Examen du 19/03/2021

MAP-STA2 : Séries chronologiques

Y. Goude - yannig.goude@edf.fr

Exercice 1 /4

Soit un processus $(X_t)_{t\in Z}$ moyenne mobile infinie tel que $X_t = \sum_{i=-\infty}^{+\infty} a_i \varepsilon_{t-i}$ ou ε_t est un bruit blanc de variance σ^2. La suite $(a_i)_{i\in Z}$ est une suite absolument sommable.

1. montrer que sa fonction d'autocovariance $\gamma(h)$ est bien définie et donner son expression. /1

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

   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:

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

2. donner la définition de sa densité spectrale, montrer qu'elle existe et que c'est une fonction réelle. /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:

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

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

3. montrer qu'il est équivalent de connaitre sa fonction d'autocovariance ou sa densité spectrale. /2

   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 $\sum_{h=1}^{+\infty}|\gamma(h)| = \sigma^2\sum_{h=1}^{+\infty}|\sum_{i=1}^{+\infty}a_i a_{i-h}| \leq \sigma^2\sum_{h=1}^{+\infty}\sum_{i=1}^{+\infty}|a_i||a_{i-h}| = \sigma^2(\sum_{i=-\infty}^{+\infty}|a_i)|)^2 < \infty$.

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 2 /4

Soit le processus suivant, supposé stationnaire, avec ε_t un bruit blanc de variance σ^2:

$$y_t = \sum_{k=1}^{p} a_k y_{t-k} + \varepsilon_t$$

1. Quel est le nom de ce processus? Donner la formule de sa fonction d'autocovariance $\gamma(h)$ et celle de sa fonction d'autocorrélation $\rho(h)$. /1

C'est un AR(p).

$$\gamma(h) = cov(y_t, y_{t-h})$$

$$\rho(h) = \gamma(h)/\gamma(0)$$

et $\rho(h)$ est définie comme la solution d'une suite récurrente d'ordre p, donc racines non-nulles de l'équation:

$$\rho(h) + \sum_{j=1}^{p}\phi_j\rho(h-j) = 0, \quad h = 1, 2...$$

soit:

$$1 + \sum_{j=1}^{p}\phi_j\lambda^{-j} = 0$$

2. Vérifiez si les deux processus suivants sont stationnaires puis calculer les autocorrélations et autocorrélations partielles d'ordre 1, 2, et 3 pour ces processus. /3

$$y_t = 0.7y_{t-1} - 0.49y_{t-2} + \varepsilon_t$$

$$y_t = 0.8y_{t-1} + \varepsilon_t$$

Il faut vérifier que les racines des polynômes caractéristiques sont de modules >1.

C'est le cas ici et:

Pour l'AR(2):

pour un AR(2) $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t$ et l'équation de récurrence vérifiée par la fonction d'autocorrelation $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$ nous permet d'obtenir: $\rho_1 = \phi_1 + \phi_2\rho_1$ et $\rho_2 = \phi_1\rho_1 + \phi_2$ soit $\rho_1 = \frac{\phi_1}{1-\phi_2}$ et $\rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$ et $\rho_3 = \phi_1\rho_2 + \phi_2\rho_1$.

Cela donne:

- $\gamma(1) = \sigma^2\frac{0.7}{1.49} = 0.47\sigma^2$, $\gamma(2) = \sigma^2(\frac{0.7^2}{1.49} - 0.49) = -0.17\sigma^2$, $\gamma(3) = 0.7\gamma_2 - 0.49\gamma_1 = -0.343\sigma^2$
- $r(1) = \frac{\gamma(1)}{\gamma(0)} = 0.47$, $r(2) = -0.49$, $r(3) = 0$

Pour l'AR(1):

- $\gamma(1) = (0.8\sigma^2)/(1-0.8^2) = 2.22\sigma^2$, $\gamma(2) = (\sigma^2 0.8^2)/(1-0.8^2) = 1.77\sigma^2$, $\gamma(3) = (\sigma^2 0.8^3)/(1-0.8^2) = 1.42\sigma^2$
- $r(1) = 0.8$, $r(2) = r(3) = 0$

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:

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

   $a^2\sigma_v^2 + (1+\phi_2^2)\sigma_u^2 > 2\phi_2\sigma_u^2$, 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

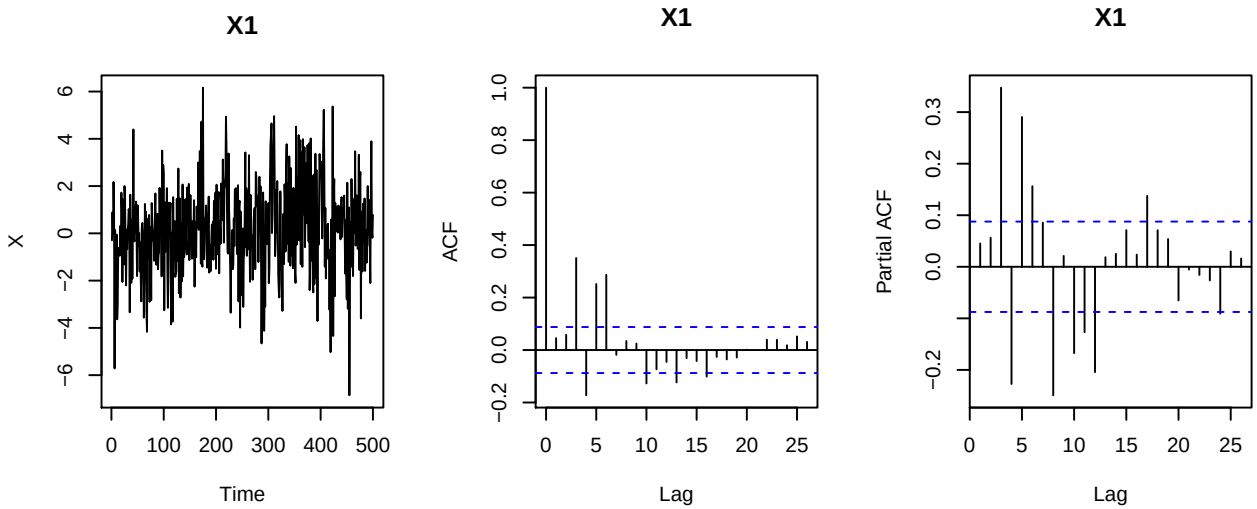
   On a $(1-\phi_1 L)(1-\phi_2 L)Y = (1-\lambda L)\varepsilon$, avce $\lambda \neq \phi_1$ et $\lambda \neq \phi_2$ c'est donc un ARMA(2,1).

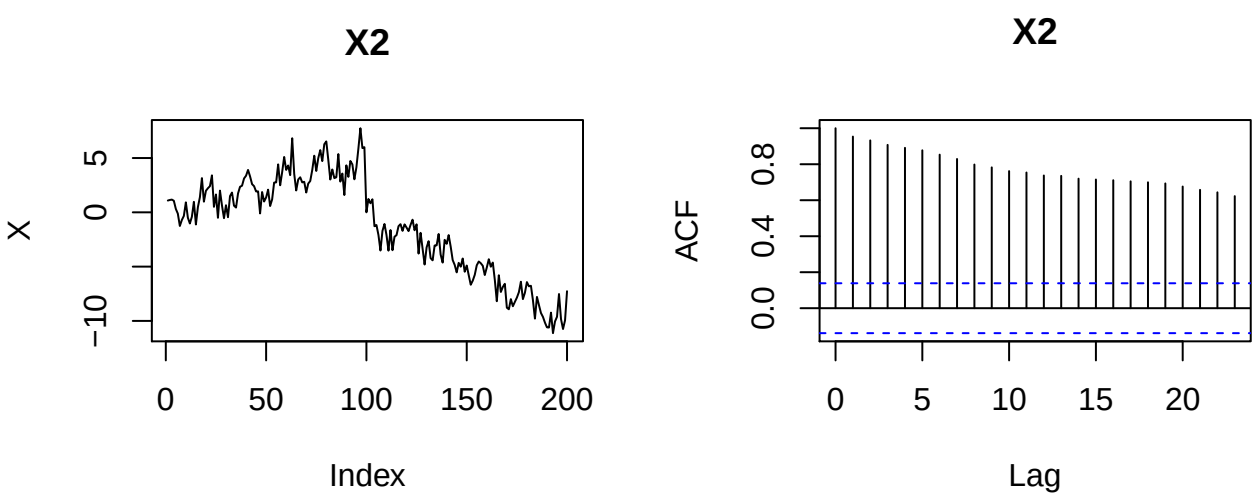
3. déterminer la prévision optimale à l'instant t notée Y_t^* et la variance de l'erreur de prévision $Var(Y_t - Y_t^*)$. /0.5

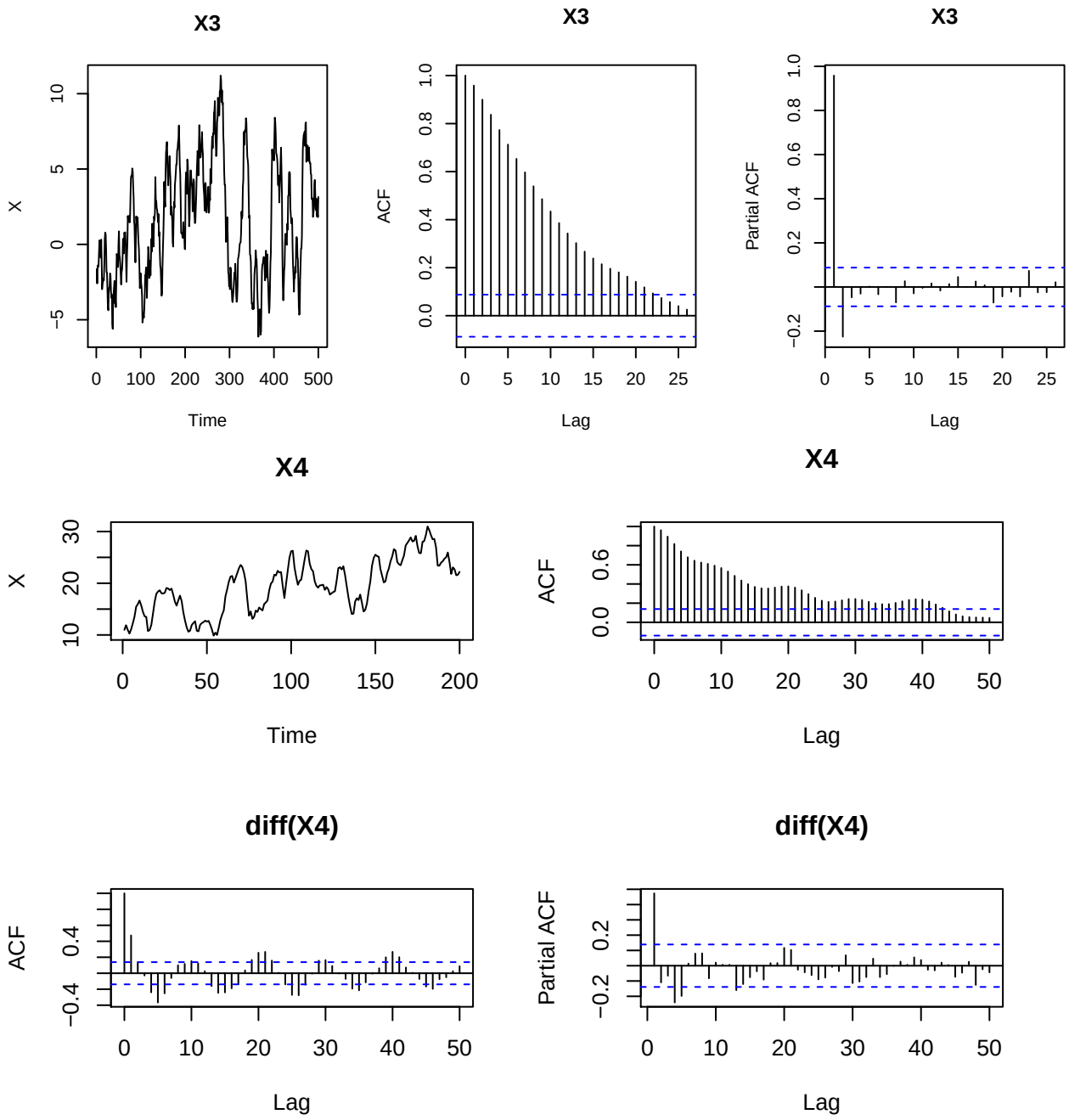
   Les racines de $\Phi(L) = (1-\phi_1 L)(1-\phi_2 L)$ sont de modules strictement supérieur à 1, on a donc un développement en moyenne mobile infinie $Y_t = \sum_{k>0} a_k \varepsilon_{t-k}$ et ε_t est l'innovation. D'ou $V(Y_t - Y_t^*) = \sigma_\varepsilon^2$ et $Y^* = Y - \varepsilon = (1 - \Phi(L)(1-\lambda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

Exercice 4 /10

Un statisticien étudie un jeux 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 (/1 pour chaque série)

- 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 7 inclu, nous sommes en présence d'un MA(6).

- 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. On distingue également une saisonnalité dans la série, également additive. 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. Ensuite, on peut modéliser la saisonnalité en ajustant une série de fourier (ou une régression sur base de spline périodique). La période peut être estimée via l'ACF ou un périodogramme. Le choix du

nombre d'harmoniques pourra se faire par une méthode de sélection de variable de type AIC ou selon des performances obtenues sur un échantillon test (sur la série corrigée de la tendance). Sur la série corrigée de la tendance et de la saisonnalité, il faudra représenter l'ACF et la PACF pour rechercher l'existance d'autocorrélation dans la série et ainsi préciser l'ordre d'un 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 3, nous sommes donc en présence d'un AR(2).
- Série 4: la série présente une tendance et une saisonnalité. Elle semble homoscédastique. L'acf décroit lentement vers 0 ce qui confirme la non-stationnarité. Il faut donc mettre en oeuvre une procédure de correction de la tendance et de la saisonnalité. Cela peut se faire par des m éthodes de régressions similaire à celles utilisées pour la série 2. On peut également différencier la série pour la corriger de sa tendance, on constate ici que l'acf de la série différenciée décroit plus rapidement vers 0 et comporte des oscillations caractéristiques d'une saisonnalité de période 20. On peut donc ensuite soit la différencier avec un opérateur de différenciation de lag 20 ou effectuer une régression sur série de Fourier (cf série 2). Sur la série corrigée de la tendance et de la saisonnalité, il faudra représenter l'ACF et la PACF pour rechercher l'existance d'autocorrélation dans la série et ainsi préciser l'ordre d'un modèle ARMA.

2. Notre statisticien s'intéresse ensuite à une autre série $X5_t$. Il propose de la modéliser par un ARMA et cherche ensuite à valider son modèle. Il obtient les sorties suivantes, à 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? /2

```
##
## Call:
## arima(x = X5, order = c(3, 0, 4), include.mean = T, method = c("ML"))
##
## Coefficients:
##           ar1     ar2     ar3     ma1     ma2     ma3     ma4  intercept
##       -0.3245  0.2876  0.0359  0.9251  0.0782  0.2621  0.2739    -0.0181
## s.e.   0.2333  0.1918  0.2394  0.2267  0.1939  0.2788  0.1476     0.3326
##
## sigma^2 estimated as 5.203:  log likelihood = -673.49,  aic = 1364.99
```

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

```
##       ar1       ar2       ar3       ma1       ma2       ma3       ma4 intercept
##      0.16      0.13      0.88      0.00      0.69      0.35      0.06      0.96
```

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: $\text{AIC}(\phi, \theta, \sigma^2) = -2\log(L(\theta, \phi, \sigma^2)) + 2k$ 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.

3. Au vu de ces résultats que doit faire notre statisticien? /2

L'hypothèse H_0 du test de student de nullité des coefficients AR(3) n'est pas rejetée au seuil 5%, idem pour la constante. Le statisticien doit donc proposer un modèle ARMA(2,4) sans la constante (option: include.mean = F) et réeffectuer les tests.

4. Il s'intéresse ensuite à une série $X6_t$ et effectue un test de Box-Pierce d'ordre 10 sur $X6_t$. Il obtient le résultat suivant:

```
##  lags statistic df    p-value
##    20  29.13311 20 0.08517126
```

à quoi correspond "df"? Quelle est la conclusion de ce test? /2

Le test de Box-Pierce permet de tester l'hypothèse que les résidus d'une série X_t suivant une modélisation ARMA(p,q) sont un bruit blanc ie, pour une série X_t et ses résidus associés $\widehat{\varepsilon}_t = \widehat{\Theta}(L)^{-1}\widehat{\Phi}(L)(1-B)^d X_t$ de fonction d'autocorrélation $\rho_\varepsilon(h)$ et son estimateur empirique asssocié:

$$H_0(h) : \rho_\varepsilon(1) = \rho_\varepsilon(2) = ... = \rho_\varepsilon(h) = 0$$

Il se base sur la statistique de Box-Pierce:

$$Q_{BP}(h) = n \sum_{j=1}^{h} \widehat{\rho}_\varepsilon(j)^2 \xrightarrow[n \to +\infty]{\mathcal{L}} \chi^2(h-k)$$

lags correspond donc à h l'ordre du test, et df le nombre de degrés de libertés de la χ^2 associée.

Au seuil 5% l'hypothèse H_0 n'est pas rejetée, il n'y a pas d'autocorrélation significative des résidus jusqu'à l'ordre 20.