Examen du 23/01/2018

MAP-STA2 : Séries chronologiques

Yannig Goude - yannig.goude@edf.fr; M. Zaouche - mounia.zaouche@gmail.com;

Barem

exercice1 sur 6 points répartis en 1.5 pour les questions 2) et 3), 1 point pour les autres.

exercice2 sur 2 points répartis en 0.5 point par question

exercice3 sur 4 points répartis en 1 point par question

exercice4 sur 8 points répartis en 3 points ppour la question 1) et 1 point pour les autres

Exercice 1

1. Soit X_t un processus stationnaire du second ordre, ε_t suite de v.a. iid centrée de variance σ^2. On fait l'hypothèse H_a que ce processus est un AR(1):

$$X_t = aX_{t-1} + \varepsilon_t$$

avec $|a| < 1$.

1. calculer l'autocorrélation d'ordre 1 : $\rho(1)$.

on a: $\gamma(0) = \sigma^2/(1 - a^2)$ *et* $\gamma(k) = a^k \gamma(0)$

2. supposons qu'on observe une trajectoire de taille n de ce processus. Définir l'estimateur empirique a_n de a. Quelle est la limite de a_n quand n tend vers l'infini?

$\widehat{a}_n(1) = \rho_n$ *l'autocorrélation empirique. On sait qu'il est consistant:* $\lim_{n\to\infty} \rho_n = \rho(1) = a$ *presque surement.*

3. suggérer un test pour valider l'hypothèse H_a.

Si le modèle est correct, $X_n - aX_{n-1} = \varepsilon_n$ *est un bruit blanc.On peut ainsi tester la validité du modèle en calculant les autocorrélations empiriques* $\rho_n^\varepsilon(k)$ *des résidus* $\varepsilon_i = X_i - a_n X_{i-1}$*. Pour une valeure raisonnable et fixée de* q*, on pose* $T_n = \rho_n^\varepsilon(1) + ... + \rho_n^\varepsilon(q)$ *. Sous* H_a*,* Tn *suit approximativement une* χ^2 *à* q *degrés de libertés. La region de rejet du test est* $T_n > t$ *ou* t *est un quantile approprié (selon le niveau du test) de la loi du* χ^2*.*

4. le processus est en fait un processus AR(2):

$$X_t = \alpha X_{t-1} + \beta X_{t-2} + \varepsilon_t$$

montrer que pour que ce processus puisse se décomposer en une somme de valeur passées de ε_t il est nécessaire d'avoir $|\beta| < 1$.

Pour cela le polynôme caractéristique $P(z) = 1 - \alpha z - \beta z^2$ *doit avoir ses racines en dehors du cercle unité. Le produit des racines vaut* $-1/\beta$ *on doit donc avoir* $|1/\beta| < 1$*.*

5. écrire les équations de Yule-Walker reliant $\rho(1)$, $\rho(2)$, α et β. Exprimer $\rho(1)$ et $\rho(2)$ en fonction de α et β.

$$\rho(1) = \alpha + \beta\rho(1)$$

$$\rho(2) = \alpha\rho(1) + \beta$$

et donc:

$$\rho(1) = \alpha/(1-\beta)$$

$$\rho(2) = \alpha^2/(1-\beta) + \beta$$

Exercice 2

Etudiez la stationnarité des processus suivants (ε_t est un bruit blanc de variance σ^2):

1. $y_t = at + \varepsilon_t$

non-stationnaire, en effet $E(y_t) = at$ dépend du temps

2. $x_t = \Delta y_t$

$\Delta y_t = a + \varepsilon_t - \varepsilon_{t-1}$ *processus MA(1) stationnaire*

3. $z_t = at^2 + bt + \varepsilon_t$

non-stationnaire, en effet $E(y_t) = at^2 + bt$ dépend du temps

4. $z_t = \Delta^2 yt$

stationnaire, en effet $z_t = \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}$ rocessus MA(1) stationnaire

Exercice 3

Soit Z_t une suite de variables aléatoires iid de loi normale de moyenne 0 et de variance σ^2. a, b et c des constantes. Déterminer lequel ou lesquels des processus ci-dessous sont stationnaires. Pour chaque processus stationnaire, calculer la moyenne et la fonction d'autocovariance.

1. $X_t = a + Z_t + cZ_{t-2}$

stationnaire, $E(X_t) = a$, $V(X_t) = \sigma^2 + c^2\sigma^2$, $\gamma(1) = 0$, $\gamma(2) = c\sigma^2$, $\gamma(k) = 0$ pour $k > 2$

2. $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$

stationnaire, $E(X_t) = 0$, $V(X_t) = \sigma^2$, comme \$ cos(u+v) = cosucosv-sinusinv\$ on a $\gamma(k) = \sigma^2 cos(kc)$

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

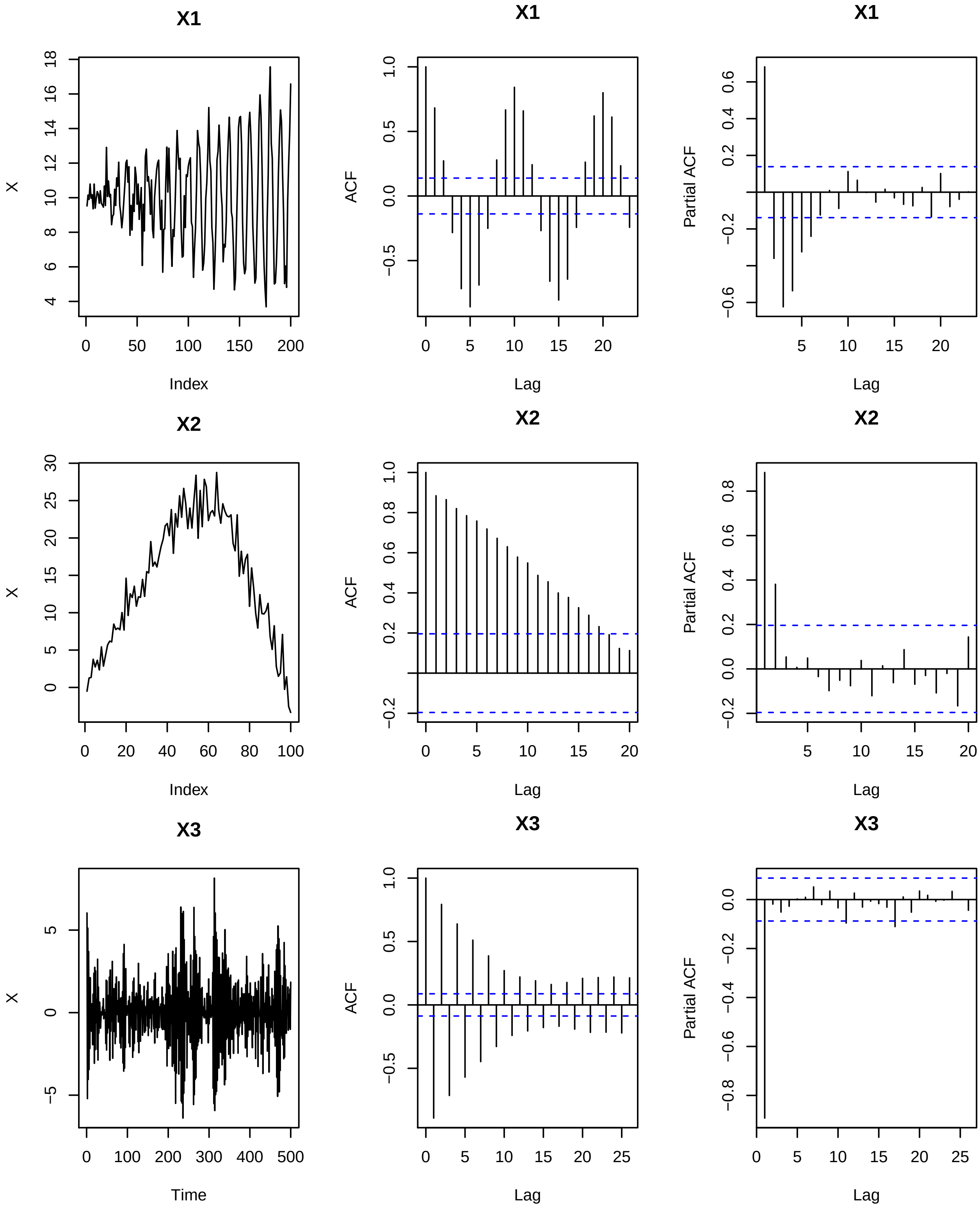
non stationnaire, $E(X_t) = 0$, $V(X_t) = \sigma^2$ mais $\gamma(k) = \sigma^2 \sin(ct)\cos(c(t-k))$ dépend de t

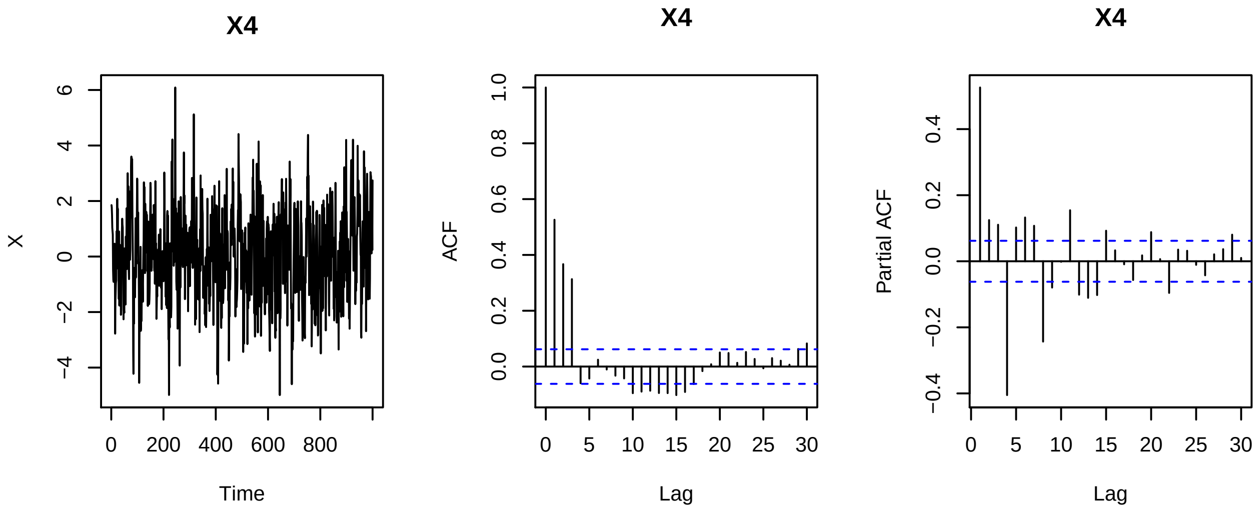
4. $X_t = Z_t Z_{t-1}$

stationnaire, $E(X_t) = 0$, $V(X_t) = \sigma^4$, $\gamma(k) = 0$ pour k différent de 0.

Exercice 4

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.

- *$X1$ la trajectoire du processus fait apparaitre une saisonnalité, confirmée par l'ACF, de période 10. D'autre part, on constate que l'amplitude de ces oscillations périodiques croit avec le temps ce qui plaide en faveur d'une tendance multiplicative. On peut donc proposer le modèle de régression suivant: $X1_t = atcos(2\pi/10) + eps$*
- *$X2$ la trajectoire du processus fait apparaitre une tendance polynomiale par morceaux, cette non-stationnarité est confirmée par l'ACF qui ne décroit pas exponentiellement vers 0. Une démarche possible est d'estimé cette tendance par régression spline ou estimateur à noyau. Ensuite il faudra étudier les résidus: ACF, PACF...*
- *$X3$ l'acf décroit exponentiellement vers 0 et la pacf montre qu'il s'agit d'un AR(1)*
- *$X4$ l'acf décroit brutalement après $q = 3$ ce qui est caractéristique d'un MA(3)*

Notre statisticien s'intéresse ensuite à une autre série X_t^5. Il propose de la modéliser par un ARMA et cherche ensuite à valider son modèle. Il obtient les sorties suivantes:

```
##
## Call:
## arima(x = X5, order = c(3, 0, 4), include.mean = T, method = c("ML"))
##
## Coefficients:
##           ar1     ar2     ar3     ma1      ma2     ma3     ma4  intercept
##       -0.2995  0.4580  0.2501  0.9064  -0.0054  0.0914  0.2061     0.2204
## s.e.   0.2213  0.1404  0.1866  0.2183   0.1672  0.1907  0.1047     0.4837
##
## sigma^2 estimated as 5.183:  log likelihood = -673.01,  aic = 1364.02

## pvalue student-test:

##       ar1       ar2       ar3       ma1       ma2       ma3       ma4
##      0.18      0.00      0.18      0.00      0.97      0.63      0.05
## intercept
##      0.65
```

2. Expliquer comment sont estimés les coefficients du modèle ARMA?

par les équations de yule-walker ou moindre carrés conditionnels, cf cours.

3. Expliquer ce que signifie la sortie "aic=1364.02".

l'aic est calculé ainsi: $AIC(\phi,\theta,\sigma^2) = -2\log(L(\theta,\phi,\sigma^2)) + 2k$ ou L est la vraisemblance du modèle. Ce critère permet d'effectuer un choix de modème parmi une famille de modèles ARMA, c'est un compromis entre un bon ajustement aux données et la dimension du modèle.

4. A quoi correspond la ligne "s.e"? Comment est calculée la p-value du test de student de nullité de chacun des coefficients?

s.e. correspond à l'écart type estimé des coefficients du modèle, diminuer d'une unité l'ordre de l'AR(p) ou du MA(q) revient à tester la significativité du coefficient ϕ_p (resp. θ_q) ce qui peut être fait par un test de student, les estimateurs $\widehat{\phi}_p$ et $\widehat{\theta}_q$ obtenus par maximum de vraisemblance ayant les mêmes propriétés qu'en régression linéaire.

5. Au vu de ces résultats que doit faire notre statisticien?

la constante et la coefficient de l'AR(3) ne sont pas significativement non nuls. Il faut réduire l'ordre de l'AR de 1 et ne pas inclure la constante dans le modèle.

6. Il s'intéresse ensuite à une série X_t^6 et effectue un test de Box-Pierce sur $X6_t$. Il obtient le résultat suivant:

```
##  Lags Statistic df     pvalue
##    20  29.13311 20 0.08517126
```

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

ce test 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:

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

$$H_1(h) : \exists k \in (1,...,h) \text{ t.q } \rho_\varepsilon(k) \neq 0$$

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 degré&s de libertés de la χ^2 associée