Projet de Séries temporelles, M1 IES-D3S
pour le 26 janvier 2024

Enseignant: Bertail Patrice

1 Exercice 1

Soit $(\varepsilon_t)_{t\in Z}$ un bruit blanc faible de variance σ^2(gaussien de variance 1 dans les simulations) A et B désigne respectivement deux variables aléatoires indépendantes de même loi (par exemple normale pour les simulations) indépendantes de $(\varepsilon_t)_{t\in Z}$, de variance finie.

Les processus $(X_t)_{t\in N}$ suivants admettent-ils une solution stationnaire au second ordre? Pour cela on calculera l'espérance et la fonction d'autocovariance si possible. Les mettre sous forme canonique quand c'est possible. Si le processus n'est pas stationnaire, proposer une méthode pour le rendre stationnaire. Représenter trois trajectoires de taille $n = 200$ de chaque processus sur un même graphique (si possible avec des couleurs différentes).

a) $X_t = \sin(t) + \varepsilon_t - 0.2 * \varepsilon_{t-1}$

b) $X_t = \varepsilon_t - \varepsilon_{t-1}$

c) $X_t = A\cos(\omega t) + B\sin(\omega t)$ pour ω fixé dans $[0, 2\pi]$, avec $EA = 0$ $EB = 0$. Discuter la stationnarité de ce processus (et on fera deux représentations graphiques différentes selon qu'il est stationnaire ou non).

d) $X_t = 2X_{t-1} - X_{t-2} + \varepsilon_t$

e) $X_t - 4X_{t-1} = \varepsilon_t - 0.25\varepsilon_{t-1}$

2 Exercice 2

Chaque groupe (d'au plus 3 personnes) doit étudier une série de prix choisie parmi p1,p2,p3,p4,...,p9. Ces prix (dit unitaires car moyennés sur une zone géographique) sont dans le fichier prix.R qu'il faut sourcer sous R. Il s'agit de données de prix hebdomadaires de produits à la consommation (les données sont anonymisées car payantes sinon), observées sur 10 ans, avec 52 semaines par années (une correction pour les années bisextiles avec 53 semaines à été appliquées pour se ramener à 52) soit au total 520 données.

a) Donner une représentation graphique du processus. Vous parait-il stationnaire? Que voit-on juste en regardant le graphique?

b) Donner les fonctions d'autocorrélation acf et pacf du processus. Donner une estimateur de la densité spectrale du processus. Proposer une décomposition du signal en trend , saisonalité avec la fonction *decompose* de R. Qu'en déduisez vous?

c) On suppose d'abord que le processus a une composante périodique déterministe de la forme

$$X_t = m + A\sin(2\pi t/T) + B\cos(2\pi t/T) + Y_t.$$

En choissisant bien T (compte-tenu des observations et du problème considéré), déterminer par une régression linéaire la valeur de m, A et B. on se ramène désormais à l'étude des résidus centrés estimés $Z_t = \widehat{Y}_t$. Tester la stationarité de Z_t.

d) Tracer l'acf et le pacf de Z_t. Est-ce un bruit blanc? En déduire un ensemble de modèles raisonnables pour Z_t Les estimer, tester leur validité. Choisir le meilleur modèle.

e) A partir du modèle retenu, prédire les valeurs du processus X_t pour le mois de janvier suivant (4 valeurs).

f) On cherche plutôt à estimer un modèle SARIMA. Que vous donne comme choix la procédure automatique *auto.arima* du package *forecast*? en déduire les prédictions du mois de janvier avec ce modèle (on pourra utiliser la fonction *forecast*) .

g) Appliquer un opérateur $(1 - B^{52})$ aux données X_t et proposer un modèle adéquat de type ARMA. En déduire les prédictions du mois de janvier avec ce modèle et les intervalles de confiance associés.

h) Quel modèle vous semble le plus pertinent : c) suivi de e) ou f) ou g)? Il y a t'il une grosse différence en terme de prédictions?

i) Simuler sous R, pour $n = 500$ le processus

$$(1 - B^{52})(1 - 2B + B^2)(1 - 2B)X_t = \varepsilon_t - \varepsilon_{t-1}, t = 1, ..n,$$

que l'on aura pris soin de mettre sous sa forme la plus réduite, en faisant l'hypothèse que les résidus sont i.i.d. gaussiens $N(0, 1)$. Tracer pour chaque simulation la série originale des X_t, la série differenciée $X_t - X_{t-1}$ et la série

$$Y_t = (1 - B)(1 - B^{52})X_t, \ t = 1, ..., n$$

(dans sa forme canonique).Que constatez vous? X_t est-elle stationnaire? Y_t?

j) donner les formes $AR(\infty)$ et $MA(\infty)$ du processus Y_t.

k) Pour la série Y_t, estimer les modèles $AR(p)$, $MA(q)$, $ARMA(p,q)$ admissibles pour la réalisation (éventuellement par la méthode du coin) et les estimer. A partir des diagnostiques sous R, déterminer les ordres et le meilleur modèle pour la série observée. Estimer les prévisions à l'ordre 1 et 2 et les intervalles de confiances correspondants.

l) Estimer les prévisions à l'ordre 1 et 2 i.e. Y_{501}, Y_{502} pour 999 simulations de Y_t en supposant qu'il s'agit d'un AR(1). En déduire les moyennes des prévisions et leurs écart-types (calculées sur les répétitions). On tracera également la distribution empirique des prévisions.Que concluer vous?