Examen du 14/03/2017

MAP-STA2 : Séries chronologiques

Yannig Goude - yannig.goude@edf.fr; Jade Giguelay jgiguela@ens-paris-saclay.fr

Consignes L'examen est constitué de 2 parties. La première correspond à des exercices à rendre sur copie, la deuxième comporte des exercices d'application pour lesquels un code r est demandé en plus d'une copie. Ce code r est à envoyer par email à l'adresse yannig.goude@edf.fr ou jgiguela@ens-paris-saclay.fr selon les groupes, avant la fin de l'examen.

Les documents ne sont autorisés que pour la partie 2.

Respecter la nomenclature suivante SVP pour le nom de fichier du code: code_nom.R

Code le code doit être lisible, commenté et écris de manière à pouvoir être exécuté par le correcteur.

Partie 1: sur copie uniquement

Exercice 1

1. Donner la définition d'un processus stationnaire fort et faible.
2. Donner 3 exemples de processus stationnaires.
3. Donner 3 exemples de processus non-stationnaires.
4. Enoncer le théorême de Wold. En quoi est il important pour la modélisation de séries chronologiques?

Exercice 2

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 $\rho(h)$.
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.

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

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

Exercice 3

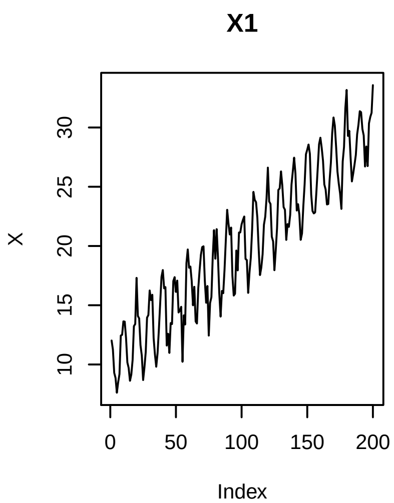
1. Donner la définition de la densité spectrale d'un processus X_t
2. Soit ε un bruit blanc et X un processus moyenne mobile d'ordre 1 défini, pour $\theta \in]-1, 1[$, par:

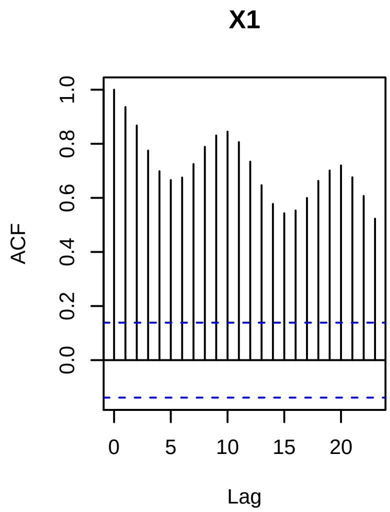
$$X_t = \varepsilon_t - \theta\varepsilon_{t-1}$$

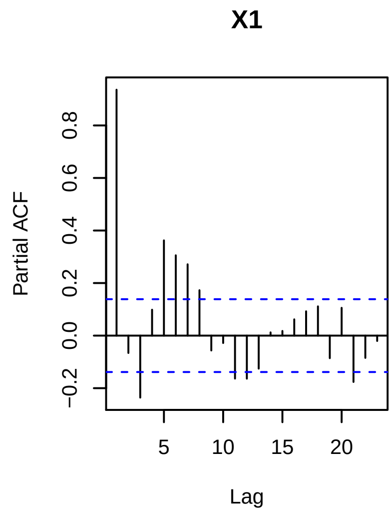
calculer sa densité spectrale.

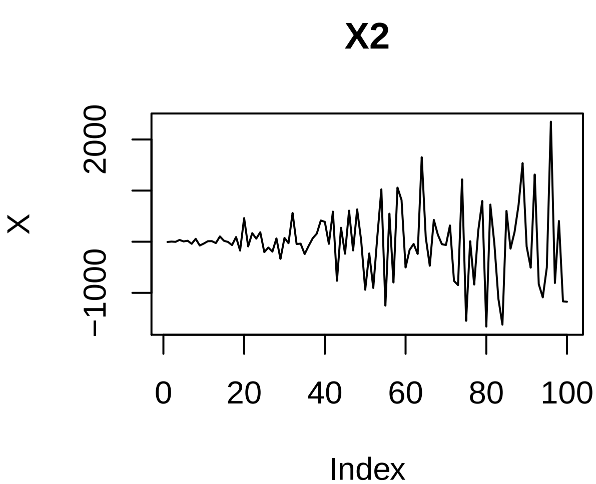
Exercice 4

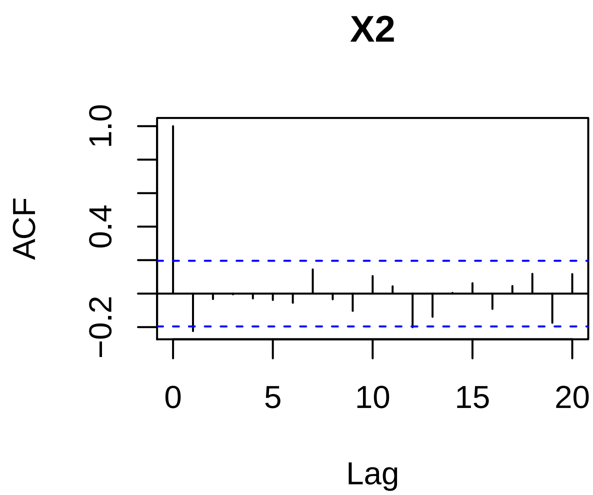
Un statisticien étudie un jeux de données composé de 4 séries temporelles pour lesquelles il a représenté/calculé les statistiques suivantes:

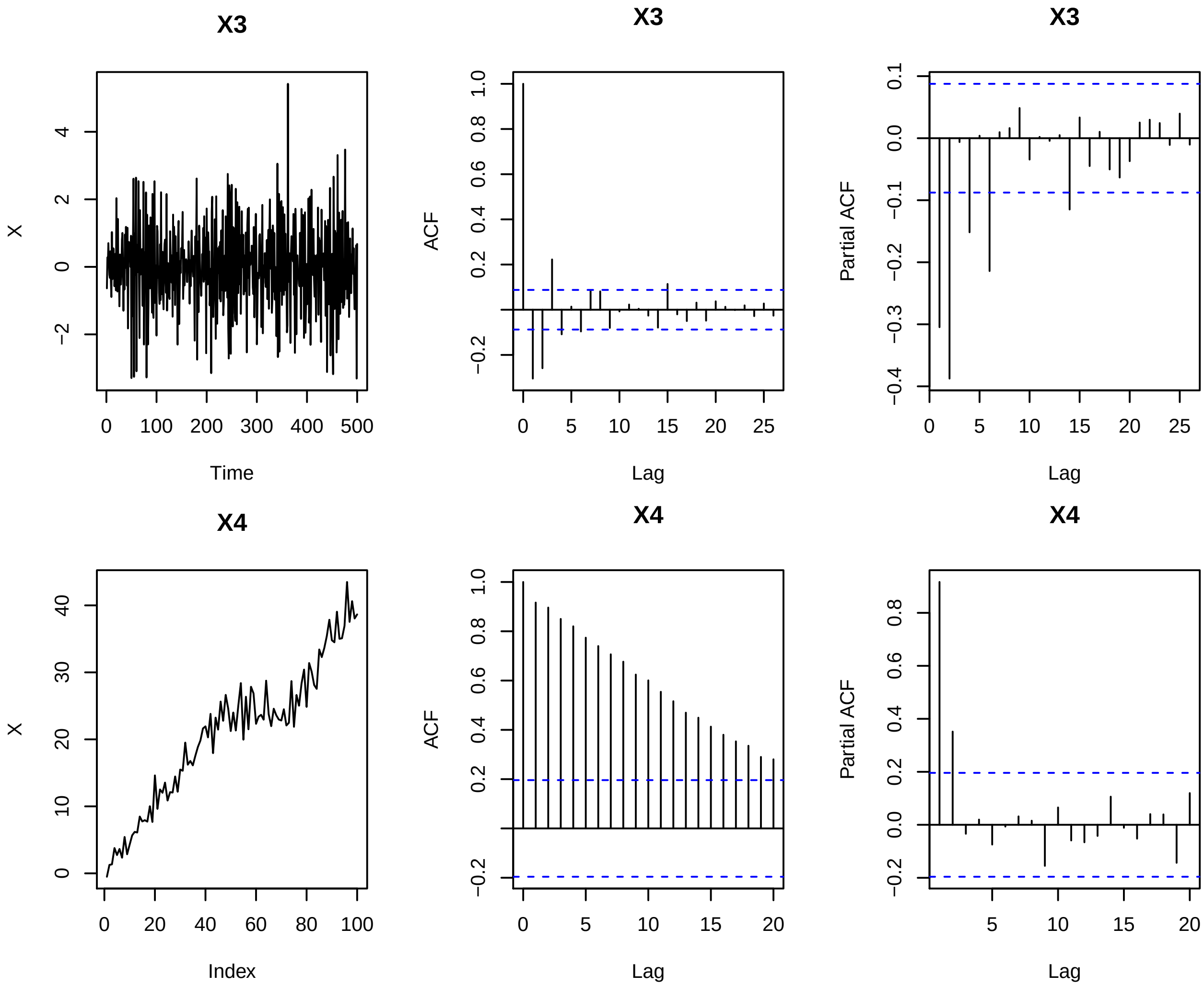












1. Quelles séries ne sont pas stationnaires (justifier)? Pour celles qui ne sont pas stationnaires proposer une méthode pour se ramener à un processus stationnaire.

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:

```
##
## 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
##      0.16      0.13      0.88      0.00      0.69      0.35      0.06
## intercept
##      0.96
```

2. A quoi correspond la ligne "s.e"? Comment est calculée la p-value du test de student de nullité de chacun des coefficients?
3. Au vu de ces résultats que doit faire notre statisticien?
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     pvalue
##    20  29.13311 20 0.08517126
```

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

Partie 2: code R

Exercice

1. Importer les données "data_conso.txt" et représenter la série "Load"" en fonction du temps (construire la variable date adéquate) ainsi que son autocorrélogramme. Cette série est elle stationnaire?
2. Séparer les données en deux jeux de données: les 600 premières observations (table data0) pour l'apprentissage de nos modèles, les 79 dernières (table data1) pour effectuer une prévision.
3. Proposer une modélisation de la tendance de la série. Construire une série Load_detrend corrigée de la tendance.
4. Calculer le spectrogramme de la série Load_detrend puis proposer une modélisation de la composante saisonnière à l'aide d'une méthode de régression sur des fonctions trigonométriques bien choisie, justifier votre choix et valider votre modèle par une analyse des résidus.
5. Par la méthode de Box-Jenkins, proposer un modèle SARIMA sur les résidus du modèle de régression choisi (pour éviter des problèmes de convergence, utiliser l'argument method="CSS" dans la fonction arima)
6. Les résidus de ce modèle SARIMA sont ils bien un bruit blanc gaussien?
7. Effectuer un test de student de significativité des coefficients du SARIMA obtenu, que dire?
8. Effectuer la prévision de la série Load de la table Data1. Representer sur un même graphique vos prévisions et les données associées. Calculer l'erreur quadratique moyenne de prévision (MSE): $\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$.
9. Reprendre la même démarche sans utliser de régression sur des fonctions trigonométriques mais avec un unique modèle SARIMA. Comparer les erreurs de prévision.