

Examen du 14/03/2016

MAP-STA2 : Séries chronologiques

Yannig Goude - yannig.goude@edf.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 avant la fin de l'examen.

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

Code le code doit être lisible et écrit de manière à pouvoir être exécuté par le correcteur. Veillez à ne pas faire de boucle.

Partie 1: sur copie uniquement

Exercice 1

Soit le modèle suivant:

Dans cet exercice, ε_t est un bruit blanc Gaussien de variance σ^2 .

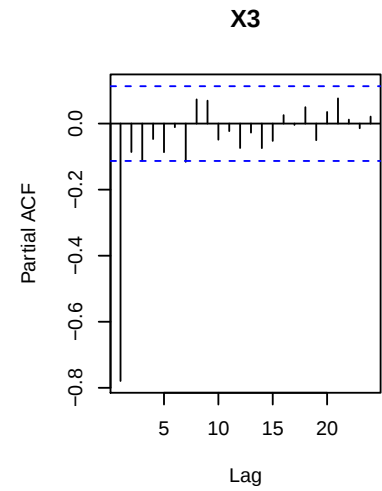
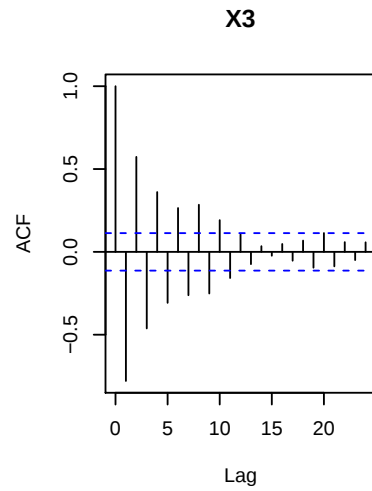
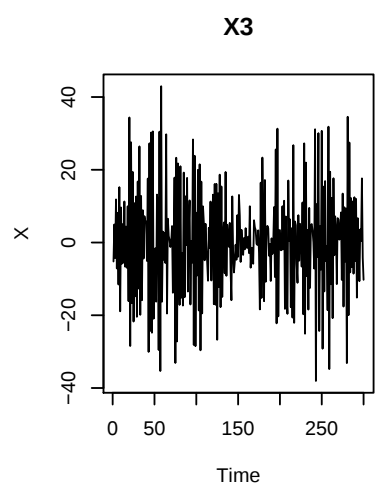
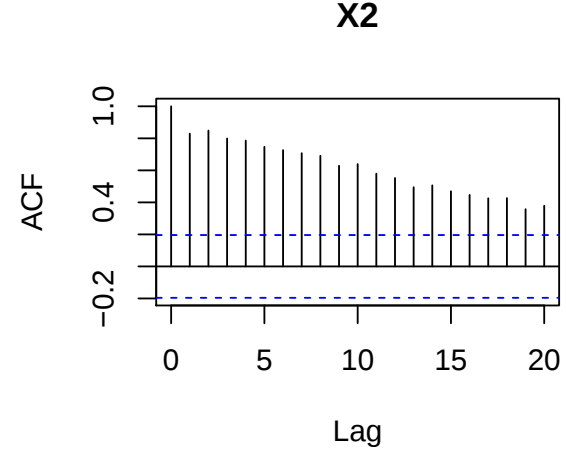
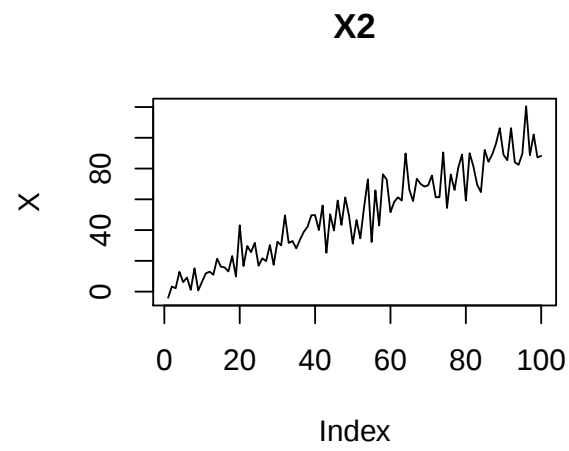
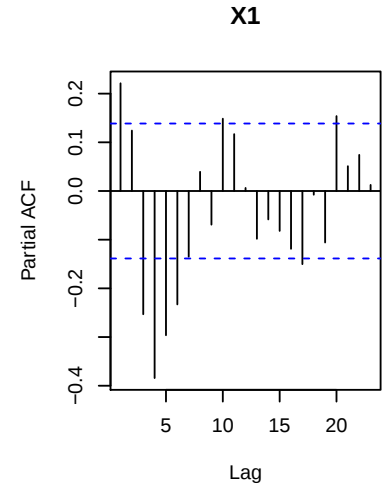
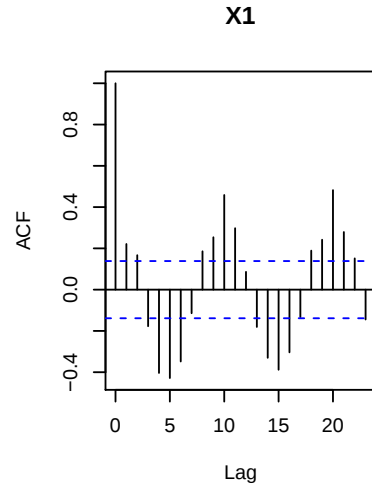
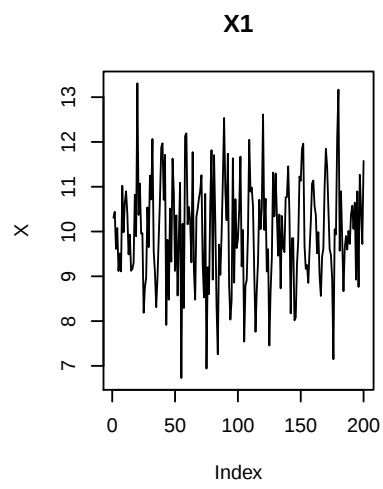
1. soit un processus $y_t = \varepsilon_t - 0.3\varepsilon_{t-1} + 0.6\varepsilon_{t-2}$. Quel est le nom de ce processus? Calculer sa fonction d'autocovariance $\gamma(h)$.
2. soit un processus $y_t = \frac{5}{6}y_{t-1} - \frac{3}{4}y_{t-2} + \varepsilon_t$. Quel est le nom de ce processus? Calculer sa fonction d'autocovariance $\gamma(h)$ et sa fonction d'autocorrélation partielle $r(h)$.

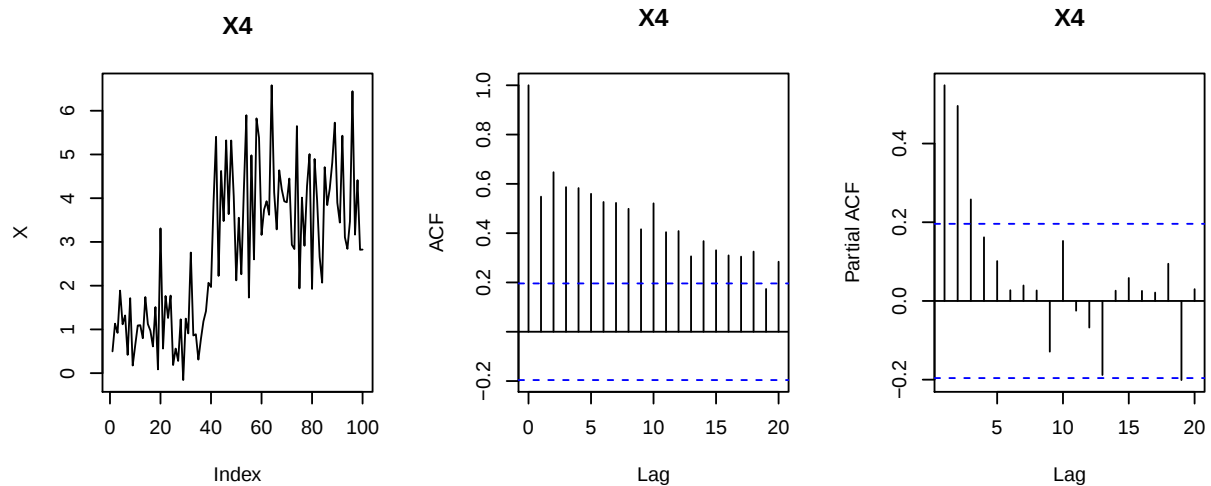
Exercice 2

1. soit un processus $X_t = A \cos(\omega t) + B \sin(\omega t)$ $t \in \mathbf{Z}$ où A et B sont des variables aléatoires indépendantes et centrées de même variance σ^2 . Montrer que X_t est stationnaire au second ordre. Calculer sa fonction d'autocovariance.
2. soit $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, calculer sa densité spectrale.
3. soit $X_t = \phi X_{t-1} + \varepsilon_t$, calculer sa densité spectrale.

Exercice 3

Un statisticien étudie un jeu 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
##      -0.3292  0.5595  0.0742 -1.0445  0.4782  0.5128 -0.5794   0.0151
## s.e.   0.2093  0.2179  0.1148  0.1984  0.2752  0.2747  0.1888   0.0568
##
## sigma^2 estimated as 3.45:  log likelihood = -616.98,  aic = 1251.96

## pvalue student-test:

##          ar1      ar2      ar3      ma1      ma2      ma3      ma4
##          0.12      0.01      0.52      0.00      0.08      0.06      0.00
## intercept
##          0.79
```

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

1. Simuler un processus ARMA(2,2) stationnaire dont l'innovation est un bruit blanc Gaussien de variance σ^2 .
2. Calculer, sans utiliser les fonction acf et pacf de R, les autocorrélations partiels de ce processus.
3. Proposer une méthode pour simuler un processus SARIMA[(0,0,2):(0,0,1)] dont l'innovation est un bruit blanc Gaussien de variance σ^2 .

Exercice 2

1. Importer les données "Nottingham_castle.txt" et représenter la série Temperature en fonction du temps ainsi que son autocorrélogramme. Cette série est elle stationnaire?
2. Séparer les données en deux jeux de données: les 220 premières observations (table data0) pour l'apprentissage de nos modèles, les 20 dernières (table data1) pour effectuer une prévision. Modéliser 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.
3. Par la méthode de Box-Jenkins, proposer un modèle SARIMA sur les résidus du modèle de régression choisi.
4. Les résidus de ce modèle SARIMA sont ils bien un bruit blanc gaussien?
5. Effectuer un test de student de significativité des coefficients du SARIMA obtenu, que dire?
6. Effectuer la prévision de la température de Data1. Représenter sur un même graphique vos prévisions et les données associées. Calculer l'erreur quadratique moyenne de prévision (MSE): $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$.
7. Reprendre la même démarche sans utiliser de régression sur des fonctions trigonométriques mais avec un unique modèle SARIMA. Comparer les erreurs de prévision.