

Feuille d'exercices no 5

Préliminaires

Créer un fichier texte dans lequel vous répondrez clairement aux questions ci-dessous, en incluant vos codes R, les résultats obtenus sous R (graphique y compris), vos interprétations, remarques ... Une fois ce TP fini, vous mettrez en forme votre compte-rendu et l'exporterez au format pdf. C'est ce qui sera demandé au partiel.

1 Simulation des processus $ARMA$

Penser à représenter chaque processus simulé.

1. Donner la définition d'un processus $ARMA_{p,q}$. Rappeler les conditions sur les coefficients pour que ce processus soit stationnaire.
2. À l'aide de la fonction `arima.sim`, simuler plusieurs processus AR_p et MA_q (p et q pas trop grands). Avant toute simulation, écrire la définition mathématique du processus à simuler et veillez à ce que les conditions de stationnarité soient respectées.
3. Observer les auto-corrélations empiriques (partielles ou non). Que constatez-vous?
4. Simuler quelques $ARMA_{p,q}$, observer et interpréter les auto-corrélations empiriques (partielles ou non).
5. Faire de même avec un modèle $ARIMA_{p,d,q}$, avec un d assez petit.

2 Identification d'un processus $ARMA$

Récupérer le fichier de données `serie1.dat`¹.

1. Ce processus vous semble-t-il modélisable par un processus $ARMA$? Pourquoi?
2. On travaille désormais avec la série obtenue en appliquant la fonction `diff` à la série. Quelle transformation a-t-on effectuée?
3. En observant les auto-corrélations empiriques et auto-corrélations partielles empiriques, proposer des modèles AR_p ou MA_q d'ordre faible pour modéliser cette série.
4. Estimer les paramètres du modèle sélectionné.
5. Tester la blancheur des résidus.
6. Conclure pour choisir un modèle.
7. S'inspirer de la démarche pour modéliser la série `serie2.dat`.

1. Toujours à l'adresse : <http://math.unice.fr/~rubentha/enseignement/>

3 Prévision dans un processus *ARMA*

Soit le processus $X_t - X_{t-1} + \frac{1}{2}X_{t-2} - \frac{1}{3}X_{t-3} = \epsilon_t$ avec ϵ_t un bruit blanc gaussien centré réduit.

1. Après avoir identifié ce processus, simuler 50 réalisations de longueur 105.
2. Pour chaque simulation, extraire les 100 premières valeurs et estimer les paramètres d'un AR_3 .
3. Pour chaque simulation, prédire les cinq valeurs suivantes.
4. Donner une estimation de l'erreur moyenne (biais) et de la variance de l'erreur de prévision à 1, 2, 3, 4 et 5 pas.
5. Recommencer en rallongeant la durée d'observation, et comparer aux résultats précédents.

4 Précipitations mensuelles à San Fransisco entre 1932 et 1966

Récupérer la série `sanfran.dat`.

1. La série semble-t-elle stationnaire? Si non faites en sorte qu'elle le soit.
2. Proposer alors un AR_p adapté. Valider votre modélisation en testant les résidus. Mettez votre idée de coté, car nous vous proposons de tester maintenant directement sur la série initiale un AR_2 avec partie saisonnière (autrement dit un $SARIMA_{2,0,0,12}$). On utilise désormais les données jusqu'à la fin de l'année 1963.
3. Estimer le modèle $SARIMA_{2,0,12}$. Afficher et tester les résidus.
4. Prévoir les précipitations de 1964, 1965 et 1966 à partir de ce modèle. Superposer sur un graphique prévision et valeurs réelles.
5. Faire de même avec le modèle AR_p sélectionné à la première question (ainsi qu'avec un lissage exponentiel de Holt-Winters, avec et sans composante saisonnière).
6. Quel est, graphiquement, le meilleur modèle pour prédire cette série?
7. Comment aurait-on pu répondre à la question précédente de façon moins subjective? Comparer les résultats à l'analyse graphique.

5 Taux d'intérêt au Royaume-Uni

Le fichier `UKinterestrates.dat` contient le spread des taux d'intérêts (différence entre taux d'intérêt à long terme et à court terme) pour le Royaume-Uni entre mars 1953 et décembre 1995. En vous inspirant des exercices précédents, proposer une modélisation de type *ARMA*, *ARIMA* ou *SARIMA*. Justifier bien votre démarche.

Appendice : mise en œuvre sous R

Auto-corrélation partielle : utiliser la fonction `pacf`.

La fonction `arima.sim(modele,n)` permet de simuler un processus $ARMA_{p,q}$ défini par

$$X_t - \sum_{k=1}^p a_k X_{t-k} = \epsilon_t + \sum_{j=1}^q b_j \epsilon_{t-j}.$$

Les paramètres a_k et b_j du processus sont précisés dans le paramètre `modele` de la fonction : `modele<-list(ar=c(a1, . . . , ap),ma=c(b1, . . . , bq))` (choisir `ma` ou `ar` pour un processus *MA* ou *AR*, si $p = 1$, écrire `ar=a1`). Pour simuler un modèle $ARIMA_{p,d,q}$ il faut ajouter le composant `order=c(p,d,q)` dans le paramètre `modele` de la fonction `arima.sim`.

La fonction `ar` permet d'estimer les paramètres d'un processus AR_p :
`out<-ar(data,aic=TRUE,order.max=NULL)`.

L'ordre p du processus auto-régressif est choisi (inférieur à `order.max`) à l'aide du critère AIC (si l'option `aic` est validée). La fonction `arima` permet d'estimer les paramètres :

- d'un $ARMA_{p,q}$: `out<-arima(serie,order=c(p,0,q))`
- d'un $ARIMA_{p,d,q}$: `out<-arima(serie,order=c(p,d,q))`
- d'un $SARIMA_{p,d,q,T}$:

`out<-arima(serie,order=c(p,d,q),seasonal=list(order=c(P,D,Q),period=T))`.

Les paramètres P, D, Q du modèle $SARIMA$ ne sont pas abordés dans ce cours, nous leur donnerons par défaut la valeur des paramètres p, d, q (pas trop grands). Parmi les sorties de cette fonction, on peut obtenir :

- `out$coef` : estimation des coefficients,
- `out$aic` : valeur du critère AIC,
- `out$resid` : estimation des résidus.

La fonction `p=predict(out,h)` permet d'effectuer une prévision à l'horizon h . Parmi les sorties de cette fonction, `p$pred` contient les prévisions, et `p$se` contient l'écart-type de l'erreur de prévision. Il n'existe pas de fonction prédéfinie pour calculer un intervalle de confiance