Séries chronologiques
M1 Statistique et économétrie
TD 4 - encore un modèle autorégressif
V. Monbet

Application directe du cours

1. Simuler le processus AR(2) suivant

$$X_t = 0.9X_{t-1} - 0.3X_{t-2} + 0.1\epsilon_t$$

   pour t variant de 1 à 5000 et avec ϵ un bruit blanc gaussaient centré et réduit.
2. Tracer la série temporelle obtenue. Tracer une estimation de la densité de probabilité de X_t correspondant à la première moitié de la série temporelle puis à la deuxième moitié. Que peut-on dire de ces graphiques ?
3. Estimer la moyenne puis la fonction d'autocovariance du processus X. Commentez. Comment peux-t-on calculer la fonction d'autocovariance théorique ?
4. Estimer la fonction d'autocorrection partielle du processus X_t. Commentez les résultats.
5. Estimer les paramètres du modèle en utilisant les estimateurs de Yule-Walker (fonction `ar` + option). Donner un intervalle de confiance pour les paramètres.

Exercice 1 : décès liés à des maladies respiratoires

On considère la série temporelle `ldeaths` disponible sous `R` qui décrit le nombre mensuel de décès liés à des maladies respiratoires au Royaume-Uni sur la période 1974-1979.

1. La commande `plot(ldeaths)` permet de tracer la série temporelle. Peut-on identifier une composante saisonnière ? Peut-on identifier une tendance ?
2. Proposer, en la justifiant, une méthode permettant de rendre la série temporelle stationnaire.
3. Peut-on supposer que la série temporelle stationnarisée est la réalisation d'un bruit blanc gaussien ? Qu'est-ce que cela signifie concrètement ?

Exercice 2 : volumes de pèche

Il s'agit de proposer un modèle pour la série des volumes de pèche (`poisson.dat`) qui permette de décrire les données puis de faire de la prévision.

1. Doit-on choisir un modèle additif ou multiplicatif ? Pourquoi ? Comment procède t'on dans ce cas ?
2. Tracer la série temporelle du logarithme des volumes de pèche. On la notera par exemple `xl`.
3. Etude de la tendance
   (a) Modéliser la tendance du logarithme des volumes de pèche. On notera `yl` la série sans tendance.
   (b) Proposer un test paramétrique permettant de vérifier que la série de contient plus de tendance linéaire additive.
   (c) Proposer un test non paramétrique permettant de vérifier que la série ne contient plus de tendance monotone additive. On pourra par exemple se baser sur un test du signe et comparer $yl[t]$ à $yl[t + T/2]$ avec T la longueur de la série temporelle.
4. Modéliser la saisonnalité du logarithme des volumes de pèche par un modèle paramétrique. On notera zl la série résiduelle et on supposera qu'elle est stationnaire. Il existe sous `R` des tests de stationnarité. Le plus utilisé est sans doute celui de la fonction `adf.test` (Augmented Dickey-Fuller Test) du package `tseries`.
   L'hypothèse nulle de ce test est que la série observée vient d'un modèle admettant une racine de module 1 (ie non stationnaire).
5. Pour les questions, précédentes, avec le packages `TSA`, on peut écrire

```
library(TSA)
P = read.table("Poissons.dat")
x = ts(P[,2],start = 1971,freq=12)
xl = log(x)
har.=harmonic(xl,1)
xreg = cbind(seq(1971,1980,by=1/12)[1:length(x)],har.)
m5.xl=arima(xl,order=c(0,0,0),xreg=xreg)
plot(m5.xl$residuals)
newhar.=harmonic(ts(rep(1,24), start=c(1980,1),freq=12),1)
newxreg = cbind(seq(1980,1982,by=1/12)[1:24],newhar.)
plot(m5.xl,n.ahead=24,newxreg=newxreg,
   type='b',ylab='Volumes de pèches',xlab='Year')
```

6. Étude de la série stationnaire zl.

   (a) Tracer la fonction d'autocorrelation de la série zl ainsi que le nuage de points de $zl(t)$ en fonction de $zl(t-1)$. Peut-on supposer que la série temporelle stationnarisée est la réalisation d'un bruit blanc gaussien ? Pourquoi ?

   (b) Tracer la fonction d'autocorrelation partielle de la série zl. Que pouvez vous en conclure ?

   (c) Ajuster des modèles $AR(k)$ pour k allant de 1 à 4 en utilisant la fonction `arima`. Quel estimateur est utilisé ?

   (d) Sélection de modèle - Calculer les critères d'Akaike (AIC) et Bayesian Information Criterion (BIC) pour les 4 modèles. Qu'en concluez-vous ? Ecrire le modèle. On rappelle que

   $$AIC = -2\log(\mathcal{L}) + 2N_{par}$$
   $$BIC = -2\log(\mathcal{L}) + N_{par}\log(T)$$

   avec T la longueur de la série temporelles, N_{par} le nombre de paramètres dans le modèle et $\mathcal{L}$ la vraisemblance.

   (e) Ré-estimer les paramètres du meilleur modèle. Vous pourrez ensuite utiliser la fonction `tsdiag` pour tracer la fonction d'autocovariance des résidus.

   (f) Validation du modèle par validation croisée. Utiliser les années 1971 à 1979 pour ajuster les paramètres du modèle. Prédire les données de l'année 1980, prédiction à un pas de temps puis deux pas de temps.
   On peut utiliser la fonction `predict.Arima`.
   Vous pourrez travailler dans un premier temps avec les logarithmes des volumes de pèches puis revenir aux volumes réels. Estimer la variance des erreurs. Discuter les résultats obtenus.

   (g) Comparer les résultats obtenus ci-dessous avec ce qu'on obtient avec le lissage exponentiel. Commenter.