

Séries Chronologiques

Université d'Angers – M2 Data Science

Guide pour le TP 3

L'objectif du TP 3 est de se familiariser avec les processus ARMA, le but étant d'acquérir une bonne connaissance empirique des notions résumées ci-dessous :

- 1) Comprendre la structure des ARMA en étant capable de les simuler.
- 2) Observer le comportement empirique de leurs ACF/PACF, en particulier dans les cas AR et MA purs.
- 3) Estimer les paramètres du modèle et tester leur significativité.
- 4) Faire un diagnostic complet du résidu obtenu et valider ou invalider le modèle.

Une fois maîtrisés, ces outils devront, à terme, être mis en pratique dans le contexte plus général des SARIMA. Ces derniers constitueront le socle du projet du module de Séries Chronologiques qui sera abordé lors du TP 4 et se poursuivra sur les dernières séances de TP.

Point 1) Quelques idées :

- Programmer une fonction `simulerARMA(n, m, Phi, Theta, s2)` qui simule un processus ARMA de taille `n`, centré en `m`, de coefficients de la partie autorégressive `Phi` (dont la dimension donne p), de coefficients de la partie moyenne mobile `Theta` (dont la dimension donne q), et dont le bruit est gaussien de variance `s2`.
- Cette fonction représente graphiquement (en soignant les détails) la série simulée puis la renvoie sous la forme d'un vecteur `X` de taille `n`.
- On pourra compléter la compréhension de ce point par l'étude de la fonction `arima.sim` qui fait le travail que vous venez de programmer, ou encore la structure de données `ts`.
- Simuler quelques processus ARMA stationnaires mais très proches de la non stationnarité (par exemple $X_t = 0.999X_{t-1} + \varepsilon_t$) et observer graphiquement leur comportement.

Point 2) Quelques idées :

- Écrire une fonction `tracerCorr(X, lag)` qui prend en entrée une série `X` et représente graphiquement l'ACF et la PACF empiriques, jusqu'au décalage `lag`. La fenêtre graphique sera séparée en deux parties : à gauche l'ACF, à droite la PACF.
- Simuler des MA purs (`Phi = c()`) et observer les ACF/PACF empiriques. Comparer avec la théorie.
- Simuler des AR purs (`Theta = c()`) et observer les ACF/PACF empiriques. Comparer avec la théorie.

Point 3) Quelques idées :

- Manipuler la fonction `arima` pour estimer les paramètres et interpréter toutes les sorties.
- Récupérer les résidus, en déduire les valeurs reconstruites (les "fitted values" ou "one-step") et superposer la série avec sa reconstruction par le modèle.
- Écrire une fonction `significativite(ARMA, alpha)` qui prend en entrée un objet ARMA issu de la fonction `arima` et affiche TRUE ou FALSE pour chaque paramètre, selon qu'il est considéré comme significatif ou non, au risque `alpha`. Il s'agit donc simplement de comparer les rapports absolus $\frac{|\text{estimations}|}{\text{écarts-types}}$ avec $u_{1-\frac{\alpha}{2}}$.

Point 4) Quelques idées :

- Récupérer les résidus de la modélisation ARMA issus de `arma`.
- Écrire une fonction `outilsNormalite(Res)` qui prend en entrée les résidus et affiche dans une même fenêtre trois graphiques : un histogramme, un QQ-plot et un nuage de points standardisé muni des bornes ± 1.96 .
- Écrire une fonction `outilsBlancheur(Res, lag)` qui prend en entrée les résidus et affiche dans une même fenêtre trois graphiques : les ACF/PACF jusqu'au lag `lag` et le nuage de points de l'instant t en fonction de l'instant $t - 1$.
- Interpréter visuellement les sorties sur plusieurs simulations.
- Interpréter les tests de Shapiro et de Ljung-Box (ou de Box-Pierce) pour un lag suffisant (3 ou 4 par exemple).
- Tenter d'autres simulations avec des innovations non gaussiennes, pour voir si tout se passe toujours aussi bien dans le diagnostic de normalité...