Introduction à l'analyse des séries temporelles

TP2

Cette seconde séance de travaux pratiques s'intéresse à l'analyse jointe de deux séries temporelles, et la réalisation de modèles de régression linéaire avec erreurs autocorrélées. Les commandes suivantes seront abordées :

- `ccf(series1, series2, lag.max = H)` permet d'afficher la fonction de covariance croisée entre deux séries temporelles, et donc d'identifier si leur corrélation présente un décalage temporelle.
- `arima.sim(model = list(ar = c(phi1, phi2), ma = c(ma1, ma2, ma3)), n)` permet de simuler un modèle ARMA ; par exemple ici, un modèle ARMA(2,3).
- `arima(series, order = c(p,d,q), seasonal = list(order = c(P,D,Q), period = S), xreg = cbind(X1, X2, X3))` : l'argument `xreg` permet de spécifier des variables explicatives. Par exemple, la commande donnée en exemple revient à estimer le modèle de régression linéaire avec erreurs autocorrélées suivant :
$$\text{series} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_t,$$
où ε_t est un bruit gaussien autocorrélé, modélisé par un processus $\text{ARIMA}(p,d,q)(P,D,Q)_S$.
- `predict(model, n.ahead = m, se.fit = TRUE, newxreg = cbind(X1, X2, X3))` : l'argument `newxreg` permet de spécifier les valeurs des variables explicatives pour les prédictions.

Partie 1 : étude de la fonction de corrélation croisée

On s'intéresse dans un premier temps au comportement de la fonction de corrélation croisée (*Cross Covariance Function*, CCF). Formellement, la CCF entre deux séries *stationnaires* x_t et y_t s'écrit

$$\rho_{xy}(h) = \text{Corr}(x_{t+h}, y_t) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}},$$

où $\gamma_{xy}(h) = \text{Cov}(x_{t+h}, y_t)$ désigne la fonction de covariance croisée entre les deux séries.

1. Simulez deux bruits blancs gaussiens (fonction `rnorm`). Tracez la CCF entre les deux séries. Commentez la figure obtenue.

2. Simulez deux bruits blancs gaussiens de longeur différente. En déduire comment sont calculées les bornes de l'intervalle de confiance (lignes pointillées bleues) affichées sur la CCF ?

3. Simulez la série $x_t = w_t$, avec w_t un bruit blanc gaussien. Simulez ensuite une série dépendante de la première, de la forme $y_t = x_{t-5} + w'_t$, avec w'_t un autre bruit blanc gaussien. Tracez et commentez la CCF entre les deux séries.

4. Simulez maintenant deux processus AR(3), avec $\phi_1 = 0.9$, $\phi_2 = -0.4$ et $\phi_3 = -0.3$. Tracez la CCF entre les deux séries. Répétez ces deux étapes plusieurs fois. Que remarquez-vous ? Les deux séries sont-elles indépendantes ?

La question précédente souligne le fait que l'analyse de la CCF n'est pas pertinente si les séries présentent des structures d'autodépendance. En particulier, deux séries saisonnières sont presque toujours corrélées. Il faut s'appuyer sur les résultats de la proposition suivante pour avancer.

Proposition La distribution asymptotique de l'estimateur $\widehat{\rho}_{xy}(h)$ de la corrélation croisée est une loi normale centrée, d'écart-type

$$\sigma_{\widehat{\rho}_{xy}} = \frac{1}{\sqrt{n}},$$

si au moins l'une des séries est un bruit blanc indépendant.

Pour étudier la CCF entre deux séries, on s'appliquera donc d'abord à retirer la structure d'autodépendance d'une des deux séries en la modélisant par un processus ARIMA, avant de tracer la CCF entre les résidus de ce modèle et l'autre série. Cette étape de modélisation ARIMA préalable à l'étude de la CCF s'appelle *blanchiment* (parfois *préblanchiment*) des données.

5. En reprenant les séries simulées de la question 3, ajustez un modèle ARIMA(3,0,0) sur une seule des deux séries. Tracez ensuite la CCF entre les résidus du modèle et l'autre série simulée. Commentez.

Partie 2 : étude des facteurs écologiques sur le recrutement de poissons

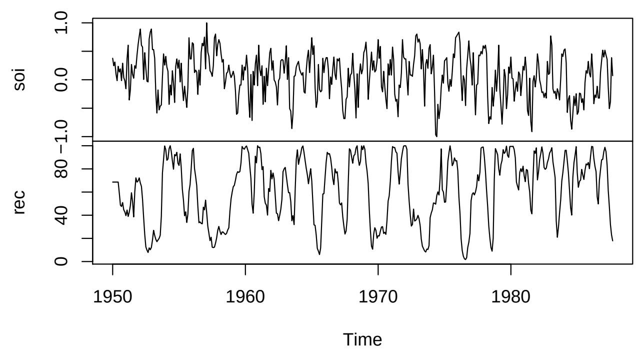
```
# Librairie à télécharger pour obtenir les données
install.packages("astsa")
library(astsa)
```

On dispose de deux séries temporelles issues d'une étude écologique, sur une période de 453 mois entre 1950 et 1987 :

- `soi` (pour *Southern Oscillation Index*) mesure les changements de pression atmosphérique associés aux températures en surface dans l'Océan Pacifique central.
- `rec` mesure le recrutement (le nombre de nouveaux poissons) dans la même région.

On souhaiterait étudier l'influence des changements de température sur la reproduction des poissons.

```
plot.ts(cbind(soi, rec), main = "")
```



1. Etude de la CCF entre `soi` et `rec`.
   a. Tracez naïvement la CCF entre les deux séries.
   b. Blanchissez la série `rec`.
   c. Tracez maintenant la CCF entre la série `rec` blanchie, et la série `soi`.
   d. Identifiez s'il existe une association entre les deux séries, et le cas échéant si cette association est décalée dans le temps.
2. Analyse de l'association entre `soi` et `rec`.
   a. À partir du décalage identifié, créez la série décalée pour `soi`, et la série correspondante pour `rec`.
   b. Effectuez une régression linéaire avec erreurs autocorrélées de la série `rec` sur `soi`.
   c. Ecrivez explicitement le modèle, et les hypothèses du modèle.
   d. Comparez les résultats du modèle à ceux que vous auriez obtenus si la structure d'autodépendance de la série `rec` n'avait pas été prise en compte.
   e. Répondez à la problématique écologique posée.
3. Prédiction de la série `rec`.
   a. Récupérez les éléments du vecteur `soi` qui vont permettre d'effectuer les prédictions futures de la série `rec`.
   b. Effectuez et tracez les prédictions de la série `rec`.
   c. Y a-t-il une probabilité que le nombre de nouveaux poissons tombe à 0 dans les mois à venir ?