

Corrigé de la feuille d'exercices no 8 (révisions)

Exercice 1.

- (1) Pour tout t , $\mathbb{E}(X_t) = \phi\mathbb{E}(X_{t-1}) + 0$. Comme X est stationnaire, $\mathbb{E}(X_t) = \mathbb{E}(X_{t-1})$, et donc, puisque $|\phi| < 1$, $\mathbb{E}(X_t) = 0$. Donc, pour tout t , $\mathbb{E}(Y_t) = \mathbb{E}(X_t) - \mathbb{E}(X_{t-1}) = 0$.
 Pour tout t , $\mathbb{E}(Y_t^2) = \text{Var}(X_t) + \text{Var}(X_{t-1}) - 2\text{Cov}(X_t, X_{t-1})$, et puisque X est stationnaire, cette quantité ne dépend pas de t . Pour tout t et $h > 0$,

$$\begin{aligned}\mathbb{E}(Y_t Y_{t+h}) &= \mathbb{E}((X_t - X_{t-1})(X_{t+h} - X_{t+h-1})) \\ &= \text{Cov}(X_t, X_{t+h}) - \text{Cov}(X_t, X_{t+h-1}) - \text{Cov}(X_{t-1}, X_{t+h}) + \text{Cov}(X_{t-1}, X_{t+h-1}),\end{aligned}$$

qui ne dépend pas de t puisque X est stationnaire. Donc Y est stationnaire.

On calcule :

$$\begin{aligned}Y_t &= \phi X_{t-1} + Z_t - (\phi X_{t-2} + Z_{t-1}) \\ &= \phi Y_{t-1} + Z_t - Z_{t-1}.\end{aligned}$$

Donc Y est un $ARMA(1,1)$.

- (2) On calcule (pour t quelconque) :

$$\text{Var}(X_t) = \phi^2 \text{Var}(X_{t-1}) + \sigma^2.$$

Donc

$$\text{Var}(X_t) = \frac{\sigma^2}{1 - \phi^2}.$$

Nous avons pour tout t :

$$\begin{aligned}\mathbb{E}(X_t X_{t-1}) &= \phi \text{Var}(X_{t-1}) + 0 \\ &= \frac{\phi \sigma^2}{1 - \phi^2}.\end{aligned}$$

D'où :

$$\begin{aligned}\text{Var}(Y_t) &= \text{Var}(X_t) + \text{Var}(X_{t-1}) - 2\text{Cov}(X_t, X_{t-1}) \\ &= \frac{2\sigma^2}{1 - \phi^2} - 2 \left(\frac{\phi \sigma^2}{1 - \phi^2} \right) \\ &= \frac{2\sigma^2(1 - \phi)}{1 - \phi^2} = \frac{2\sigma^2}{1 + \phi}.\end{aligned}$$

Exercice 2. Calculons (on admet que les séries sont convergentes, on remarque que ce sont des séries dans lesquelles beaucoup de termes se simplifient) :

$$\begin{aligned}X_t - \frac{2}{3}X_{t-1} &= \sum_{k \geq 0} \left[\frac{1}{2^k} \left(X_{t-k} - \frac{7}{6}X_{t-k-1} + \frac{1}{3}X_{t-3} \right) \right] \\ &= \sum_{k \geq 0} \left[\frac{1}{2^k} \left(Z_{t-k} - \frac{1}{4}Z_{t-k-1} - \frac{1}{8}Z_{t-k-2} \right) \right] \\ &= Z_t + \frac{1}{4}Z_{t-1}.\end{aligned}$$

Exercice 5.

- (1) Pour tracer le périodogramme : `k=kernel("daniell",4); spec.pgram(sncf,k,taper=0,log='no')`.
On obtient la figure 0.1.

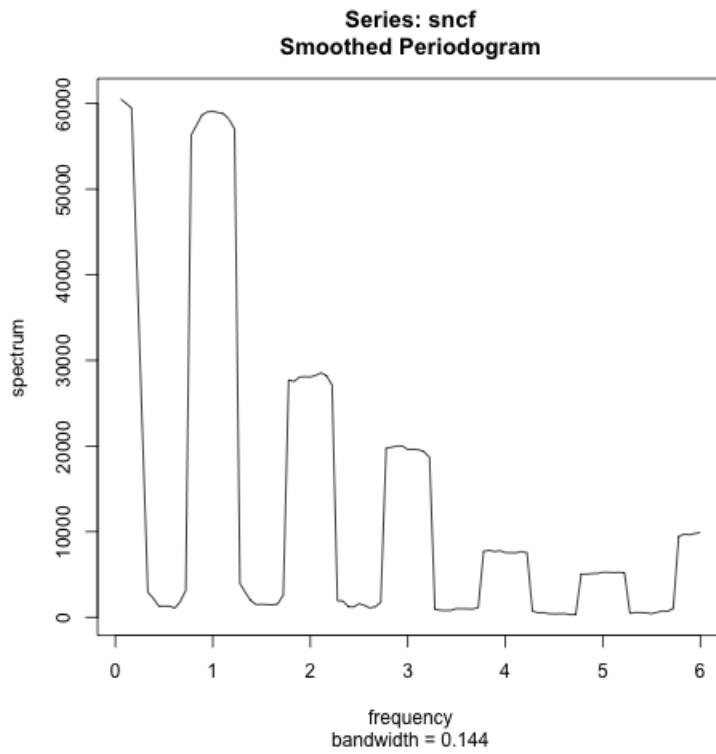


FIGURE 0.1. périodogramme

Rappel : les abscisses vont de 0 à π (l'échelle affichée n'est pas correcte). Le pic le plus important se trouve donc en $\pi/6$, ce qui correspond à une période $2\pi/(\pi/6) = 12$ (ce qui est bien cohérent avec le graphique de la série et les autocorrélations (`png(filename=~"/Dropbox/sylvain-rubenthala/par(mfrow=c(2,1)); plot(sncf); acf(sncf,lag.max=50); dev.off()`, figure 0.2)

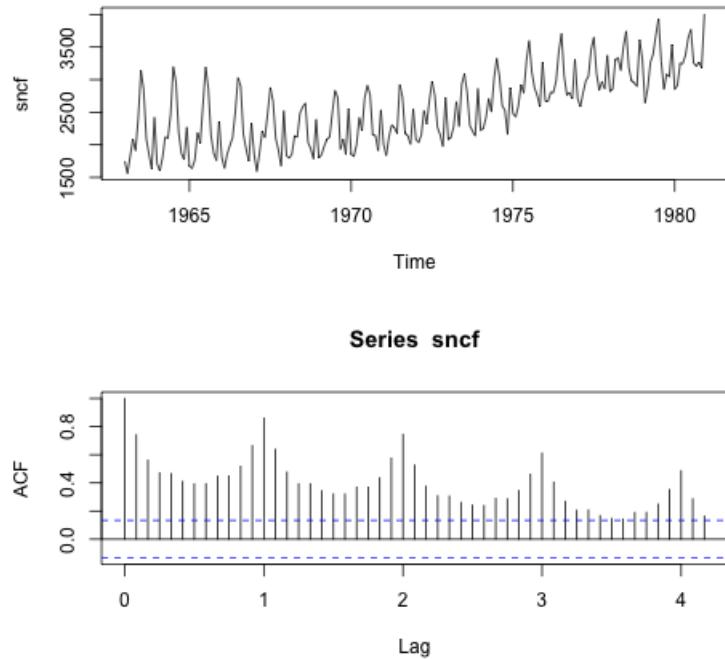


FIGURE 0.2. graphique et ACF

- (2) Calculons les séries demandées : $y1=diff.ts(sncf,lag=1)$; $y2=diff.ts(sncf,lag=12)$; $y3=diff.ts(y1,lag=12)$.
 - On commence par $y1$ (figure 0.3).

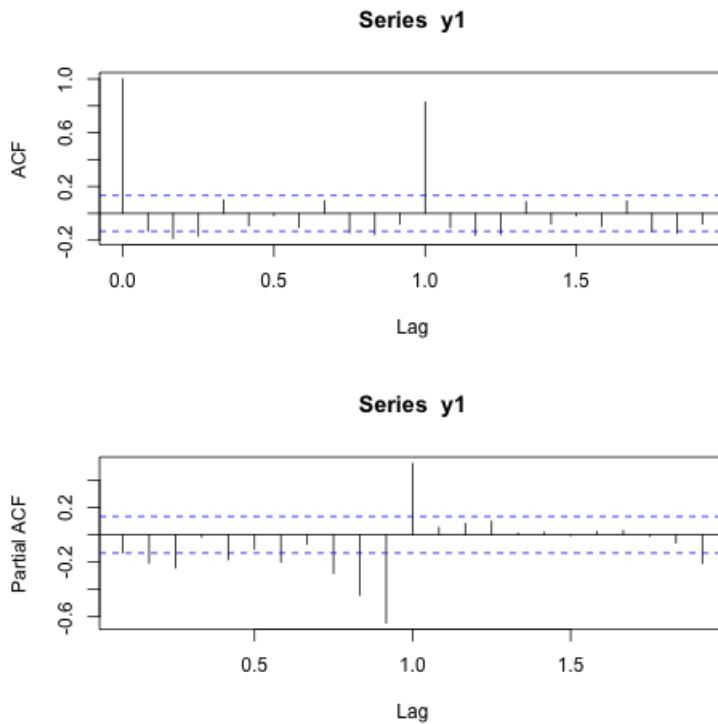


FIGURE 0.3.

On se dit que Y_1 pourrait être un $AR(3)$, et donc que X pourrait être un $ARIMA(3, 1, 0)$. On estime les coefficients par : `out<-arima(sncf, order=c(3,1,0))`. On fait un test de niveau 0,05 pour savoir si les résidus forment un bruit blanc. L'instruction `Box.test(out$resid, lag=10)` renvoie une p -valeur de 0,002345 (le test calcule une statistique et la p -valeur est la probabilité d'observer une statistique aussi grande sous l'hypothèse « bruit blanc »). Comme la p -valeur est plus petite que 0,05, on rejette l'hypothèse « bruit blanc ». Notre processus ne peut pas être un $ARIMA(3, 1, 0)$.

- On continue avec Y_2 (figure 0.4).

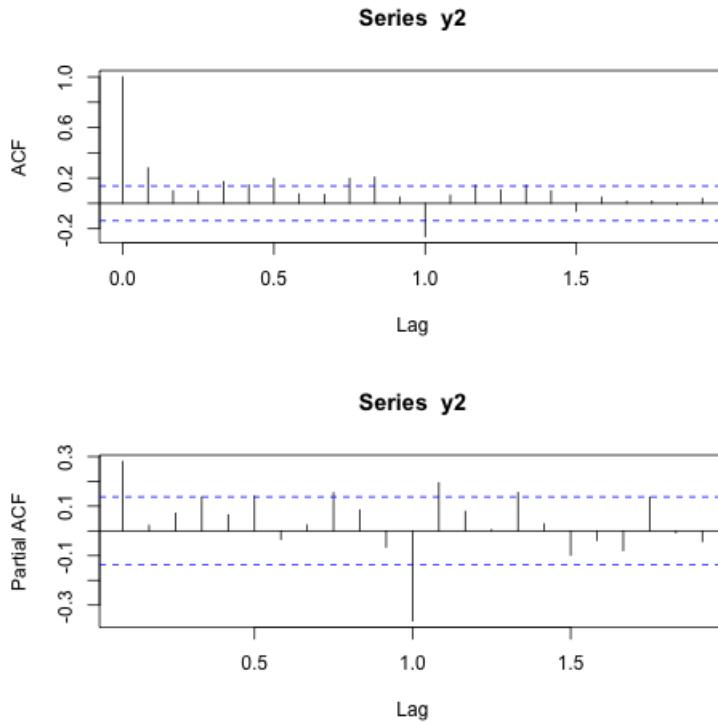


FIGURE 0.4.

On se dit que Y_2 pourrait être un $AM(1)$ et donc que X pourrait être un $SARIMA(0, 0, 1, 12)$.
On estime les coefficients par `out<-arima(sncf, order=c(0, 0, 1), seasonal=list(order=c(0, 0, 1), period=12))`.
On fait un test de niveau 0,05 pour savoir si les résidus forment un bruit blanc.
L'instruction `Box.test(out$resid, lag=10)` renvoie une p -valeur de $2,2 \times 10^{-16}$, donc on rejette l'hypothèse « bruit blanc ». Notre processus ne peut pas être un $SARIMA(0, 0, 1, 12)$.
– On continue avec Y_3 (figure)

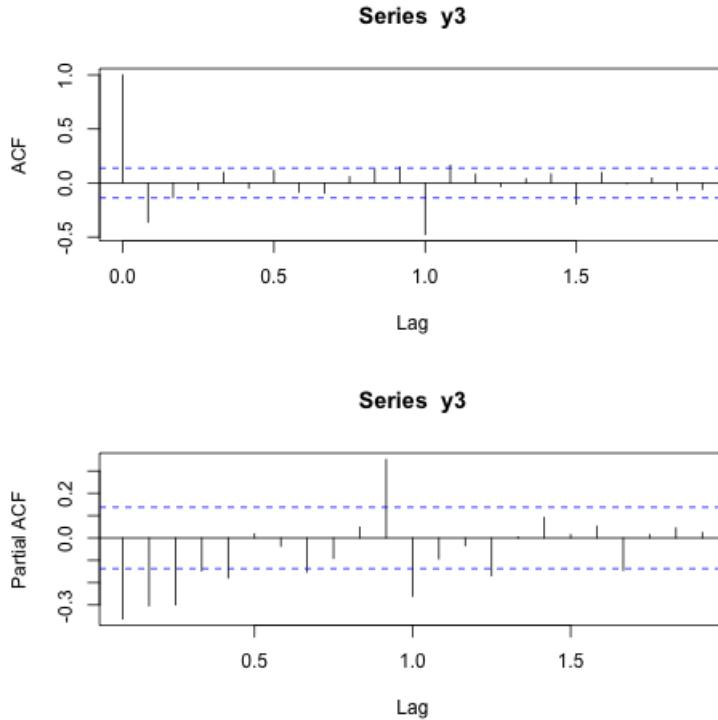


FIGURE 0.5.

On se dit que Y_3 pourrait être un $MA(2)$ et que donc X pourrait être un $SARIMA(0, 1, 2, 12)$.
On estime les coefficients par `out<-arima(sncf, order=c(0,1,2), seasonal=list(order=c(0,1,2), period=12))`.
On fait un test de niveau 0,05 pour savoir si les résidus forment un bruit blanc.
L'instruction `Box.test(out$resid, lag=10)` renvoie une p -valeur de 0,4596 ($> 0,05$) donc on reste sur l'hypothèse « bruit blanc ». Donc notre processus X a bien l'air d'être un $SARIMA(0, 1, 2, 12)$.

- (3) On exécute une boucle

```

p=0
q=0
out<-arima(sncf, order=c(0,1,2), seasonal=list(order=c(0,1,2), period=12))
m=out$aic
for (i in 0:2)
{
  for (j in 0:2)
  {
    if (i+j>0)
    {
      out<-arima(sncf, order=c(i,1,j), seasonal=list(order=c(i,1,j), period=12))
      if (out$aic>m) { p=i q=j }
    }
  }
}
cat(p,q)

```

Le meilleur couple maximisant le critère AIC est $p = 2, q = 0$ (qui ne correspond pas aux graphiques ACF et PACF ci-dessus).

- (4) On exécute le code (graphique dans la figure 0.6)

```
out02<-arima(sncf,order=c(0,1,2),seasonal=list(order=c(0,1,2),period=12))
outpred<-predict(out02,n.ahead=12)
pred=outpred$pred
ts.plot(sncf,pred,col=c('black','red'))
```

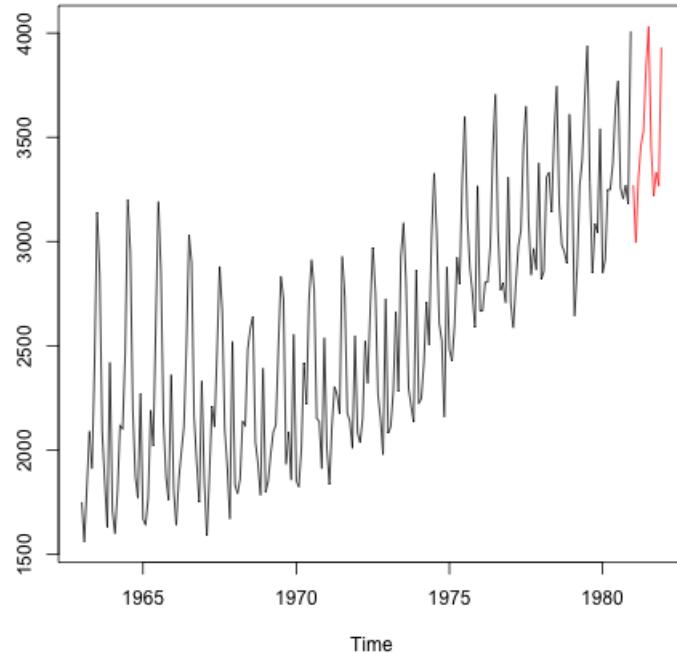


FIGURE 0.6.

Pour avoir un graphique convaincant, il faudrait calculer des intervalles de confiance.