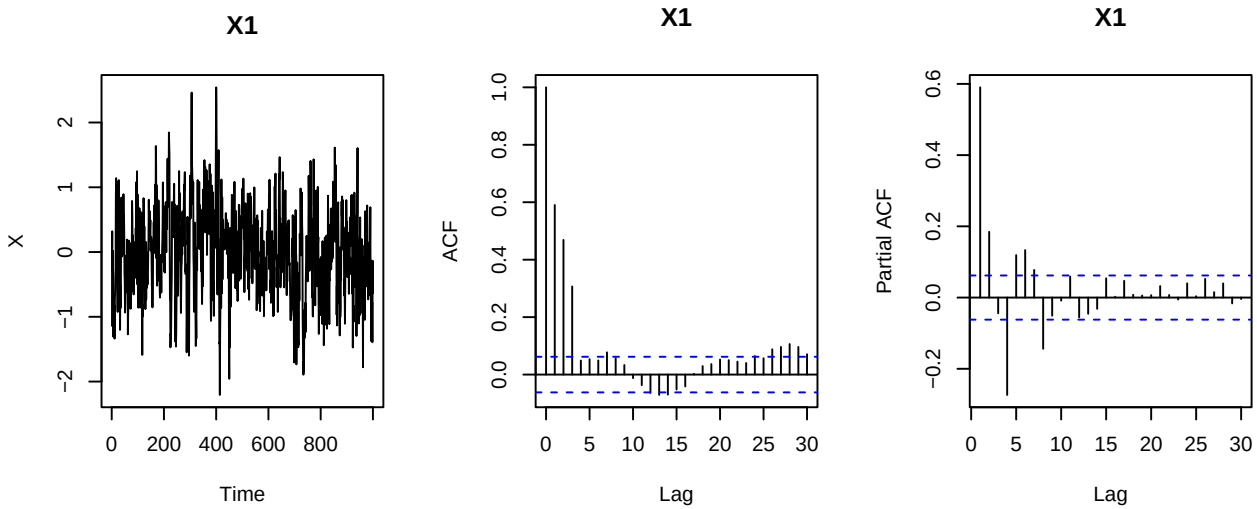
$$= \frac{1}{2\pi} \sum_{j=1}^{+\infty} \gamma(j) \int_{-\pi}^{\pi} \mathrm{e}^{i\omega(h-j)} d\omega$$

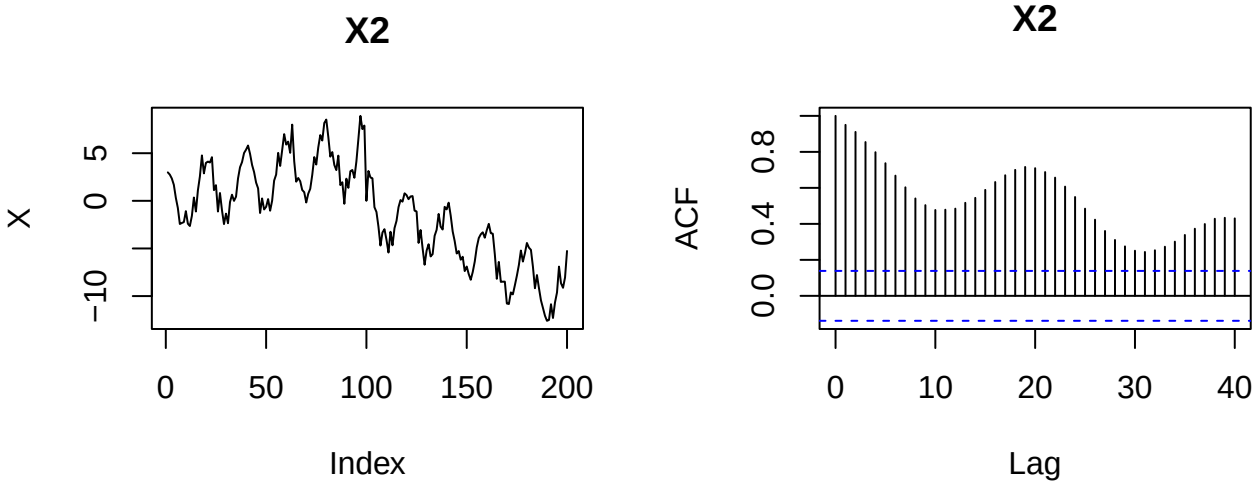
le terme $\int_{-\pi}^{\pi} \mathrm{e}^{i\omega(h-j)} d\omega$ vaut 2π si $j = h$, 0 sinon.

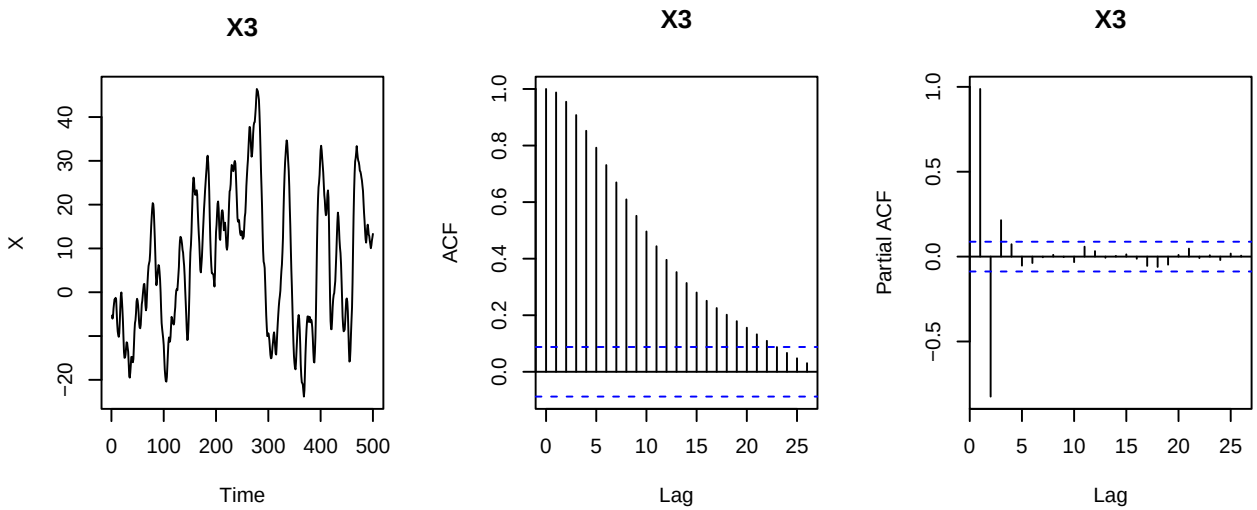
d'ou $\gamma(h) = \int_{-\pi}^{\pi} f(\omega) \mathrm{e}^{i\omega h} d\omega$.

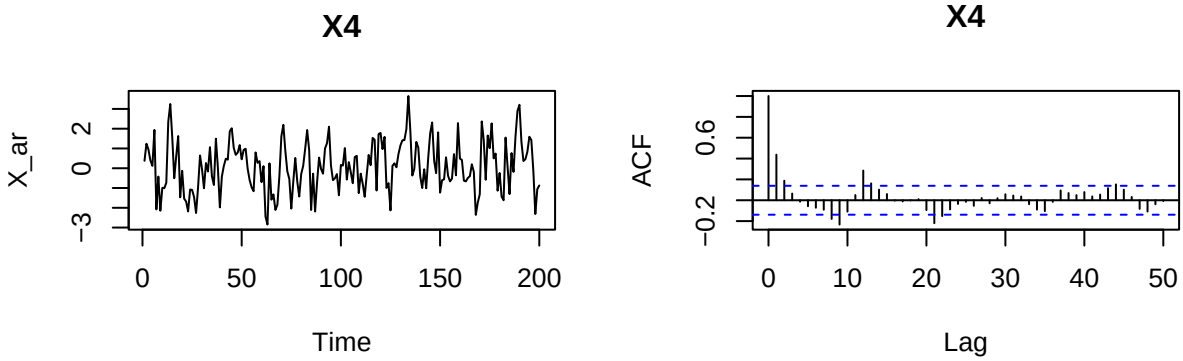
Exercice 5 /10

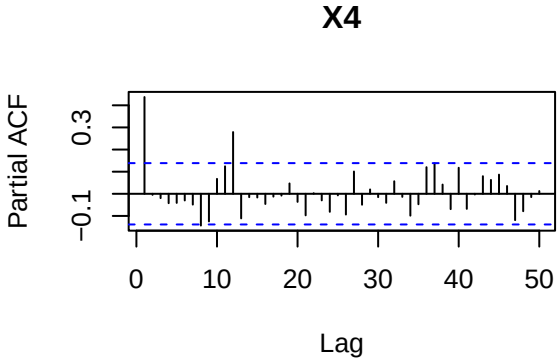
Un statisticien étudie un jeu de données composé de 4 séries temporelles pour lesquelles il a représenté/calculé les statistiques suivantes:











1. Proposer une démarche de modélisation pour chacune de ces séries, justifier. /4

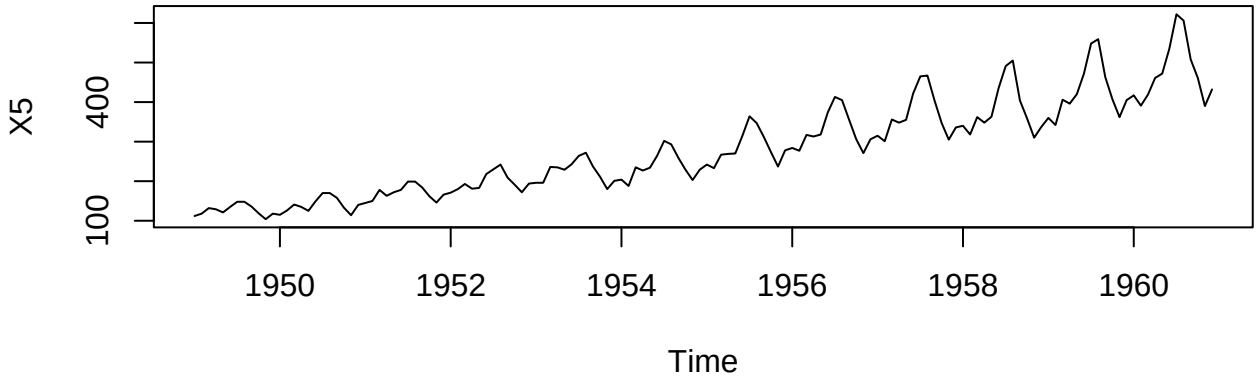
- Série 1: la série ne présente pas de comportement évident de type tendance ou saisonnalité. Sa variance semble également stable dans le temps. L'ACF s'annule à partir du lag 4 inclu, nous sommes en présence d'un MA(3).

- Série 2: au vu de la décroissance lente de l'ACF la série n'est pas stationnaire, cela est également visible sur la réprésentation graphique du processus qui indique la présence d'une tendance additive linéaire en deux parties: croissante pour t<100, décroissante pour t>100 ainsi qu'une saisonnalité de période 20. Il faut dans un premier temps corriger la série de sa tendance en utilisant une régression sur des bases de spline de degré 1 en introduisant un noeud en t=100. Sur la série corrigée de la tendance il faudra modéliser la partie saisonnière par exemple par régression sur série de fourier. Enfin, il faudra représenter' l'ACF et la PACF pour rechercher l'existance d'autocorrélation dans la série et

ainsi préciser l'ordre d'un éventuel modèle ARMA.

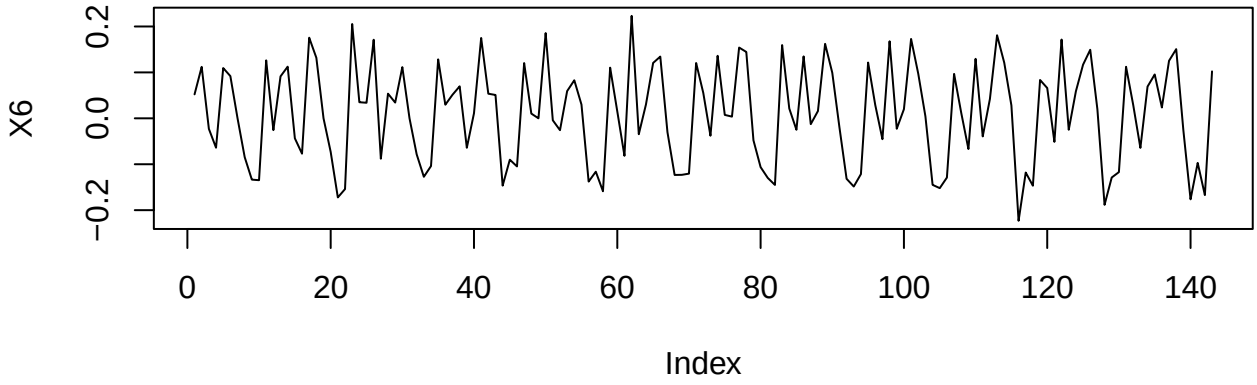
- Série 3: la série ne présente pas de comportement évident de type tendance ou saisonnalité. Sa variance semble également stable dans le temps. L'ACF décroit rapidement vers 0, on peut donc supposer que c'est un processus stationnaire. La PACF s'annule à partir du lag 4 inclu, nous sommes donc en présence d'un AR(3).
- Série 4: la série ne présente pas de non-stationarités évidentes (tendance ou cycles). L'ACF et la PACF montre une composante SARIMA associé à une périodicité d'ordre 12. Il n'est pas nécessaire de différencier la série car l'ACF décroit rapidement (modulo 12 et de base). Le PACF modulo 12 s'annule à partir de l'ordre 24 inclu. Le PACF s'annuleà partir du rang 1 inclu. Nous proposons un SAR(1,0,0) 12 (1,0,0).

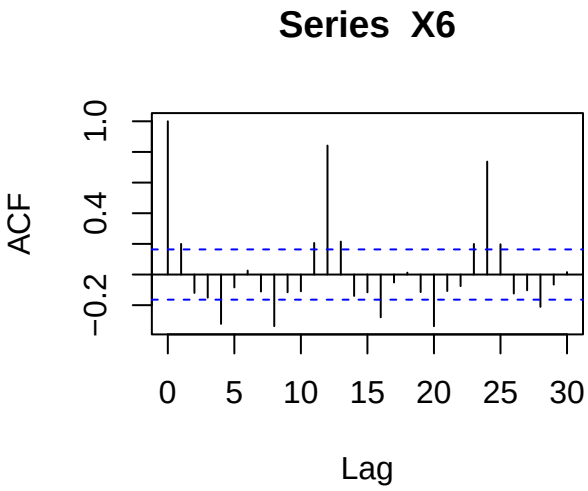
2. Voulant se frotter à de vraies données, notre statisticien s'intéresse ensuite à la série $X5$ représentée ci-dessous. Quelle première transformation est nécessaire pour stationnariser la série ? /0.5

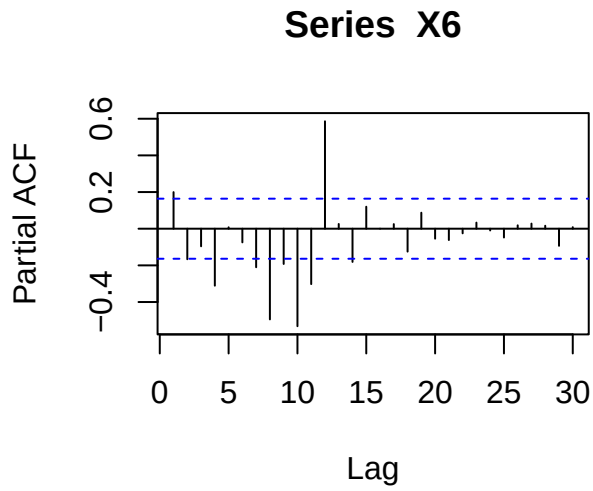


La première étape est la modélisation de la tendance de la série. On voit clairement qu'il existe une tendance croissante multiplicative. Il faut passer au log puis différencier la série ainsi obtenue.

3. Après avoir effectuer cette première étape, notre statisticien obtient la série $X6$ suivante dont il représente l'autocorrélogramme et l'autocorrélogramme partiel. Que pouvez vous dire sur X6? /1







On voit clairement apparaitre une saisonnalité annuelle (12 observations).

4. Notre statisticien propose le modèle suivant

```
## Series: X6
## ARIMA(0,0,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 = 0.001369:  log likelihood = 244.7
## AIC=-483.39   AICc=-483.2   BIC=-474.77
```

Quelles instructions a-t-il effectué pour obtenir ce modèle (code ou pseudo-code attendu)? /1

Il a tout d'abord transformé ses données en objet ts en indiquant la périodicité des données, via l'instruction

```
X6 <- ts(X6, frequency = 12)
```

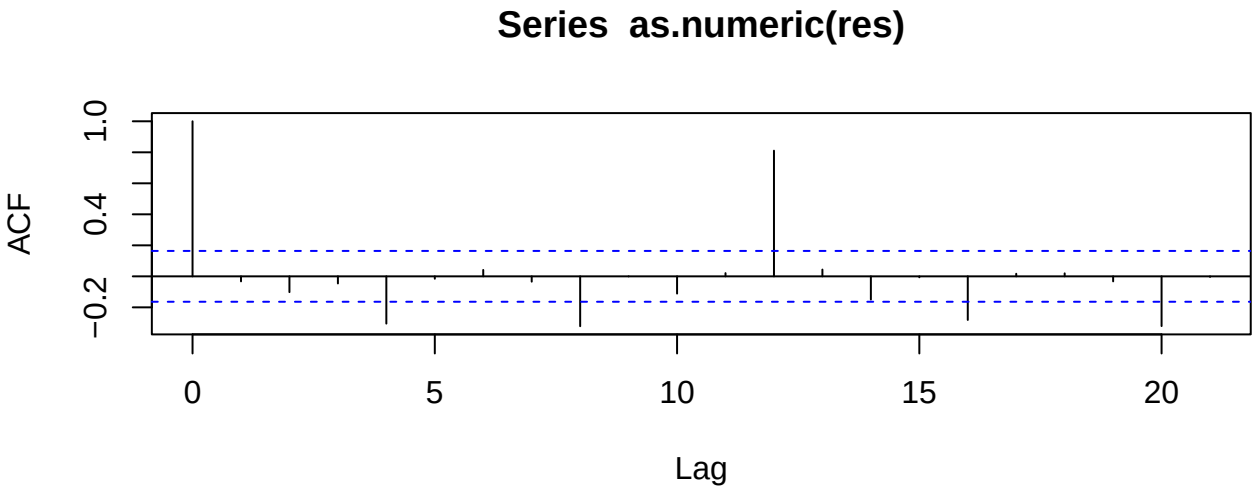
Après avoir identifé les ordres du modèle ARIMA il a lancé l'instruction:

```
arima(X6,order = c(0,0,1), seasonal = list(order = c(0,1,1)),include.mean = F)
```

Une autre possibilité, sans passer par la classe ts:

```
arima(X6,order = c(0,0,1), seasonal = list(order = c(0,1,1), period = 12),include.mean = F)
```

5. Avant d'obtenir ce modèle, notre statisticien avait représenté l'ACF des résidus d'un arima(0,0,1) estimé sur la série X6:



Etes-vous d'accord avec la modélisation proposée à la question 4.? Pourquoi? /1

On observe que l'acf s'annule après l'ordre $q = 1$, la PACF obtenue sur les résidus du modèle ARIMA(0,0,1) s'annulle presque dès le rang 1 d'ou le choix d'un MA(1). Pour la partie saisonnière, l'ACF (modulo 12) ne décroit pas suffisamment vite vers 0, d'ou le choix d'une différentiation d'ordre 1. Il faut effectuer cette différentiation puis représenté l'ACF pour choisir l'ordre de cette partie MA

6. Notre statisticien décide de regarder plus en détails son modèle. A quoi correspond la ligne "s.e"? Comment est calculée la p-value du test de student de nullité de chacun des coefficients? A quoi correspond AIC? /1

```
## Series: X6
## ARIMA(0,0,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 = 0.001369:  log likelihood = 244.7
## AIC=-483.39   AICc=-483.2   BIC=-474.77
```

```
## pvalue student-test:
```

```
##          ma1         sma1
## 7.380434e-06 2.575717e-14
```

La sortie s.e. correspond aux écarts types estimés des paramètres du modèle. Les p-values affichées correspondent au test de student de significativité des coefficients du modèle au seuil 0.05:

$$\frac{|\widehat{\phi}_p|}{\sigma_{\widehat{\phi}_p}} < 1.96$$

L'AIC est calculé ainsi: $\mathrm{AIC}(\phi, \theta, \sigma^2) = -2\log(L(\theta, \phi, \sigma^2)) + 2k$ ou L est la vraisemblance du modèle, k le nombre de paramètre du modèle ($p+q+1$ si la constante est inclue, $p+q$ sinon). Ce critère permet d'effectuer un choix de modèle parmi une famille de modèles ARMA, c'est un compromis entre un bon ajustement aux données et la dimension du modèle.

7. Au vu de ces résultats que doit conclure notre statisticien? /0.5

L'hypothèse H_0 du test de student de nullité des coefficients est rejetée au seuil 5%. Il peut donc conserver ce modèle.

8. Quels autres diagnostiques aurait-il pu effectué pour valider son modèle? /1

- test de box pierce pour vérifier la blancheur des résidus.
- test de shapiro wilk ou qq plot pour vérifier l'hypothèse Gaussienne.
- test de Dickey-Fuller augmenté pour la stationnarité.