Séries Temporelles: TP1

Commentaires sur le TP

- Il y a 7 exercices:
  - Exercice 0: Manipulations basiques sur la série "AirPassengers" (disponible dans la distribution basique de R).
  - Exercice 1 Simulations dans le modèle T+S+X et décomposition
  - Exercice 2 Simulations dans différents modèles, calcul des ACFs et PACFs, c'est un premier pas vers la sélection de modèles
  - Exercice 3 Processus AR(2) et leurs ACFs
  - Exercice 4 Prédiction dans un modèle AR(1)
  - Exercice 5 Prédiction dans un modèle MA via l'algorithme des innovations
  - Exercice 6: Analyse de la série "AirPassengers".
- Avant de commencer, charger les packages R: `ggfortify`, `astsa`, `forecast`.

Exercice 1: Manipulations de base sur la série "AirPassengers" (disponible dans la distribution basique de R)

Nous allons travailler sur la série "AirPassengers" (disponible dans la distribution basique de R).

1. Chargez les packages
1. Chargez les données et afficher leur description.
1. Le type `ts`
   - Vérifiez le type de l'objet `AirPassengers`
   - Lisez l'aide sur l'objet `ts`
   - Méthodes associées à cet objet
3. Représenter la série
4. Que font ces commandes ?

```
start(AirPassengers)
end(AirPassengers)
frequency(AirPassengers)
deltat(AirPassengers)
summary(AirPassengers)
cycle(AirPassengers)
boxplot(AirPassengers~cycle(AirPassengers))
```

5. Obtenez les ACF et PACF de cette série.
6. Création d'un objet `ts`. Supposez maintenant que vous avez une série brute `as.numeric(AirPassengers)` et l'information sur la date du premier enregistrement et sur la fréquence d'enregistrement (`start = c(1949,1)`, `frequency=12`). Pouvez vous recréer un objet `ts` ?

Exercice 2: Simulation

1. Simulez 1200 réalisations d'un bruit blanc gaussien $WN(0, 1)$. Représentez la série, notée BB, ses ACF et PACF
2. Choisissez et additionnez à BB une tendance polynomiale T et une saisonnalité S dont les amplitudes sont raisonnables comparées à la variance du bruit. Créez une série ST de type `ts`, en supposant que les observations ont commencé en Janvier 1900 et sont mensuelles.
3. Représentez les séries BB,S,T,TS et comparez avec les résultats de la fonction `slt`. Commentez.
4. Qu'est ce qui est représenté par la fonction `monthplot` ? Quelle information sera vous apporte sur la série ?

Exercice 3: Séries temporelles, ACF, PACF

1. Le code suivant génère des réalisations (of size $n = 500$) de 4 séries temporelles T_1, T_2, T_3, T_4. Pouvez vous déduire de leurs ACF et PACF les modèles qui ont été utilisés ?

```
n = 500
T1 = arima.sim(n=n,list(ma = c(3, 1)),sd = sqrt(0.1796))
T2 = arima.sim(n=n,list( ar = c(-0.9,-0.5)),sd = sqrt(0.1796))
T3 = cumsum(rnorm(n) + 0.2)
T4 = 2*cos(2*pi*1:n/50 + 0.6*pi) + rnorm(n,0,0.5)
```

2. Changez n pour 200 les conclusions sont-elles si claires ?

Exercice 4: Les processus AR(2) et leur ACF

(from Brockwell and Davis p91 - example 3.2.4)

Considerez un modèle AR(2) stationnaire

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \omega_t$$

où ω est un bruit blanc gaussien avec variance 1. Notons ξ_1 et ξ_2 les racines de $1 - \phi_1 z - \phi_2 z^2$. Dans ce cas, l'ACF vérifie

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2)$$

avec les conditions initiales $\rho(0) = 1$ et $\rho(1) = \phi_1/(1 - \phi_2)$ et est donc donnée par

$$\rho(h) = \frac{(\xi_2^2 - 1)\xi_1^{1-h} - (\xi_1^2 - 1)\xi_2^{1-h}}{(\xi_1\xi_2 + 1)(\xi_2 - \xi_1)}$$

quand $\xi_1 \neq \xi_2$. Nous illustrons dans cet exercice les différents comportements de l'ACF ρ.

Vous trouverez ci-dessous un code qui génère un "stem plot", vous en aurez besoin pour représenter les vrais ACF.

```
stem <- function(x,y,pch=16,linecol=1,clinecol=1,...){
if (missing(y)){
    y = x
    x = 1:length(x) }
    plot(x,y,pch=pch,...)
    for (i in 1:length(x)){
       lines(c(x[i],x[i]), c(0,y[i]),col=linecol)
    }
```

```
    lines(c(x[1]-2,x[length(x)]+2), c(0,0),col=clinecol)
}
#An example
x <- seq(0, 2*pi, by = 0.2)
stem(x,sin(x), main = 'Default style')
```

1. Écrivez

- une fonction `roots2ar` qui calcule les coefficients ϕ_1 et ϕ_2 à partir des racines ξ_1 et ξ_2.
- une fonction `true_acf` qui calcule l'ACF au point h à partir de ξ_1 et ξ_2

2. Choisssez des coefficients ϕ_1 et ϕ_2 tels que le processus est non-causal.
   - Essayez de simuler $n = 50$ observations à partir de la fonction `arima.sim`. Commentez le commentaire.
   - Simulez $n = 50$ observations à partir de la fonction `filter`. Vous devez simuler $n+20$ observations et garder seulement les 50 dernières pour vous débarrasser du problème d'initialisation.
   - Représentez la série et son ACF empirique. Que se passe-t-il ?
3. Un AR(2) causal à racines complexes
   - Choisissez 2 paires de racines complexes et conjuguées, l'une avec des modules proches de 1 mais telles que le processus est causal, l'autre avec des modules plus éloignés de 1.
   - Dans chaque cas, calculez ϕ_1 et ϕ_2 et simulez 500 observations des processus et représentez les.
   - Représentez les vraies ACFs et leurs versions empiriques et comparez les graphiques.
4. Choisissez 2 paires de racines réelles, une paire proche de 1 et l'autre plus éloignée.
   - Faire les mêmes graphiques qu'à la question 3.
   - Comparez les graphiques des ACF des questions 3 et 4. Quelle est la principale différence ? difference?

Exercice 5: Prédiction pour la série `lh` data

Nous travaillons maintenant avec la série `lh`.

1. Visualisez les données, leurs ACF et PACF.
2. Nous choisissons un modèle AR(1)
$$X_t = \phi X_{t-1} + \omega_t$$
   pour cette série(cf TP2 pour une justification).
   - Estimez la moyenne.
   - Proposez une estimation de ϕ et σ^2 basée sur les résultats des fonctions `acf` et `pacf` (vous aurez besoin de l'option `type="correlation"` de la fonction `acf`).
   - Comparez vos résultats avec `arima(lh, order = c(1, 0, 0))`. Pourquoi, à votre avis, y a-t-il une (petite) différence ?

- En utilisant les équations de prédiction, le fait que, pour un AR(1), $\gamma(h) = \phi\gamma(h-1)$, et vos estimations de $\gamma(0)$ et ϕ
  - calculez la matrice Γ_n (utilisez la fonction `toeplitz`) et son inverse(fonction `solve`)
  - et les vecteurs $\gamma_n^{(m)} = \big(\gamma(m), \ldots, \gamma(m+n-1)\big)^\top$ pour $m = 1$ à 12

- Proposez une prédiction pour les 12 prochaines valeurs de la série (ne pas oubliez de soustraire au préalable la moyenne)
- Construisez des intervalles de prédiction.
- Comparez vos résultats avec la méthode `predict` appliquée au résultat de la fonction `arima`.

Exercice 6 : Prédiction dans un modèle MA via l'algorithme des innovations

Nous considérons dans cet exercice les données `varve` disponibles dans le package `astsa`.

1. Représentez la série, ses ACF et PACF. Donnez des arguments pour motiver le choix de la transformation par log pour stationnariser la série.
2. Représentez la nouvelle série, ses ACF et PACF. La série vous semblet-elle stationnaire ? Expliquez pourquoi la différenciation devrait stationnariser le log de varve.
3. Différenciez le log de varve, représentez la nouvelle série, ses ACF et PACF et expliquer le choix du modèle MA(1) pour cette nouvelle série.
4. Estimations dans le modèle MA(1) $X_t = \omega_t + \theta\omega_{t-1}$.
   - Estimez θ (vous pouvez utiliser la fonction `polyroot`). Choisissez la solution qui rend le MA inversible.
   - er $\sigma^2 = \mathrm{Var}(\omega_t)$ (n'oubliez pas l'option `type = "covariance"` de la fonction `acf`)
   - Comparez vos estimations avec les résultats de `arima(lh, order = c(0, 0, 1))`
5. L'algorithme des innovations pour un modèle MA(1).
   - Codez l'algorithme des innovations et déduisez une prédiction et un intervalle de prédiction pour la valeur suivante.
   - Comparez vos résultats avec la méthode `predict` appliquée à la fonction `arima`.
   - Admettez que $X^n_{n+m} = \theta_{n+m-1,1}X_n$ où les $\theta_{n+m-1,1}$ sont déterminés par le même algortihme. Déduisez en les prédictions et leurs erreurs pour les 10 valeurs suivantes de la série `varve` elle-même.
   - Comparez avec ceux de la méthode `predict` appliquée à la fonction `arima`.

Exercice 6: Analyse de la série "AirPassengers"

Nous allons travailler sur la série "AirPassengers" (disponible dans la distribution basique de R).

1. Chargez les packages
1. Chargez les données et afficher leur description.
1. Le type `ts`
   - Vérifiez le type de l'objet `AirPassengers`
   - Lisez l'aide sur l'objet `ts`
   - Méthodes associées à cet objet
3. Représenter la série
4. Que font ces commandes ?

```r
start(AirPassengers)
end(AirPassengers)
frequency(AirPassengers)
deltat(AirPassengers)
summary(AirPassengers)
cycle(AirPassengers)
boxplot(AirPassengers~cycle(AirPassengers))
```

5. Obtenez les ACF et PACF de cette série.
6. Création d'un objet `ts`. Supposez maintenant que vous avez une série brute `as.numeric(AirPassengers)` et l'information sur la date du premier enregistrement et sur la fréquence d'enregistrement (`start = c(1949,1)`, `frequency=12`). Pouvez vous recréer un objet `ts` ?
8. Proposer une transformation de la série de façon à stabiliser la variance.
9. On se propose d'estimer la tendance par moindres carrés en utilisant un modèle linéaire. Tracer la tendance estimer sur le graphe de la série. Commenter
10. On se propose d'estimer la tendance par moyenne mobiles. Tracer la tendance estimée sur le graphe de la série. Commenter
11. Proposer une décompostion additive de la série en utilisant la commande

```r
res=decompose(data2,type="additive")
```

12. On se propose de stationnariser la série en appliquand l'opérateur retard d'ordre 12 pour enlever la saisonnalité puis l'opérateur retard d'ordre 1 sur cette nouvelle série. Mettre en oeuvre avec les commandes et expliquer comment fonctionne cette méthode de stationnarisation.

```r
y=diff(data2,lag=12,differences=1)
x=diff(y,lag=1,differences=1)
```

13. Comparer les ACF et PACF de la série initiale et de la série stationnarisée.
14. Mettre en oeuvre plusieurs modèles ARMA(p,q) sur la série initiale en utilisant des commandes du type

```r
z4<-arima(data2, order=c(1,1,1), seasonal=list(order=c(1,1,0), period=12))
```

15. On se propose de sélectionner le meilleur modèle à l'aide du critère AIC. Présenter ce critère dans le cadre des séries temporelles et déduire des sorties R précédentes le meilleur modèle $ARMA(p,q)$.
16. Etudier les résidus du modèle sélectionné à l'aide de la commande:

```r
tsdiag(z3, gof.lag=20)
```

17. Etudier la prédiction obtenue à partir du modèle sélectionné à l'aide de la commande:

```r
prev<-predict(z3, n.ahead=12)
```