ECM 3A-DDEFi 2024-2025

Statistique et décision TP3 [1]

Objectifs :

1. *Comprendre et analyser les propriétés des séries chronolgiques.*
2. *Prédiction à partir de données réelles à l'aide des séries chronolgiques.*

Analyse de séries temporelles via simulation numérique

Nous allons simuler des données afin d'illustrer les résultats du cours sur les séries temporelles. A l'aide des simulations on chercher à mieux comprendre la qualité et l'efficacité des méthodes d'estimation et des tests statistiques vus en cours.
1) Soit $(Z_n)_{n\in\mathbb{Z}}$ un bruit blanc gaussien de variance $\sigma^2 = 0,01$.

1. Simuler une trajectoire de $n = 100$ points suivant le modèle ARMA(1,2) suivant:
$$X_n = -0,5X_{n-1} + Z_n + 0,1Z_{n-1} - 0,3Z_{n-2}. \quad (1)$$
2. A l'aide d'un test statistique vérifier la stationnarité des données (test Augmented Dickey-Fuller, Box-Pierce). Commenter vos résultats.
3. A l'aide de la fonction *auto.arima* du logiciel *R*, estimer les paramètres du modèle $ARMA(p, q)$ à savoir, p,q, σ et les coefficients du modèle ARMA.
4. Tracer une estimation de la fonction d'auto-corrélation et de l'auto-corrélation partielle.
5. Que peut-on dire des résultats de l'estimation?
6. A l'aide de la fonction *tsdiag* observer les résidus du modèle. Commenter vos résultats tout en s'appuyant sur les résultats du test Ljung-Box, Breusch-Pagan, Durbin-Watson, Shapiro-Wilk, Breusch-Godfrey.

2) Simuler 1000 trajectoires suivant le modèle défini dans l'équation (1) et enregistrer les données.

1. Estimer les paramètres du modèle correspondant à chaque trajectoire avec la fonction *auto.arima*.
2. Donner le nombre total de fois où le paramètre p est estimé à 1?
3. Donner le nombre total de fois où le paramètre q est estimé à 2?
4. Pour les cas où la fonction *auto.arima* estime $p = 1$ et $q = 2$ calculer l'erreur moyenne quadratique de l'estimateur des paramètres σ, ϕ_1, θ_1 et θ_2. Donner le boxplot associé chacun des estimateurs. Interpréter les résultats.
5. Pour un cas où la fonction *auto.arima* estime $p > 1$ ou $q > 2$ estimer les paramètres du modèle avec la fonction *arima* et tester la significativité des paramètres estimés.
6. Pour un cas où la fonction *auto.arima* estime $p < 1$ ou $q < 2$ estimer les paramètres du modèle et appliquer un test Ljung and Box pour étudier les résidus. Commenter vos résultats.

[1] Version du 3/10/2022

3) Considérons les données associées à la consommation énergétique des appareils ménagers dans une maison basse consommation. Les données que nous allons étudier ont été recueillies dans une maison située à Stambrudges (Belgique). L'énergie (kWh) utilisée par les appareils ménagers à été mesurée toutes les 10 min entre le 11 janvier 2016, 17h et le 27 mai 2016, 18h. Toutes les données sont accessibles librement sur le site UCI Machine Learning Repository
([http://archive.ics.uci.edu/dataset/374/appliances+energy+prediction]) Nous ne nous intéresserons qu'à la seconde colonne (Appliances) des données importées.

1. Tracer la série obtenue. Qu'en pensez-vous ?
2. La fonction *window* permet d'extraire les données correspondant à une certaine plage de temps. Tracer les données sur un mois ou sur une semaine.
3. Une saisonnalité à la fois journalière et mensuelle est identifiable. Tout d'abord considérons l'estimation de la saisonnalité s_t. Soit le modèle suivant pour la saisonnalité:

   $$s_t = a_0 + a_1 \cos(2\pi t/168) + a_2 \sin(2\pi t/168) + a_3 \cos(2\pi t/24) + a_4 \sin(2\pi t/24)$$

   où le temps t est mesuré en heures (NB : une semaine compte $24 \times 7 = 168$ heures).

   (a) Trouver les coefficients a_0, a_1, a_2, a_3 et a_4 en applicant la commande lm aux données "Appliances" ou au logarithme des données ou à la puissance α des données ($\alpha = \pm 0,5$ ou $\alpha = \pm 1/3$).

   (b) Tracer dans une même fenêtre : la série originale, les valeurs estimées de s_t et les résidus. (Utiliser les commande $\cdots$\$fitted.values et $\cdots$\$ residuals).
4. Considérons la série transformée $\log(X_t)$ ou X_t^α. (avec $\alpha = \pm 0,5$ ou $\alpha = \pm 1/3$ par exemple). Tout d'abord ramenons nous à l'étude d'une série stationnaire en utilisant la différenciation via la fonction $diff$ sur une fenêtre de 24h.
5. Etudier la stationnarité de la série. Etudier si le bruit est gaussien. À la suite de cette étape, on pourra éventuellement être amené à changer la transformation de la série initiale.
6. Observer les fonctions d'autocorrelation et d'autocorrelation partielle et proposer un modèle de type ARMA(p,q) avec p et q le plus petit possible.
7. A l'aide de la fonction *arima* estimer les paramètres du modèle proposé dans dans la question précédente.
8. De manière générale, quel modèle ARIMA propose la fonction *auto.arima*?
9. En se basant sur des tests statistiques choisir le modèle le plus adapté.
10. À l'aide de la fonction $forecast$ du package forecast, proposer des prédictions pour la série transformée.