Introduction à l'analyse des séries temporelles

Banque d'exercices corrigés

Cette banque d'exercices corrigés a pour vocation d'illustrer les propriétés du cours, et la rigueur demandée dans la résolution des exercices. Les exercices présentés sont généralement plus difficiles que ce qui est attendu au moment des devoirs. Il existe vraisemblablement des erreurs dans les corrigés des exercices, n'hésitez pas à m'en informer le cas échéant.

Partie 1 : stationnarité, fonction d'autocovariance

Exercice 1

Motivation On montre que le modèle de régression linéaire simple ne produit pas un processus stationnaire, et que la différentiation à l'ordre 1 est une technique suffisante pour retirer une tendance linéaire du modèle.

On considère la série temporelle définie par

$$x_t = \beta_0 + \beta_1 t + w_t,$$

où β_0 et β_1 désignent des constantes, et w_t un bruit blanc gaussien de variance σ^2.

1. x_t est-elle stationnaire ?
2. Montrer que le processus $y_t = x_t - x_{t-1}$ est stationnaire.
3. ** Montrer que l'espérance du processus moyenne mobile

$$y_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j}$$

est $\beta_0 + \beta_1 t$ et donner une expression simplifiée de sa fonction d'autocorrélation.

Solution

1. L'espérance de la série, donnée par $\mathbb{E}[x_t] = \beta_0 + \beta_1 t$, n'est pas constante (elle dépend de t). La série n'est donc pas stationnaire.
2. On a

$$y_t = (\beta_0 + \beta_1 t + w_t) - (\beta_0 + \beta_1 (t-1) + w_{t-1}) = \beta_1 + w_t - w_{t-1}.$$

Cette série est d'espérance constante, égale à $\mathbb{E}[y_t] = \beta_1$, et sa fonction d'autocovariance vaut :

$$\begin{aligned} \gamma(t,t) &= \mathbb{E}[(w_t - w_{t-1})^2] = \mathbb{E}[w_t^2] + \mathbb{E}[w_{t-1}^2] = 2\sigma^2, \\ \gamma(t,t-1) &= \mathbb{E}[(w_t - w_{t-1})(w_{t-1} - w_{t-2})] = -\mathbb{E}[w_{t-1}^2] = -\sigma^2, \\ \gamma(t,t-k) &= \mathbb{E}[(w_t - w_{t-1})(w_{t-k} - w_{t-k-1})] = 0 \qquad \text{si } k > 1, \end{aligned}$$

où on a utilisé le fait que, pour $s \neq t$, $\mathbb{E}[w_s w_t] = \mathbb{E}[w_s]\mathbb{E}[w_t] = 0$. Comme la fonction d'autocovariance $\gamma(s,t)$ ne dépend que du décalage $|s-t|$ et pas de s ni de t (en effet, soit $|s-t| = 0$ ou 1 et l'autocovariance vaut $2\sigma^2$ ou $-\sigma^2$ respectivement, soit $|s-t| > 1$ et l'autocovariance vaut 0), on conclut que la série est stationnaire.

3. On écrit

$$\begin{aligned} y_t &= \frac{1}{2q+1}\sum_{j=-q}^{q} \beta_0 + \beta_1(t-j) + w_{t-j} \\ &= \frac{1}{2q+1}\left((2q+1)\beta_0 + (2q+1)\beta_1 t + \beta_1 \sum_{j=-q}^{q}(-j) + \sum_{j=-q}^{q} w_{t-j}\right) \\ &= \beta_0 + \beta_1 t + \frac{1}{2q+1}\sum_{j=-q}^{q} w_{t-j}, \end{aligned}$$

où on a utilisé le fait que pour la somme sur $(-j)$, chaque terme et son opposé apparaissent une fois : $\sum_{j=-q}^{q}(-j) = \sum_{j=1}^{q}(-j) + \sum_{j=1}^{q}(+j) = 0$.

On en déduit que $\mathbb{E}[y_t] = \beta_0 + \beta_1 t$, et que

$$\gamma(t, t+h) = \frac{1}{(2q+1)^2}\mathrm{Cov}\left(\sum_{i=-q}^{q} w_{t-i}, \sum_{j=-q}^{q} w_{t+h-j}\right).$$

Les seuls termes non nuls sont ceux tels que $t - i = t + h - j$, c'est-à-dire $j = i + h$. Il y en a $2q+1-h$. Il vient que

$$\gamma(t, t+h) = \frac{2q+1-h}{(2q+1)^2}\sigma^2.$$

On peut ensuite calculer la fonction d'autocorrélation :

$$\rho(t, t+h) = \frac{\gamma(t, t+h)}{\sqrt{\gamma(t,t), \gamma(t+h, t+h)}},$$

où $\gamma(t,t) = \gamma(t+h, t+h) = \frac{1}{2q+1}\sigma^2$. Il vient que

$$\rho(t, t+h) = \frac{2q+1-h}{2q+1} = 1 - \frac{h}{2q+1}.$$

Exercice 2

Motivation On étudie les propriétés de moment (espérance, variance, fonction d'autocovariance) de la marche aléatoire avec dérive.

On considère une marche aléatoire avec dérive,

$$x_t = \delta + x_{t-1} + w_t,$$

où la condition initiale est donnée par $x_0 = 0$, et w_t désigne un bruit blanc gaussien de variance σ^2.

1. * Montrer que le modèle peut être écrit comme $x_t = \delta t + \sum_{k=1}^{t} w_k$.
2. * Calculer l'espérance et la fonction d'autocovariance de x_t.
3. x_t est-elle stationnaire ?
4. ** Montrer que, pour h fixé, $\rho(t-h, t) = \sqrt{\frac{t-h}{t}} \to 1$ lorsque $t \to \infty$. Comment cela se manifeste-t-il ?
5. Suggérer une transformation pour rendre la série stationnaire.

Solution

1. On montre que :

$$\begin{aligned} x_t &= \delta + x_{t-1} + w_t \\ &= \delta + (\delta + x_{t-2} + w_{t-1}) + w_t \\ &= 2\delta + x_{t-2} + w_{t-1} + w_t. \end{aligned}$$

   Puis, par récurrence :

$$\begin{aligned} x_t &= \delta + x_{t-1} + w_t \\ &= 2\delta + x_{t-2} + w_{t-1} + w_t \\ &= 3\delta + x_{t-3} + w_{t-2} + w_{t-1} + w_t \\ &= \dots \\ &= \delta t + x_0 + w_1 + w_2 + \dots + w_t \\ &= \delta t + \sum_{k=1}^{t} w_k. \end{aligned}$$

2. On a, pour tout $t \geq 1$,

$$\mathbb{E}[x_t] = \mathbb{E}\left[\delta t + \sum_{k=1}^{t} w_k\right] = \delta t + \sum_{k=1}^{t} \mathbb{E}[w_k] = \delta t,$$

   où la deuxième égalité découle de la linéarité de l'espérance.

   Pour la fonction d'autocovariance, on considère $s \neq t$. On a

$$\begin{aligned} \gamma(s,t) = \mathrm{Cov}(x_s, x_t) &= \mathbb{E}\left[\sum_{k=1}^{s} w_k, \sum_{l=1}^{t} w_l\right] \\ &= \sum_{k=1}^{s}\sum_{l=1}^{t} \mathbb{E}[w_k w_l] \\ &= \sum_{k=1}^{\min(s,t)} \mathbb{E}[w_k] = \min(s,t)\,\sigma^2. \end{aligned}$$

3. Comme l'espérance de x_t n'est pas constante pour $\delta \neq 0$, la série n'est pas stationnaire. Par ailleurs, même avec $\delta = 0$, la fonction d'autocovariance de la série n'est pas invariante par translation. Par exemple, on a $\gamma(1,2) = \sigma^2 \neq \gamma(2,3) = 2\sigma^2$, alors que dans les deux cas, le décalage entre les deux pas de temps considérés est de 1. En conséquence, la série n'est pas stationnaire (quelque soit la valeur de δ).

4. On commence par calculer la variance de la série :

$$\mathrm{Var}(x_t) = \mathrm{Var}\left(\delta t + \sum_{k=1}^{t} w_k\right) = \sum_{k=1}^{t} \mathrm{Var}(w_k) = t\sigma^2,$$

   où la deuxième égalité découle du fait que les w_k sont indépendants.

   Ensuite, la fonction d'autocorrélation se calcule directement :

$$\rho(t-h,t) = \frac{\gamma(t-h,t)}{\sqrt{\gamma(t-h,t-h)\gamma(t,t)}} = \frac{(t-h)\sigma^2}{\sqrt{(t-h)\sigma^2 \cdot t\sigma^2}} = \sqrt{\frac{t-h}{t}}.$$

Par ailleurs, pour h fixé, comme $h/t \to 0$,

$$\rho(t-h,t) = \sqrt{\frac{t-h}{t}} = \sqrt{1-\frac{h}{t}} \to \sqrt{1} = 1.$$

Cela a pour conséquence que, lorsqu'on regarde la fonction d'autocorrélation (ACF) d'une marche aléatoire avec dérive, les pics d'autocorrélation tendent vers 1 lorsque le nombre d'observations est grand.

5. En effectuant une différentiation d'ordre 1, c'est-à-dire en étudiant les variations de la série, on obtient $\nabla x_t = x_t - x_{t-1} = \delta + w_t$. Il s'agit d'une série stationnaire, de fonction d'autocovariance $\gamma(h)$ nulle pour tout $h \neq 0$.

Exercice 3

Motivation On montre qu'un processus autorégressif d'ordre 1 peut s'écrire comme un processus moyenne mobile, ce qui permet d'établir un certain nombre de propriétés intéressantes.

On considère un processus AR(1) de la forme $x_t = \phi x_{t-1} + w_t$, avec $|\phi| < 1$ une constante, w_t un bruit blanc gaussien de variance σ^2 et condition initiale $x_0 = w_0$.

1. * Montrer que l'on peut écrire $x_t = \sum_{j=0}^{t} \phi^j w_{t-j}$.
2. En déduire l'espérance de x_t.
3. Calculer la variance de x_t.
4. * Montrez que, pour $h \geq 0$, $\mathrm{Cov}(x_{t+h}, x_t) = \phi^h \mathrm{Var}(x_t)$.
5. La série est-elle stationnaire ?
6. * Argumenter que, lorsque $t \to \infty$, le processus devient "asymptotiquement stationnaire".
7. * Discuter comment ces résultats peuvent être utilisés pour simuler n observations d'un processus **stationnaire** AR(1) à partir d'un vecteur de gaussiennes $\mathcal{N}(0,1)$ i.i.d.
8. * On suppose désormais que $x_0 = w_0/\sqrt{1-\phi^2}$. Ce processus est-il stationnaire ?

Solution

1. On procède par récurrence. On a :

$$\begin{aligned}
x_t = \phi x_{t-1} + w_t &= \phi(\phi x_{t-2} + w_{t-1}) + w_t \\
&= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\
&= \phi^2(\phi x_{t-3} + w_{t-2}) + \phi w_{t-1} + w_t \\
&= \phi^3 x_{t-3} + \phi^2 w_{t-2} + \phi w_{t-1} + w_t \\
&= \dots \\
&= \phi^t x_0 + \phi^{t-1} w_1 + \dots + \phi w_{t-1} + w_t \\
&= \sum_{j=0}^{t} \phi^j w_{t-j}.
\end{aligned}$$

2. Directement, il vient $\mathbb{E}[x_t] = \sum_{j=0}^{t} \mathbb{E}[w_{t-j}] = 0$.
3. Comme les w_{t-j} sont indépendants, on a

$$\mathrm{Var}(x_t) = \sum_{j=0}^{t} \phi^{2j} \sigma^2 = \sigma^2 \frac{1-\phi^{2(t+1)}}{1-\phi^2},$$

où la dernière égalité provient du fait que $\sum_{j=0}^{t} \phi^{2j}$ est la somme d'une suite géométrique de raison ϕ^2.

4. En utilisant les résultats de la question 1, on a
$$\mathrm{Cov}(x_{t+h}, x_t) = \mathrm{Cov}\left(\sum_{i=0}^{t+h} \phi^i w_{t+h-i}, \sum_{j=0}^{t} \phi^j w_{t-j}\right).$$
Les seuls termes non nuls sont ceux tels que $t+h-i = t-j$, c'est-à-dire $j+h = i$. On en déduit que
$$\mathrm{Cov}(x_{t+h}, x_t) = \sum_{j=0}^{t} \phi^{j+h}\phi^j\sigma^2 = \phi^h \sum_{j=0}^{t} \phi^{2j}\sigma^2.$$
On reconnait dans le terme de droite la variance de x_t, d'où le résultat souhaité.

5. Comme la variance dépend de t, la série n'est pas stationnaire.

   Note : Bien que ce résultat puisse surprendre, il est à noter ici que la série commence à $x_0 = w_0$, alors que dans le cas des séries définies dans le cours, la série "ne commence pas" : on l'observe uniquement à partir de $t = 0$, ou $t = 1$, mais en réalité elle existe déjà depuis une infinité de pas de temps.

6. Lorsque $t \to \infty$, on a $\mathrm{Var}(x_t) = \sigma^2 \frac{1-\phi^{2(t+1)}}{1-\phi^2} \to \sigma^2 \frac{1}{1-\phi^2}$ (car $|\phi| < 1$). La variance est alors asymptotiquement constante, et la fonction d'autocovariance est asymptotiquement indépendante par translation. On en déduit que, en un sens, le processus devient stationnaire lorsque $t \to \infty$.

7. Comme la série devient asymptotiquement stationnaire, on peut simuler N observations de la série, avec $N > n$, puis retirer les $N - n$ premières observations. Lorsque N est grand, alors les n dernières observations de la série seront presque stationnaires. Plus précisément : *(i)* on génère N variables gaussiennes v_t i.i.d. suivant la loi normale $\mathcal{N}(0, 1)$; *(ii)* on multiplie ces variables par σ, $w_t = \sigma v_t$, pour obtenir des réalisations d'une loi normale $\mathcal{N}(0, \sigma^2)$; *(iii)* puis on utilise soit la formule de récurrence de l'AR(1), soit celle établie en question 1 pour simuler des observations x_t du modèle ; *(iv)* enfin, on retire les $N - n$ premières observations et on ne garde que les n dernières.

   Note : cette technique de simulation est en pratique souvent utilisée. Les $N - n$ observations qui sont retirées sont appelées réalisations de *burn-in* (on parle aussi de période de *burn-in*).

8. En reprenant la démonstration de la question 1, on obtient que $x_t = \phi^t x_0 + \sum_{j=1}^{t} \phi^j w_{t-j}$. On peut désormais reprendre le calcul de la variance de la série dans ce cas, en notant que x_0 est indépendant des w_j, pour $j \geq 1$:
$$\begin{aligned}\mathrm{Var}(x_t) &= \phi^{2t}\mathrm{Var}(x_0) + \sum_{j=1}^{t} \phi^{2j}\sigma^2 \\ &= \sigma^2 \frac{\phi^{2t}}{1-\phi^2} + \sigma^2 \frac{1-\phi^{2t}}{1-\phi^2} \\ &= \sigma^2 \frac{1}{1-\phi^2}.\end{aligned}$$
On remarque que cette fois, la variance est constante. La série est donc stationnaire.

   Note : lorsque $x_0 = w_0/\sqrt{1-\phi^2}$, cela donne à x_0 la même distribution que si la série avait commencé depuis une infinité de pas de temps. Cela permet alors de simuler x_0 comme si la série existait dans le passé, ce qui confère à l'ensemble de la série sa stationnité.

Exercice 4

Motivation On met en évidence que le pic d'autocorrélation d'un MA(1) ne dépasse pas 1/2 en amplitude.

Pour un MA(1), $x_t = w_t + \theta w_{t-1}$, montrer que $|\rho(1)| \leq 1/2$ pour toute valeur de θ. Pour quelle valeur de θ est-ce que $\rho(1)$ atteint son maximum ? son minimum ?

Solution

On calcule la variance et la fonction d'autocovariance de la série :

$$\gamma(0) = \mathrm{Var}(w_t + \theta w_{t-1}) = \sigma^2 + \theta^2\sigma^2 = (1+\theta^2)\sigma^2$$
$$\gamma(1) = \mathrm{Cov}(w_t + \theta w_{t-1}, w_{t-1} + \theta w_{t-2}) = \theta\sigma^2.$$

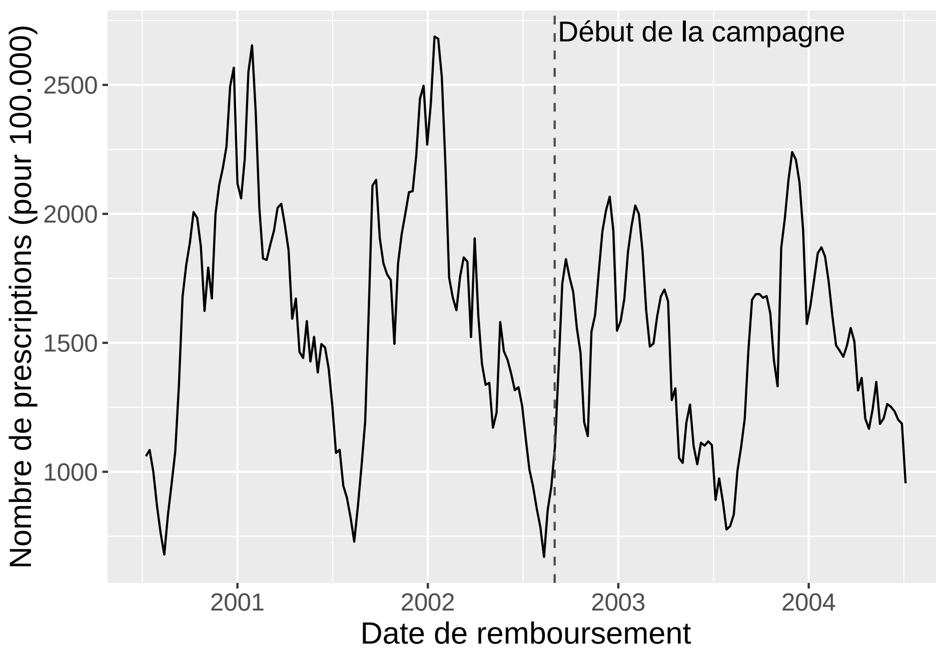
Il vient ensuite que

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta}{1+\theta^2}.$$

La fonction $\theta \mapsto \theta/(1+\theta^2)$ est impaire, vaut 0 en $\theta = 0$ et lorsque $\theta \to \infty$. Par ailleurs, sa dérivée est $\theta \mapsto (1-\theta)^2/(1+\theta^2)^2$, est positive sur $[-1, 1]$ et négative ailleurs. On en déduit que $\theta \mapsto \theta/(1+\theta^2)$ est décroissante sur $[-\infty, -1]$, croissante sur $[-1, 1]$ puis décroissante sur $[1, +\infty]$, et donc qu'elle atteint son minimum en $\theta = -1$, auquel cas elle vaut $-1/2$, et son maximum en $\theta = 1$, auquel cas elle vaut $1/2$.

Partie 2 : analyse de modèles ARIMA

Motivation On cherche à étudier l'impact de la campagne de sensibilisation "Les antibiotiques, c'est pas automatique !", débutée en 2002 et reconduite chaque hiver jusqu'en 2005. Pour cela, on dispose de la série hebdomadaire du nombre de prescriptions de bêta-lactamines et macrolides au niveau métropolitain (pour 100 000 habitants), et on cherche à déterminer de combien le nombre de prescriptions a diminué entre la période pré-campagne, et la période post-campagne.



Pour étudier l'évolution du nombre de prescriptions antibiotiques, on propose d'effectuer une analyse statistique de la série en ajustant un modèle de régression périodique de la forme :

$$x_t = \beta_0 + \beta_0' 1_{\{t>\tau\}} + (\beta_1 + \beta_1' 1_{\{t>\tau\}}) \cos\left(\frac{2\pi t}{52}\right) + (\beta_2 + \beta_2' 1_{\{t>\tau\}}) \sin\left(\frac{2\pi t}{52}\right) + \varepsilon_t,$$

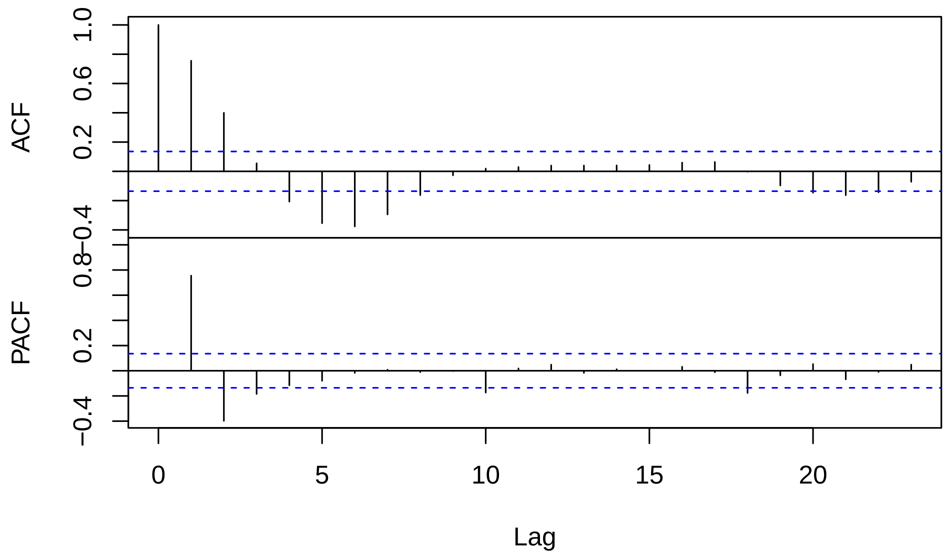
où les β_i et β_i' désignent des paramètres du modèle, τ la date de début de la campagne de sensibilisation, et (ε_t) un bruit gaussien.

1. Dans le modèle, les termes associés à $\cos(2\pi t/52)$ et $\sin(2\pi t/52)$ permettent de modéliser la saisonnalité de la série. Proposer une autre méthode qui aurait pu permettre de retirer la saisonnalité de la série. Quel défaut peut présenter cette autre méthode ?

2. * Comment interpréter les paramètres β'_0, β'_1 et β'_2 ?

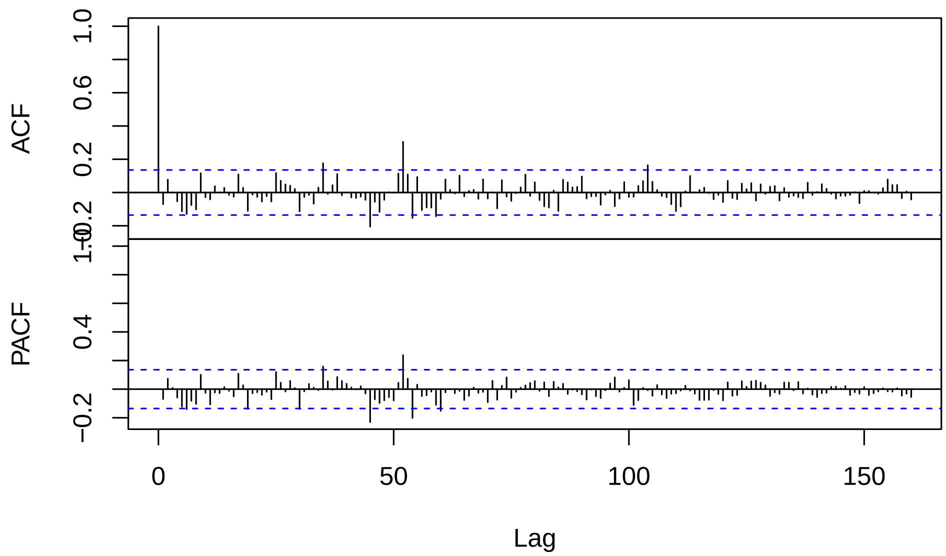
3. Dans un premier temps, on ajuste le modèle présenté plus haut sans imposer de structure de dépendance sur le bruit ε_t. Les fonctions d'autocorrélation et d'autocorrélation partielle des résidus du modèle sont données sur la figure suivante. Comment doit-être ajusté le modèle pour tenir compte de la structure de dépendance du bruit ?

Series x



4. On ajoute au modèle la structure de dépendance identifiée. Le modèle est estimé de nouveau, et les fonctions d'autocorrélation et d'autocorrélation partielle correspondantes sont données par la figure suivante. Un test de Ljung-Box est effectué pour déterminer si les résidus du modèle sont décorrélés.
   a. Rappeler l'hypothèse nulle testée.
   b. En s'aidant du test effectué sous R, déterminer combien de décalages ont été testés.
   c. A l'aide des figures et du test de Ljung-Box, déterminer si la structure de dépendance doit être ajustée et, le cas échéant, comment l'ajuster.

Series x



```
##
##  Box-Ljung test
##
## data:  model$residuals
## X-squared = 121.12, df = 58, p-value = 2.371e-06
```

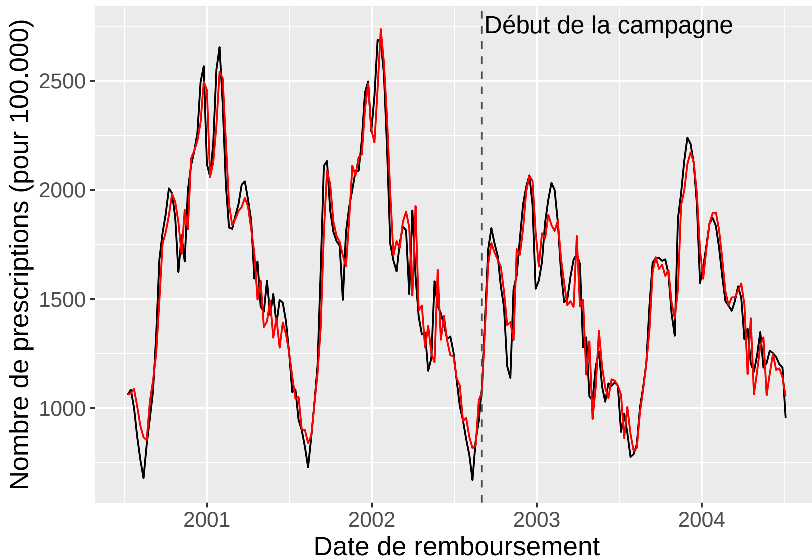
5. Au final, deux modèles sont retenus, dont les sorties R sont données ci-dessous.
   a. Comment choisir lequel des deux modèles retenir ?
   b. Ecrire l'équation de récurrence vérifié par le bruit ε_t du modèle retenu.

```
##
## Call:
## arima(x = presc, order = c(2, 0, 0), seasonal = list(order = c(1, 0, 0), period = 52),
##     xreg = cbind(cos, sin, cst_post, cos_post, sin_post))
##
## Coefficients:
##          ar1      ar2    sar1  intercept        cos       sin   cst_post
##       0.9272  -0.2580  0.4387  1676.4594  -609.1529  -75.8045  -214.6476
## s.e.  0.0717   0.0724  0.0717    46.5319    64.9084   65.4769    52.5279
##       cos_post  sin_post
##       187.0338   89.5180
## s.e.   75.0853   72.5441
##
## sigma^2 estimated as 15402:  log likelihood = -1310.19,  aic = 2640.39
```

```
##
## Call:
## arima(x = presc, order = c(2, 0, 0), seasonal = list(order = c(0, 0, 1), period = 52),
##     xreg = cbind(cos, sin, cst_post, cos_post, sin_post))
##
## Coefficients:
##          ar1      ar2    sma1  intercept        cos       sin   cst_post
##       0.9835  -0.3147  0.3305  1673.4115  -612.9778  -81.4485  -211.5181
## s.e.  0.0677   0.0687  0.0727    42.9291    60.0755   60.8501    56.1242
##       cos_post  sin_post
##       193.9411  102.7965
## s.e.   80.4549   77.7429
##
## sigma^2 estimated as 16483:  log likelihood = -1314.78,  aic = 2649.57
```

6. À supposer que la diminution du nombre de prescriptions peut être attribuée à la campagne de sensibilisation, donner un intervalle de confiance sur le nombre de prescriptions annuelles qui ont pu être ainsi évitées.

Remarque : la figure ci-dessous représente graphiquement l'ajustement du modèle retenu (en rouge par rapport à la série initiale représentée en noir).



Solution

1. Une différentiation d'ordre 52 (puisque la série est hebdomadaire) aurait permis de retirer la saisonnalité de la série. Cependant, la série des variations posséderait alors 52 pas de temps de moins que la série originelle, puisque les variations ne peuvent être calculées pour les 52 premiers pas de temps.

2. Ces paramètres correspondent à des modifications des paramètres β_0, β_1 et β_2 après que la campagne de sensibilisation a commencé. Ils peuvent donc être interprétés comme les différences entre les périodes pré- et post-campagne, sur le niveau moyen (pour β_0') et sur la structure périodique (pour β_1' et β_2') de la série.

3. L'autocorrélation du bruit est non nulle pour des décalages faibles, et décroit rapidement vers 0. L'autocorrélation partielle est non nulle pour les décalages 1 et 2, puis s'annule. Il faut donc ajouter au modèle de régression linéaire une structure d'autodépendance de type AR(2).

4. a. Dans le test de Ljung-Box, l'hypothèse nulle testée est la suivante :

$$\mathcal{H}_0 : \quad \forall h = 1, \ldots, H, \quad \rho(h) = \mathrm{Corr}(x_t, x_{t-h}) = 0,$$

   où H détermine le nombre de décalages testés et est choisi par l'utilisateur.

   b. La statistique du test de Ljung-Box suit une loi du chi-deux à $H - p - q - P - Q$ degrés de liberté, où p, q, P et Q désignent les ordres des processus AR et MA estimés dans la structure de dépendance. Ici, la sortie R donne `df = 58`, et le modèle ajusté sur le bruit est un AR(2). On en déduit que 60 décalages ont été testés.

   c. Pour un seuil α donné (par exemple 0.01), le test de Ljung-Box rejette l'hypothèse nulle. Par ailleurs, l'ACF et la PACF révèlent un pic d'autocorrélation non nul à l'ordre 52 et il existe peut-être un pic non nul à l'ordre 104 sur l'ACF. Ces observations incitent à ajuster le modèle en ajoutant une structure résiduelle saisonnière au bruit, du type $AR(1)_{52}$, de sorte que le modèle de dépendance sur le bruit ε_t est un $\mathrm{ARIMA}(2,0,0)(1,0,0)_{52}$.

5. a. Si les fonctions d'autocorrélation des deux modèles ne présentent plus de pics significativement différents de 0, alors le modèle retenu est déterminé par l'AIC, qui est une mesure de l'ajustement du modèle, pénalisée par le nombre de paramètres. Ici, l'AIC du premier modèle est inférieure de 9 points. On retiendra donc ce modèle.

b. Le modèle retenu pour ε_t est un $\mathrm{ARIMA}(2,0,0)(1,0,0)_{52}$ qui vérifie la formule de récurrence

$$(1 - \varphi_1 B^{52})(1 - \phi_1 B - \phi_2 B^2)\varepsilon_t = w_t,$$

où B désigne l'opérateur retard et w_t un bruit blanc gaussien.

6. Pour déterminer l'impact de la campagne de sensibilisation, il faut s'intéresser au paramètre β_0' qui rend compte de la diminution de l'espérance de la série (il existe aussi une diminution de l'amplitude des termes périodiques, mais à l'échelle d'une année, ces termes périodiques se compensent). Comme β_0' est estimé par maximum de vraisemblance, la loi limite de l'estimateur est donnée par

$$\frac{\widehat{\beta}_0' - \beta_0'}{\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_0')}} \sim \mathcal{N}(0,1).$$

On en déduit donc un intervalle de confiance asymptotique au niveau $1-\alpha$ pour β_0' (voir note de bas de page[1]) :

$$\mathrm{IC}_{1-\alpha}(\beta_0') = \left[\widehat{\beta}_0' \pm u_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_0')}\right],$$

où u_γ désigne le quantile d'ordre γ de la loi normale centrée réduite. Avec $\alpha = 0.05$, et les sorties du modèle retenu, on trouve

$$\mathrm{IC}_{0.95}(\beta_0') = [-318; -112]\,.$$

C'est-à-dire, à l'échelle d'une année, la campagne de sensibilisation a permis d'éviter $-52\widehat{\beta}_0' = 11\ 162$ prescriptions antibiotiques pour 100 000, avec un intervalle de confiance donné par

$$\mathrm{IC}_{0.95}(-52\beta_0') = [5\ 808; 16\ 515]\,.$$

[1] Cet intervalle de confiance est construit de manière usuelle : par conséquence de la loi limite de l'estimateur, on a, lorsque le nombre d'observations tend vers l'infini,

$$\mathbb{P}\left(-u_{1-\alpha/2} \leq \frac{\widehat{\beta}_0' - \beta_0'}{\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_0')}} \leq u_{1-\alpha/2}\right) \to 1-\alpha.$$

En manipulant l'inégalité à l'intérieur de la probabilité, on obtient :

$$\mathbb{P}\left(\widehat{\beta}_0' - u_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_0')} \leq \beta_0' \leq \widehat{\beta}_0' + u_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_0')}\right) \to 1-\alpha,$$

ce qui correspond à l'intervalle de confiance asymptotique souhaité.