Université Paris I, Panthéon - Sorbonne

Master M.A.E.F. 2016 – 2017

TP de Séries Temporelles n^0 3 : Identification du bruit - Processus ARMA

Dans ce TP, on travaille sur la partie bruit (aléatoire) d'une série temporelle, et plus particulièrement quand ce bruit est un processus ARMA, pour identifier sa loi, estimer sa covariance.

Identification du bruit

	Indépendance du bruit
x=rnorm(50)	On simule un bruit blanc gaussien.
r=acf(x)	S'affiche le corrélogramme de x.
	Celui-ci est le calcul des covariances empiriques divisées par la variance de x
	Les traits verticaux représentent ces corrélations empiriques pour des longueurs (lag) de taille $0, 1, 2, ...$ Les traits horizontaux pointillés représentent l'intervalle de confiance à 95% de $Z/\sqrt{n}$, où Z est une v.a. $\mathcal{N}(0, 1)$ et n la longueur de x. Lorsque le coefficient de corrélation est dans cet intervalle on pourra le considérer comme nul. La série x est-elle un bruit blanc?
y=x[1:49]+x[2:50]	On génère une nouvelle série y. Théoriquement quelle est la distribution de y?
	Est-ce un bruit blanc? Calculer le coefficient de corrélation de longueur 0 et 1.
ry=acf(x)	Corrélogramme de y. A-t-on un bruit blanc?
rho=ry$acf[1]	Calcul du coefficient de corrélation empirique de longueur 1.
	Le comparer avec le résultat théorique.
	Tests d'adéquation d'un bruit blanc
x=5*rnorm(50)-2	Quelle est la loi de x?
hist(x,nclass=11)	Histogramme à 11 classes (première méthode d'estimation de la loi de x).
qqnorm(x)	Test par la droite de Henry. Conclusion ?
y=3*runif(100)+1	Quelle est la loi de y?
qqnorm(y)	Conclusion?
dn=ks.test(x,"pnorm",-2,5)	Test de Kolmogorov-Smirnov. On compare x à une loi $\mathcal{N}(-2, 5^2)$.
	Le résultat intéressant fourni est la $p - value$. D'une manière générale, l'hypothèse H_0 d'un test pourra être acceptée lorsque $p - value > 0.05$.
ddn=ks.test(y,"pexp",3)	Test sur y. À quoi peut-on s'attendre ?
	Tester maintenant que y suit une loi uniforme en précisant les paramètres de cette loi.
ks.test(x,y)	Permet de tester si x et y suivent une même loi. Conclusion?

Processus ARMA

Dans la suite, on simule certains processus ARMA, puis on estime les différents paramètres l'identifiant (paramètres qui sont les coefficients des 2 polynômes P et Q) et son ordre.

	Simulation automatique d'un ARMA
X=arima.sim(100,model=list(ar=-.3,ma=.7))	On simule un processus ARMA. Ecrire sous forme récurrente le processus simulé. Tracer la trajectoire ainsi simulée.
	Représenter son corrélogramme.
	Simuler un ARMA[2,2] (choisir les coefficients).
	Simuler avec un bruit uniforme sur $[-1, 1]$.

Estimation des paramètres d'un ARMA

ord=c(1,0,1)	Ordre supposé de l'ARMA dont on estime les paramètres.
Coeff=arima(X,ord)	On estime les paramètres de l'ARMA avec un estimateur de maximum de vraisemblance Pour cela, on doit donner l'ordre supposé de l'ARMA (ord). Quelle est la variance estimée du bruit blanc ? Faire l'estimation avec un ordre plus grand. Résultats ? Augmenter la taille de la trajectoire et recommencer. Conclusions ?
Coeff$var.coef	Matrice de covariance des estimateurs des paramètres. Que se passe-t-il lorsque la taille de la trajectoire augmente ?

Corrélogrammes

	Simuler une trajectoire de taille 100, $X100$, du processus ARMA $X = (X_n)_{n\in\mathbb{Z}}$ tel que : $(1-0.4B).(1-0.5B)X = \varepsilon$, où B est l'opérateur retard et $\varepsilon = (\varepsilon)_{n\in\mathbb{Z}}$ est un bruit blanc gaussien de variance 2.
acf(X100)	Corrélogramme de $X100$. Déterminer le corrélogramme théorique de $X100$. Comparer.
pacf(X100)	Fonction de corrélation partielle empirique de $X100$. (le coefficient de corrélation partielle de X d'ordre j est le coefficient $\rho_j(j)$ de X_{n-j} obtenu quand on fait la régression linéaire de X_n par $(X_{n-1},..,X_{n-j})$. Dans le cas d'un AR, les coefficients de corrélation partielle s'identifient avec les coefficients de l'AR Augmenter la taille de la trajectoire et recommencer. Conclusions ?

Estimation des paramètres et de l'ordre

	Estimer les paramètres de $X100$ pour une trajectoire de taille 100, (taper $Coeff = arima(...)$) en précisant un ordre $(2,0,0)$ (le 0 du milieu correspond à un paramètre distinguant les ARIMA, que l'on prendra par la suite toujours égal à 0). Vérifier la convergence et donner la matrice des estimateurs des paramètres.
tsdiag(Coeff)	Reconstruit le bruit blanc à partir des coefficients estimés, c'est-à-dire que si $P(B).X = Q(B).\varepsilon$, l'estimation par MLE permet d'obtenir $\widehat{P}$ et $\widehat{Q}$, et donc $\dfrac{\widehat{P}(B)}{\widehat{Q}(B)}X100 = \widehat{\varepsilon}100$. On vérifie alors si le bruit blanc "reconstitué" $\widehat{\varepsilon}100$ est bien un bruit blanc. La convergence de la procédure *arima.mle*, la "blancheur" de $\widehat{\varepsilon}100$ et surtout permettent de vérifier si l'ordre choisi était le bon.
Box.test(Coeff$res,lag=10)	On peut faire un test de type portemanteau pour tester l'indépendance des résidus (l'hypothèse H_0 est celle de l'indépendance $->$ bon ordre=p.value>0.05) On veut faire le test avec différente valeur de l'option lag, à choisir > 5 si possible Recommencer ces procédures en choisissant pour ordre $(1,0,0)$, $(0,0,1)$, $(0,0,2)$, $(2,0,1)$, $(3,0,3)$. Noter la valeur du AIC. Conclusions ?
Z=autoarmafit(X100)	Procédure automatique d'estimation de l'ordre (AIC...) et des coefficients
Z$best.order	
Z$best.model	Conclusions?

———————————————————————-

Estimation des paramètres et de l'ordre à partir des corrélations

———————————————————————-

On travaille à partir de $X100$ déjà simulé et on estime les paramètres et l'ordre.

MA=acf(X100) — À partir du corrélogramme, on peut estimer l'ordre d'un MA modélisant $X100$ par l'entier k tel que les coefficients de corrélation $\rho(j)$,$j = 0, 1, .., k$ sont considérés comme non nuls.

AR=pacf(X100) — À partir du corrélogramme partiel, on peut estimer l'ordre d'un AR modélisant $X100$ par l'entier k tel que les k premiers coefficients de corrélation partielle sont considérés comme non nuls. Conclusions ? Recommencer avec la taille 1000.

ParamAR=ar(X100) — Cette fois-ci, on fait comme si la série $X100$ était un AR et on estime les coefficients de cet AR à partir des équations de Yule-Walker (utilisant les corrélations empiriques). Que trouvez-vous? (ordre, coefficients)?
Considérer la série des résidus ($ParamAR\$res$), c'est-à-dire le bruit blanc reconstitué à partir de $X100$ et des coefficients estimés. Est-ce un bruit blanc?
Quelle estimation des paramètres donne les meilleurs résultats (plusieurs trajectoires...) ?

Exercice

Simuler une trajectoire de taille 500 du processus $Y = (Y_n)_{n\in\mathbb{Z}}$ tel que $Y_{n+1} = -0.8Y_n + \varepsilon_n - 0.9\varepsilon_{n-1}$, avec $\varepsilon = (\varepsilon)_{n\in\mathbb{Z}}$ est un bruit blanc gaussien de variance 1. Quel type de processus est Y ? Estimer l'ordre de ce processus à partir des procédures indiquées précédemment. Conclusions ?

Prédiction pour des processus ARMA

————————————————————

Prédictions pour un processus ARMA

————————————————————

Considérer Z comportant les 490 premières valeurs de Y précédemment simulée. Estimer les coefficients par maximum de vraisemblance approchée (taper $ParaZ = arima(...)$), avec l'ordre $(1, 0, 1)$.
On veut prévoir les 10 données suivantes de Z à partir du modèle estimé précédent.

pZ=predict(ParaZ,n.ahead=10) — On a ainsi estimé les 10 dernières données de Y à partir des 490 premières.

pZ$pred — Valeurs prévues pour des 10 données.

pZ$se — Ecarts-type des estimateurs de ces 10 données.
Quelle est la loi de $\widehat{Y}_{491}$?
Comparer les 10 données prévues avec les 10 valeurs réelles. Conclusion?
Expliquer la convergence des valeurs prédites.
Faire la même chose en modélisant Y par un AR (remplacer la commande $ParaZ$ par $ParaAR$). A-t-on amélioré la prédiction?