Master IES mention D3S

Projet de séries chronologiques

Ibrahima DIALLO 39014912

Ziyi LIU 37013377

Soit Xt le processus stationnaire solution de l'équation

$X_t = 0.4X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$,

où ε_t est un bruit blanc Gaussien de variance 1 et θ = (moyenne de vos mois de naissance)/13 (vous pouvez arrondir la valeur de θ à 2 chiffres après la virgule).

Quelques Rappels

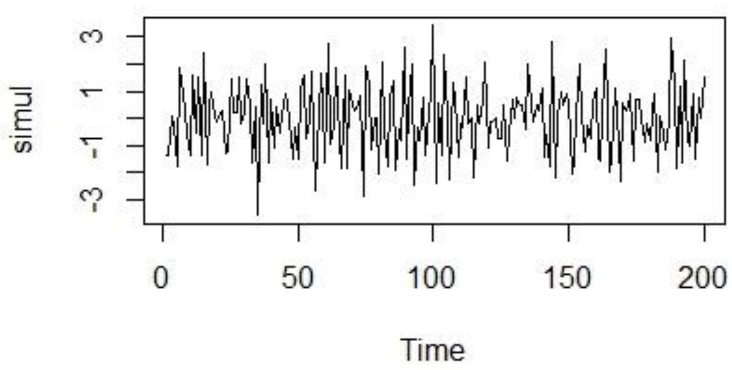
Une série $(X_t)_t$ est stationnaire si ses propriétés probabilistes sont les mêmes que celles de la série $(X_{t+h})_t$, pour tout entier h.

Le bruit blanc gaussien (ε_t) est un bruit blanc qui suit une loi normale de moyenne et variance données.

Question 1

Nous avons fait une simulation une réalisation de longueur n = 200 du processus stationnaire solution de cette équation. Le calcul de θ donne 1,23. Nous avons un ARMA (1,0,1) dont le p=1 car on a une seule variable explicative X_{t-1} et un q=1(ε_{t-1}).

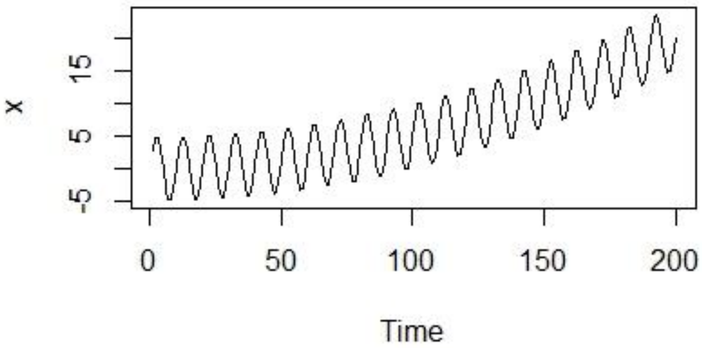
Nous obtenons un simul qui est stationnaire de moyenne E(Xt)=0.



(1) Plot de «simul »

Question 2

Nous avons créé une nouvelle série chronologique comme la somme de x + simul.



Le plot montre que la nouvelle série est non stationnaire et qu'on a un modèle additif . La non stationnarité peut être expliquée par la présence d'une ou plusieurs tendances dans la série.

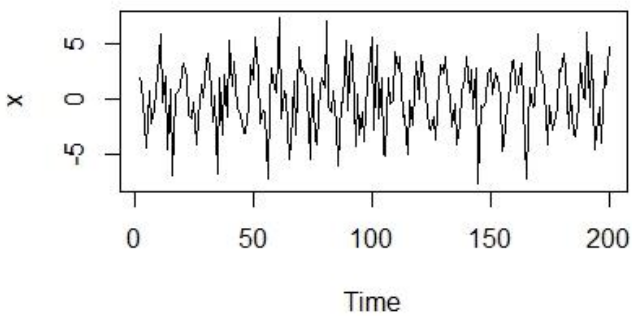
Pour confirmer notre analyse, nous avons effectué un test kpss.test qui montre bien que la série est non stationnaire.

```
        KPSS Test for Level Stationarity

data:  difft
KPSS Level = 0.019014, Truncation lag parameter = 4, p-value = 0.1
```

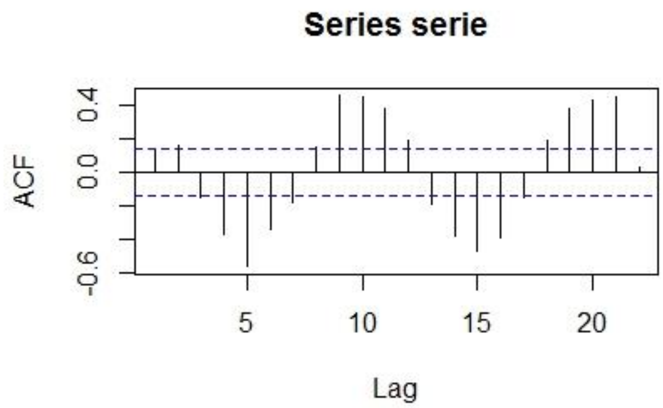
Question 3 : Ajustez le modèle SARIMA (Diapo 221 et 222)

Etape 1 : tracer la série pour proposer une structure.



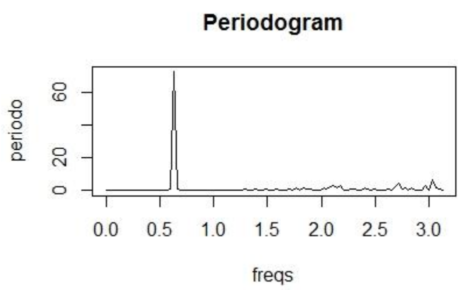
Nous constatons que la variance de la série n'est pas constante.

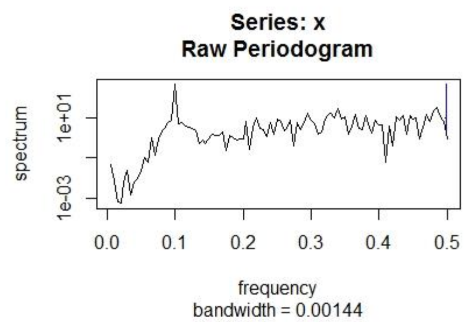
Etape 2: Nous avons tracer l'acf de la série et on a détecter la présence d'une tendance ou une composante saisonnière d'ordre 10. On constate les mêmes phénomènes qui se répètent par 10 ans.



Etape 3: Nous avons éliminé la tendance en procédant en 3 étapes:

(1) Calcul du periodog et Spectrum: le periodogram montre la taille du pic qui indique la dominance d'une période par rapport à une autre mais là il est difficile à expliquer car on a des composantes de types triangulaires (dents de scie) alors il est moins performant. Par contre le Spectrum montre plus précisément que la fréquence du max (le pic le plus grand) est égale 0.1=1/n d'où on en déduit que la période est égale à 10(n=10).

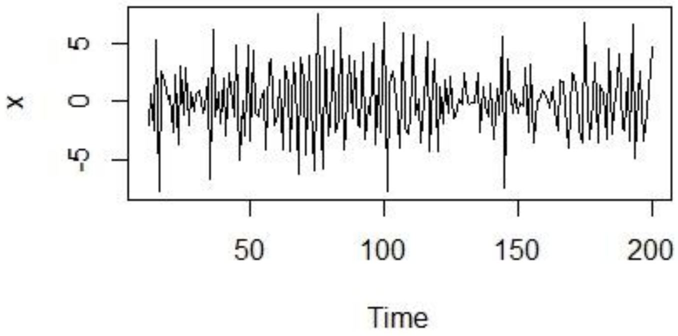




(2) Validation de la période r à l'aide du test de saisonnalité (Saison.test): Le test montre que la période=10 ce qui nous permet de valider nos résultats précédents.

```
     P'eriode         d.f.          Tobs      p-valeur
1.000000e+01 9.000000e+00 9.312057e+01 3.330669e-16
```

(3) Différenciation de la série pour faire disparaitre la composante saisonnière avec (diff(, lag=r))



(1) Plot de la série différenciée

Etape 4 : Nous avons tracer un nouveau acf pour voir s'il reste une tendance. On remarque qu'il reste de la tendance dans la série. Les tests des montées et des discordances (PtDisc.test) confirment bien l'existence d'une tendance avec un p-valeur inférieur au seuil alpha égale à 5%.

Nous avons encore différencié la série jusqu'à faire disparaitre la tendance et cette fois ci on a bien une série stationnaire. Les tests de montées et des discordances confirment bien la disparition de la tendance.

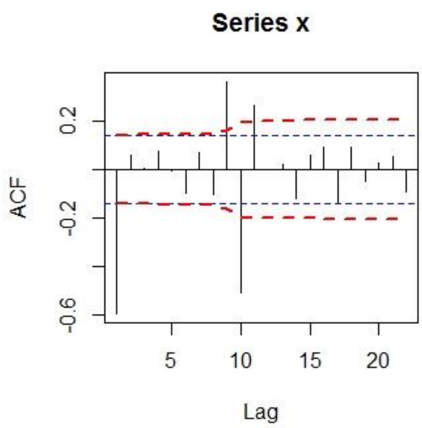
```
> #### Le résultat du test des montées
> PtMont.test(serie)#p-valeur = 0.8005421
          n          nM        stat     p-valeur
199.0000000 104.0000000   1.2247449   0.2206714
> #### Le résultat du test des discordances
> PtDisc.test(serie)#p-valeur = 0.8875603
          n          nD         stat      p-valeur
 199.0000000 9858.0000000   -0.0159704    0.9872580
```

1. On détermine le modèle de la série différentielle

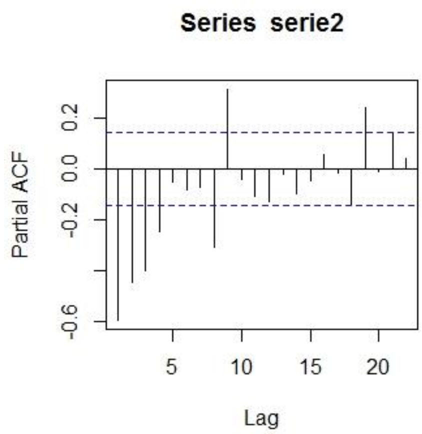
   On trace l'acf (acfMA), la pacf et l'eacf pour déterminer les ordres p et q maximaux.

   Si la série différenciée est un processus ARMA(p,q),alors la série initiale suit un modèle

   SARIMA(p,d,q)(0,1,0)[r].



(1)L'acf (acfMA)



(2) Le pacf

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o o o x x x  o  o  o
1 x o o o o o o o x x x  o  o  x
2 x o x o o o o o o x x  o  o  o
3 x x x x o o o o o x x  o  x  o
4 x x x o o o o o o x o  o  o  o
5 x x o o o o o o o x o  x  o  o
6 x o o x o o o o o x o  x  o  o
7 x x o x o x o o o x o  o  x  o
```

(3) Le eacf

Nous pouvons savoir que notre p= 2, et q = 1, donc on va faire un modèle SARIMA(2,d,1)(0,1,0)[12], pour choisir le meilleur modèle, on utilise la fonction 'auto.arima()' de la package 'forecast'.

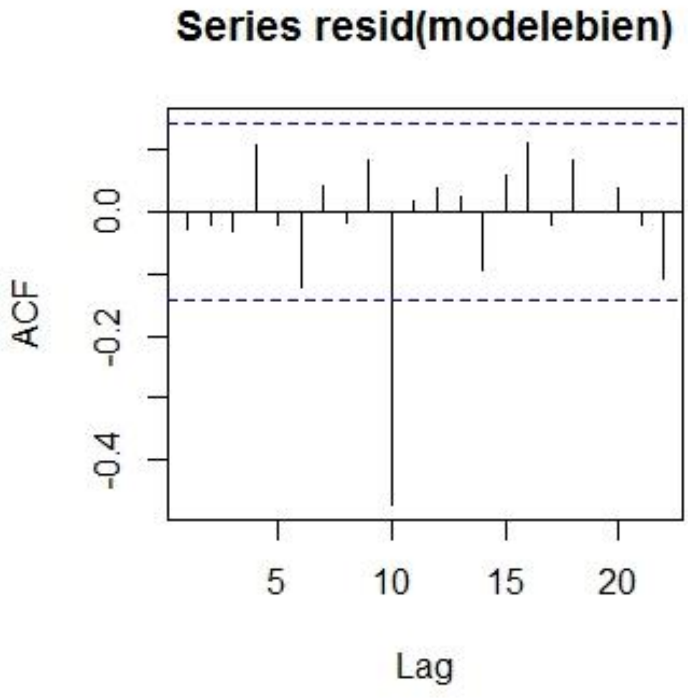
```
Series: serie2
ARIMA(2,0,1) with zero mean

Coefficients:
          ar1      ar2      ma1
      -0.3972  -0.1982  -0.9404
s.e.   0.0740   0.0737   0.0232

sigma^2 estimated as 3.536:  log likelihood=-387.64
AIC=783.29   AICc=783.51   BIC=796.26
```

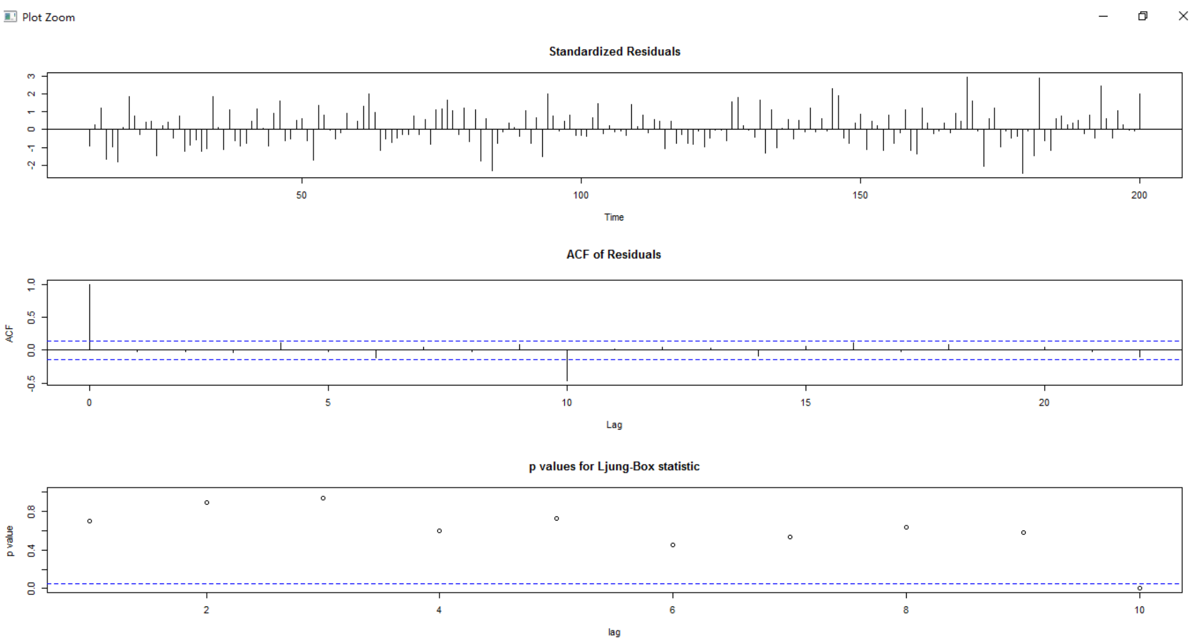
Donc, pour ce résultat-là, nous allons choisir le modèle ARIMA(2,0,1) (0,1,0) [10] qui est de moyenne 0 et pour ce modèle-là, le AIC = 783.29 et BIC = 796.26.

On regarde la distribution de la série du “modèle bien” :



Nous avons bon acf qui ne présente aucun signe de tendance

Et nous allons valider le modèle avec la commande tsdiag(modele) :



Par la distribution des résidus, nous savons que les résidus sont bruit blanc, et par l’acf des résidus, nous pouvons savoir que dès lag = 0, il n’y a des autorégression, et donc c’est à dire que on a bien simulé le modèle, et puis nous trouvons aussi que les p-valeurs se situent à un bon niveau aussi.

On essaie de prédire avec ce modèle (sans intervalle de confiance) :

```
> prevmodele
    Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
201      -5.211013 -7.620999 -2.801026 -8.896769 -1.525257
202       1.324642 -2.700277  5.349561 -4.830941  7.480225
```

Donc pour le X201, le modèle prédire à -5.211013, pour X202, il prédit à 1.324642.

Pour vérifier si ce modèle est bon , donc nous devons vérifier Si les innovations (modele$res) sont gaussiennes(Nous utilison 2 test ici.) :

Shapiro-Wilk normality test :

```
        Shapiro-Wilk normality test

data:  modelebien$residuals
W = 0.99543, p-value = 0.8391
```

H0: Les données de l'échantillon ne sont pas significativement différentes de la distribution normale.

H1: les données de l'échantillon sont très différentes de la distribution normale.

P—valeur est 0.8391, qui est supérieur à 0.05, donc on ne rejette pas H0, donc les innovations suivent la loi gaussinne.

Jarque Bera Test :

```
        Jarque Bera Test

data:  serie2
X-squared = 0.91428, df = 2, p-value = 0.6331
```

H0 : les données suivent une loi normale.

H1 : les données ne suivent pas une loi normale.

P—valeur est 0.6331, qui est supérieur à 0.05, donc on ne rejette pas H0, donc les innovations suivent la loi gaussinne.

Donc avec les 2 résultats, nous pouvons bien savoir que les innovations de notre modèle suive la loi gaussienne.

Donc maintenant on va estimer les prévisions à l’ordre 1 et 2, i.e., pour X201,X202(avec l'intervalle confiance) :

```
> prediction
    Point Forecast   Lo 99.5      Hi 99.5
201      -5.211013 -10.4897  0.06767721
202       1.324642  -7.4913 10.14058433
```

Ici, nous avons presque la meme prédiction avec les prédictions au dessus. Et puis on trace le graphique des prédictions :