Examen du 15/01/2019

MAP-STA2 : Séries chronologiques

Yannig Goude - yannig.goude@edf.fr; Ajmal Loodally - ajmalloodally@hotmail.com;

Exercice 1

Exercice 1

1. énoncer le théorême de Wold. En quoi est il important pour la modélisation de séries chronologiques?
2. donner 3 exemples de processus stationnaires et justifier.
3. donner 3 exemples de processus non-stationnaires et justifier.
4. qu'est ce que la fonction l'autocorrélation d'un processus? Proposer un estimateur empirique et un exemple de code R réalisant cette estimation.
5. qu'est ce que la fonction l'autocorrélation partiel d'un processus? Proposer un estimateur empirique et exemple de code R réalisant cette estimation.

Exercice 2

Soit U_n un bruit blanc centré de variance σ^2 et soit Y_n le processus défini par

$$Y_n = U_n + 2U_{n-1}, n \in Z$$

1. déterminer la densité spectrale de Y_n.
2. déterminer la fonction d'autocovariance de Y_n.
3. calculer la variance τ^2 de l'erreur de prédiction à horizon 1 conditionnellement au passé.
4. on pose $V_n = \sum_{j=0}^{\infty}(-2)^{-j}Y_{n-j}$. Exprimer V_n en fonction de $U_k, k \leq n$.
5. calculer la variance et la fonction d'autocovariance de V_n.
6. exprimer Y_n en fonction de V_n et V_{n-1}.
7. soit $\widehat{Y}_{n+1}$ la prédiction de Y_{n+1} conditionnellement au passé $Y_k, k \leq n$. Donner l'erreur de prédiction et le risque quadratique moyen de prévision à un pas.

Exercice 3

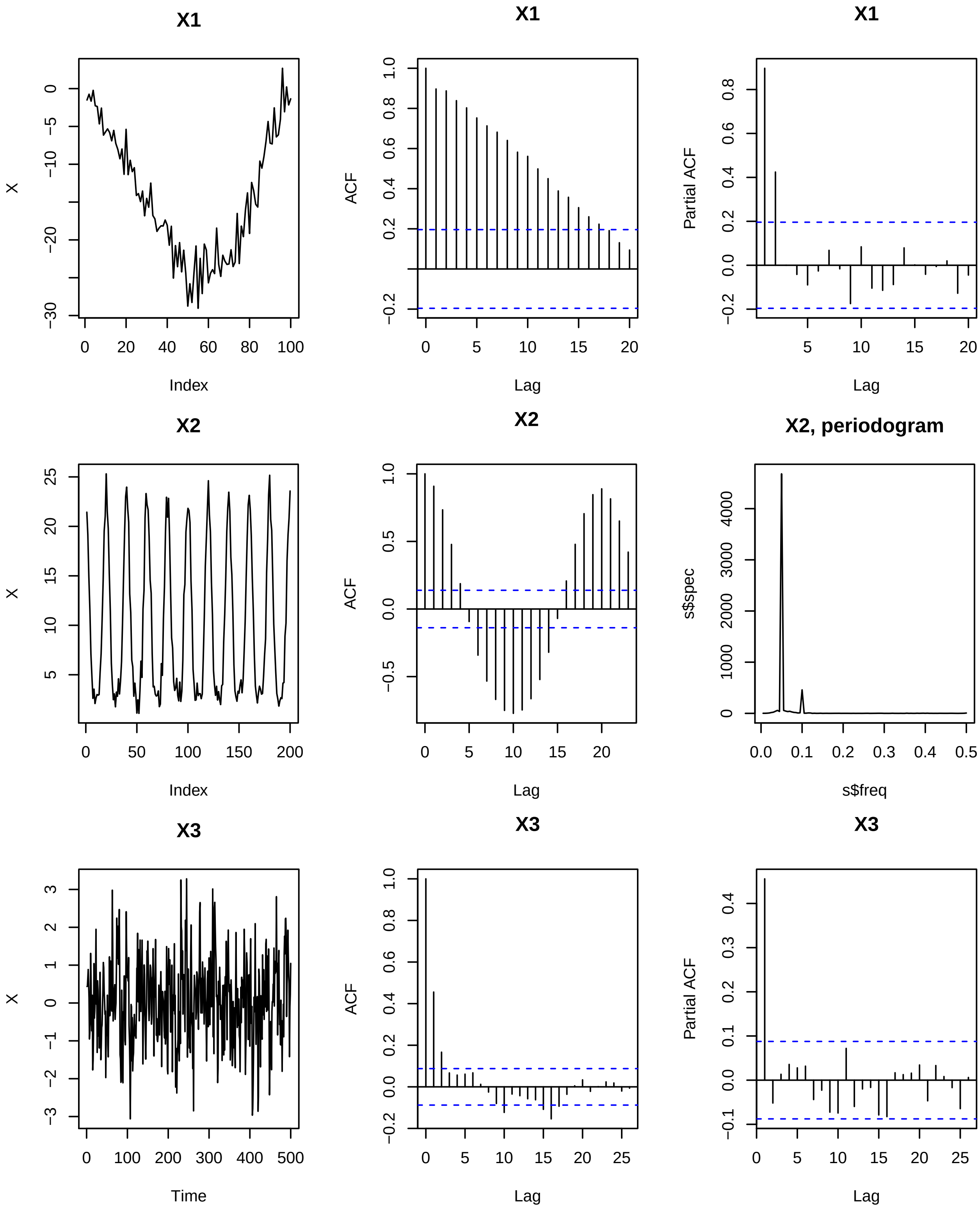
Soit le processus y_t suivant, supposé stationnaire, avec ε_t un bruit blanc de variance σ^2:

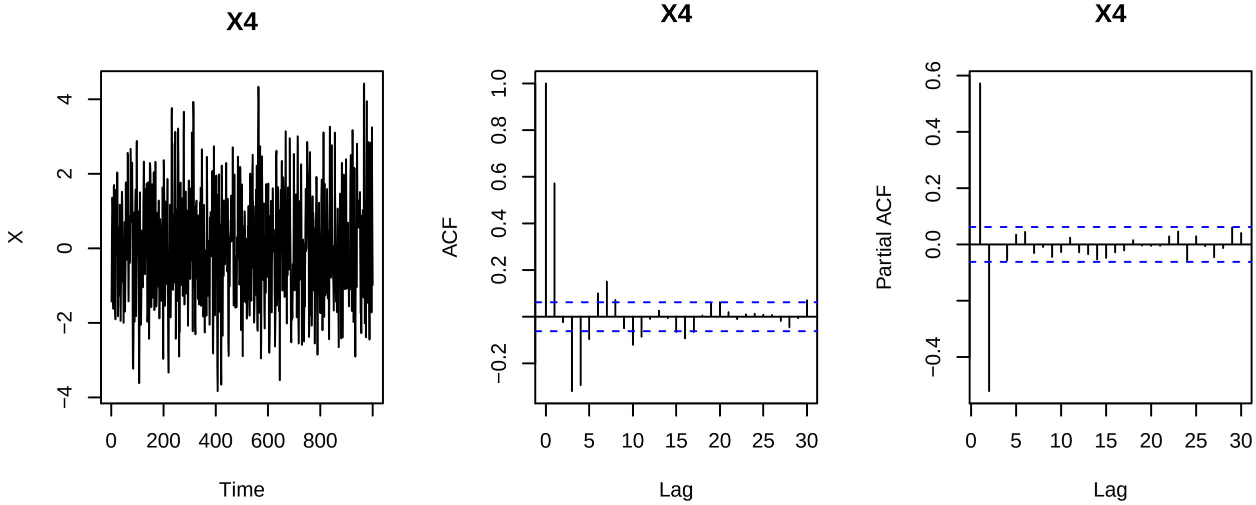
$$y_t = \sum_{k=1}^{p} a_k y_{t-k} + \varepsilon_t$$

1. quel est le nom de ce processus?
2. à quelle(s) condition(s) ce processus est il stationnaire?
3. on fait l'hypothèse que $p = 1$. A quelle(s) condition(s) ce processus est il stationnaire?
4. dans ce cas, exprimer l'écriture moyenne mobile infinie de y_t.

Exercice 4

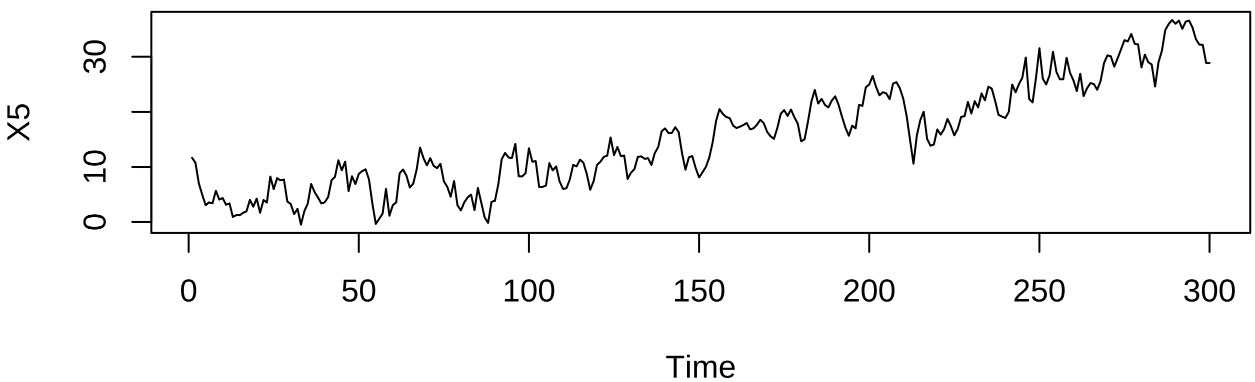
Un statisticien étudie un jeux de données composé de 4 séries temporelles pour lesquelles il a représenté/calculé les statistiques suivantes:


X1
X
Index
X1
ACF
Lag
X1
Partial ACF
Lag
X2
X
Index
X2
ACF
Lag
X2, periodogram
s$spec
s$freq
X3
X
Time
X3
ACF
Lag
X3
Partial ACF
Lag



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

Notre statisticien s'intéresse ensuite à une autre série X_t^5 représentée ci-dessous.



Il propose et cherche ensuite à valider un modèle. Il obtient les sorties suivantes:

```
##
## Call:
## arima(x = X5, order = c(3, 1, 5), include.mean = F, method = c("ML"))
##
## Coefficients:
##          ar1     ar2      ar3      ma1      ma2      ma3     ma4      ma5
##       0.4211  0.1058  -0.0110  -0.3375  -0.1663  -0.5026  0.4278  -0.2350
## s.e.  0.3656  0.2426   0.1545   0.3613   0.2370   0.1593  0.2237   0.1854
##
## sigma^2 estimated as 3.671:  log likelihood = -619.71,  aic = 1257.42
```

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

```
##  ar1  ar2  ar3  ma1  ma2  ma3  ma4  ma5
## 0.25 0.66 0.94 0.35 0.48 0.00 0.06 0.20
```

2. Préciser le modèle choisi par notre statisticien. A-t-il raison de choisir un ordre de différentiation de 1?
3. Expliquer comment sont estimés les coefficients de ce modèle?
4. Quel critère de choix de modèle a-t-il choisi? Expliquer comment celui-ci a été obtenu.
5. A quoi correspond la ligne "s.e"? A quoi correspondent les p-values affichées?

6. Au vu de ces résultats que doit faire notre statisticien?
7. Il souhaite ensuite valider son modèle, quels autres outils de diagnostique lui proposez vous?