Université de Strasbourg
M2 - Statistique
Séries Temporelles

Année 2011/2012
Ségolen Geffray
geffray@math.unistra.fr

TP: simulation de séries temporelles et étude de données réelles

Il est nécessaire d'installer les packages `R` *suivants:* `tseries`, `forecast`, `urca`, `lmtest` *et* `TSA`, *afin d'effectuer les exercices de cette feuille. Un récapitulatif des fonctions utiles de* `R` *se trouve à la fin.*

Exercice 1.
Dans chacun des cas suivants, simuler des données selon le modèle imposé. En occultant ensuite cette information, ajuster un ou plusieurs modèles (soit de type AR/MA/ARMA/ARIMA/SARIMA, soit de type régression avec tendance et/ou saisonnalité déterministe et résidus issus d'un bruit blanc ou d'un processus AR/MA/ARMA) sur la série simulée en prêtant attention à la phase exploratoire (représentation de la série initiale, de la série différenciée ou de la série résiduelle, du corrélogramme et du corrélogramme partiel,...) et aux diagnostics à effectuer (tests de blancheur, de stationnarité, d'hétéroscédasticité,...). A chaque fois, investiguer les effets d'une tentative de stationnarisation par différentiation et par régression.
Les séries à simuler sont de longueur $n = 105$. Ajuster le modèle sur les 100 premières observations puis utiliser le modèle ajusté pour faire de la prévision pour les 5 dernières valeurs. Comparer les résultats obtenus avec une prévision obtenue par la méthode de Holt-Winters. Recommencer avec $n = 510$ (resp 1050): ajuster le modèle sur les 500 (resp 1000) premières observations puis utiliser le modèle ajusté pour faire de la prévision pour les 10 (resp 50) dernières valeurs.

1. Simuler une série temporelle selon un modèle MA(1) de paramètre θ_1 égal à -0.9, puis à -0.5, puis à 0.5 et enfin à 0.9 et de paramètre σ^2 égal à 1 puis 10. Utiliser (ε_t) bruit blanc gaussien de variance égale à 1 puis 10, puis utiliser (ε_t) bruit blanc fort de loi de Student à 2 puis 5 degrés de liberté et enfin utiliser (ε_t) bruit blanc fort de loi de exponentielle de paramètre 1 puis 0.05.

2. Simuler une série temporelle selon un modèle MA(1) de paramètre θ_1 égal à 1.01, puis à 1.02 et enfin à 1.05 et de paramètre σ^2 égal à 1 puis 10. Utiliser (ε_t) bruit blanc gaussien de variance égale à 1 puis 10, puis utiliser (ε_t) bruit blanc fort de loi de Student à 2 puis 5 degrés de liberté et enfin utiliser (ε_t) bruit blanc fort de loi de exponentielle de paramètre 1 puis 0.05.

3. Simuler une série temporelle selon un modèle AR(1) de paramètre ϕ_1 égal à -0.9, puis à -0.5, puis à 0.5 et enfin à 0.9 et de paramètre σ^2 égal à 1 puis 10. Utiliser (ε_t) bruit blanc gaussien de variance égale à 1 puis 10, puis utiliser (ε_t) bruit blanc fort de loi de Student à 2 puis 5 degrés de liberté et enfin utiliser (ε_t) bruit blanc fort de loi de exponentielle de paramètre 1 puis 0.05.

4. Simuler une série temporelle selon un modèle AR(1) de paramètre ϕ_1 égal à 1.01, puis à 1.02 et enfin à 1.05 et de paramètre σ^2 égal à 1 puis 10. Utiliser (ε_t) bruit blanc gaussien de variance égale à 1 puis 10, puis utiliser (ε_t) bruit blanc fort de loi de Student à 2 puis 5 degrés de liberté et enfin utiliser (ε_t) bruit blanc fort de loi de exponentielle de paramètre 1 puis 0.05.

5. Simuler une série temporelle selon le modèle

$$X_t = \frac{5}{6}X_{t-1} - \frac{1}{6}X_{t-2} + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

6. Simuler une série temporelle selon le modèle

$$X_t = \frac{3}{2}X_{t-1} - \frac{1}{2}X_{t-2} + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

7. Simuler une série temporelle selon un modèle AR(2) de paramètres (ϕ_1, ϕ_2) égaux à (5/6,-1/6), puis à (5/6,-0.9) et enfin à (-1.12,-0.5) et de paramètre σ^2 égal à 1 puis 10.

8. Simuler une série temporelle selon un modèle ARMA(1,2) de paramètres $(\phi_1, \theta_1, \theta_2)$ égaux à (0.2,-0.8,0.016) et de paramètre σ^2 égal à 1 puis 10.

9. Simuler une série temporelle selon un modèle ARIMA(1,1,2) de paramètres $(\phi_1, \theta_1, \theta_2)$ égaux à (0.2,-0.8,0.016) et de paramètre σ^2 égal à 1 puis 10.

10. Simuler une série temporelle selon un modèle $\text{SARIMA}(1,1,2) \times (1,0,1)_{12}$ de paramètres $(\phi_1, \theta_1, \theta_2)$ égaux à () et de paramètre σ^2 égal à 1 puis 10.

11. Simuler une série temporelle selon un modèle $\text{SARIMA}(0,1,1) \times (0,1,1)_{12}$ de paramètres $(\theta_1, \Theta_)$ égaux à (-0.2,0.5) et de paramètre σ^2 égal à 1 puis 10.

12. Simuler une série temporelle selon un modèle $\text{SARIMA}(1,0,1) \times (0,1,1)_{12}$ de paramètres $(\phi_1, \Phi_1, \theta_1, \theta_2, \Theta_1)$ égaux à (0.2,0.4,-0.8,0.016,-0.5) et de paramètre σ^2 égal à 1 puis 10.

13. Simuler une série temporelle selon un modèle de marche aléatoire sans dérive, de paramètre σ^2 égal à 1 puis 10.

14. Simuler une série temporelle selon un modèle de marche aléatoire avec dérive, de paramètre de dérive égal à 4 et de paramètre σ^2 égal à 1 puis 10.

15. Simuler une série temporelle selon le modèle

$$X_t = t + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

16. Simuler une série temporelle selon le modèle

$$X_t = t.\exp(\varepsilon_t)$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

17. Simuler une série temporelle selon le modèle

$$X_t = 10\cos\left(\frac{\pi t}{6}\right) + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

18. Simuler une série temporelle selon le modèle

$$X_t = 10\cos\left(\frac{\pi t}{6}\right) + t + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

19. Simuler une série temporelle selon le modèle

$$X_t = 10\cos\left(\frac{\pi t}{6}\right) t + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

20. Simuler une série temporelle selon le modèle

$$X_t = 10\cos\left(\frac{\pi t}{6}\right) t \exp(\varepsilon_t)$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

21. Simuler une série temporelle selon le modèle

$$X_t = 10\cos\left(\frac{\pi t}{6}\right) + 4\sin\left(\frac{\pi t}{6}\right) + t + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

22. Simuler une série temporelle selon le modèle

$$X_t = 10\cos\left(\frac{\pi t}{6}\right) + 4\sin\left(\frac{\pi t}{3}\right) + t + \varepsilon_t$$

avec (ε_t) bruit blanc de variance égale à 1 puis 10.

23. Simuler une série temporelle selon le modèle

$$X_t = 50 + 3t + \varepsilon_t$$

avec (ε_t) processus AR(1) donné par

$$\varepsilon_t = 0.8\varepsilon_{t-1} + w_t$$

avec (w_t) bruit blanc de variance égale à 1 puis 10.

24. Simuler une série temporelle selon le modèle

$$X_t = 70 + 2t - 3t^2 + \varepsilon_t$$

avec (ε_t) processus AR(1) donné par

$$\varepsilon_t = 0.5\varepsilon_{t-1} + w_t$$

avec (w_t) bruit blanc de variance égale à 1 puis 10.

25. Simuler une série temporelle selon le modèle

$$X_t = 0.2X_{t-1} + \varepsilon_t$$

avec les (ε_t) sont des v.a.i.i.d. de loi $\mathcal{N}(0, (1+t)\sigma^2)$ avec $\sigma^2 = 1$ puis $\sigma^2 = 10$.

26. Simuler une série temporelle selon le modèle
$$X_t = X_{t-1} + \varepsilon_t$$
avec les (ε_t) sont des v.a.i.i.d. de loi $\mathcal{N}(0, (1+t)\sigma^2)$ avec $\sigma^2 = 1$ puis $\sigma^2 = 10$.
27. Simuler une série temporelle selon le modèle
$$X_t = 0.2\varepsilon_{t-1} + \varepsilon_t$$
avec les (ε_t) sont des v.a.i.i.d. de loi $\mathcal{N}(0, (1+t)\sigma^2)$ avec $\sigma^2 = 1$ puis $\sigma^2 = 10$.
28. Simuler une série temporelle selon le modèle
$$X_t = \begin{cases} \mu_1 + \sigma_1\varepsilon_t & \text{si } X_{t-1} > 0 \\ \mu_2 + \sigma_2\varepsilon_t & \text{si } X_{t-1} \leq 0 \end{cases}$$
avec les (ε_t) sont des v.a.i.i.d. de loi $\mathcal{N}(0, (1+t)\sigma^2)$ avec $\sigma^2 = 1$ et $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = (1, -1, 1, 1,)$ puis (1,1,1,4) puis (1,-1,1,4).
29. Simuler une série temporelle selon le modèle
$$X_t = \frac{\exp(\lambda_t\varepsilon_t - \lambda_t^2/2) - 1}{\sqrt{\exp(\lambda_t^2) - 1}}$$
où les ε_t sont i.i.d. de loi $\mathcal{N}(0, 1)$ et où $\lambda_t = \delta_0 + \delta_1\lambda_{t-1} + w_t$ où les w_t sont i.i.d. de loi $\mathcal{N}(0, 1)$. Utiliser $(\delta_0, \delta_1) = (1, 0.3)$ puis (1,0.5) puis (1,0.7) et enfin (1,0.9).

Exercice 2.
Dans chacun des cas suivants, importer les données dans l'espace de travail. Ajuster un ou plusieurs modèles (soit de type AR/MA/ARMA/ARIMA/SARIMA, soit de type régression avec tendance et/ou saisonnalité déterministe et résidus issus d'un bruit blanc ou d'un processus AR/MA/ARMA) sur la série simulée en prêtant attention à la phase exploratoire (représentation de la série initiale, de la série différenciée ou de la série résiduelle, du corrélogramme et du corrélogramme partiel,...) et aux diagnostics à effectuer (tests de blancheur, de stationnarité, d'hétéroscédasticité,...). A chaque fois, investiguer les effets d'une tentative de stationnarisation par différentiation et par régression.
Ajuster le modèle sur toutes les observations sauf quelques-unes puis utiliser le modèle ajusté pour faire de la prévision pour ces quelques dernières valeurs. Comparer les résultats obtenus avec une prévision obtenue par la méthode de Holt-Winters.

1. Le jeu de données `AirPassengers` est disponible dans `R`. Il contient le nombre de réservations (en milliers) sur les vols internationaux de la compagnie américaine Pan Am pour la période 1949-1960.
2. Le jeu de données `LakeHuron` est disponible dans `R`.
3. Le jeu de données `lynx` est disponible dans `R`.
4. Le jeu de données `co2` est disponible dans `R`. Il s'agit de données de concentration atmosphérique en CO_2 recueillie de 1959 à 1997.

5. Le jeu de données `global` est disponible sur internet à l'adresse suivante:
   http://www.massey.ac.nz/ pscowper/ts/global.dat
   Il contient un indicateur global de la température planétaire mensuelle moyenne calculé depuis 1856.

6. Le jeu de données `wave` est disponible sur internet à l'adresse suivante:
   http://www.massey.ac.nz/ pscowper/ts/wave.dat
   Il contient des données relatives à la hauteur des vagues par rapport au niveau de l'eau au repos dans une piscine munie d'un simulateur de vagues en mm. L'intervalle de temps est $0.1s$ et la durée de la série est $39.7s$.

7. Le jeu de données `Fontdsdt` est disponible sur internet à l'adresse suivante:
   http://www.massey.ac.nz/ pscowper/ts/Fontdsdt.dat
   Il contient des données relatives au débit mensuel en m^3/s d'une réserve d'eau potable entre janvier 1909 et décembre 1980.

8. Le jeu de données `Maine` est disponible sur internet à l'adresse suivante:
   http://www.massey.ac.nz/ pscowper/ts/Maine.dat
   Il contient le taux de chômage mensuel dans l'état américain du Maine de janvier 1996 à âout 2006.

9. Le jeu de données `USunemp` est disponible sur internet à l'adresse suivante:
   http://www.massey.ac.nz/ pscowper/ts/USunemp.dat
   Il contient le taux de chômage mensuel des USA de janvier 1996 à octobre 2006.

10. Le jeu de données `pounds_nz` est disponible sur internet à l'adresse suivante:
    http://www.massey.ac.nz/ pscowper/ts/pounds_nz.dat
    Il contient les taux de change trimestriels de la livre anglaise en dollars néo-zélandais entre 1991 et le 3ème trimestre de l'année 2000.

11. Le jeu de données `cbe` est disponible sur internet à l'adresse suivante:
    http://www.massey.ac.nz/ pscowper/ts/cbe.dat
    Il contient des données relatives à la production australienne mensuelle d'électricité (en millions de kWh), de bière (en millions de L) et de chocolat (en tonnes) pour la période 1958-1990.

Remarque: Pour simuler " à la main" une série de longueur n issue d'un processus autorégressif d'ordre p de paramètres $\phi_1, ..., \phi_p$ et de variance des innovations σ^2, on peut procéder de la façon suivante:

- fixer p valeurs initiales arbitraires $x_1, ..., x_p$,
- générer des innovations i.i.d. $(\varepsilon_t)_{t=1,...,n+n_0}$,
- calculer récursivement les valeurs:

$$X_t = \sum_{t=1}^{p} \phi_k X_{t-k} + \varepsilon_t \, ,$$

- éliminer les n_0 premières valeurs ainsi générées.

En pratique, les ordres de grandeur sont $n = 1000$ et $n_0 = 200, 300$ ou 500 selon la distance dans le plan complexe des racines du polynôme de la composante AR et le cercle unité.

`read.table` / `scan`	importe les données
`attach`	permet d'utiliser le nom des variables en colonnes
`str`	donne la structure d'un jeu de données
`tsp`	donne le début, la fin et la fréquence d'une série temporelle
`ts`	produit un objet `R` de type "time series"
`window`	extrait un sous-ensemble d'une série
`time`	extrait le temps d'une série
`cycle`	retourne la saison pour chaque valeur d'une série
`aggregate`	créée une série aggrégée sur une période de façon à ôter l'effet saisonnier
`ts.plot`	trace une ou plusieurs séries
`decompose`	décompose la série en tendance, saisonnalité et bruit en utilisant des moyennes mobiles
`stl`	décompose une séries en utilisant un lissage `loess` (régression par polynômes locaux)
`filter`	filtre une série temporelle
`HoltWinters`	met en oeuvre la méthode de Holt-Winters (lissage exponentiel simple, H-W avec tendance ou non, avec saisonnalité ou non)
`predict`	fonction qui prend en entrée un modèle ajusté ou une méthode et de nouvelles données, rangés un `data.frame` pour obtenir des prédictions
`set.seed`	fixe un germe pour le générateur de nombres aléatoires de façon à ce que les simulations soient reproductibles
`rnorm`	simule des variables aléatoires de loi gaussienne
`arima.sim`	simule des données selon un modèle de type ARIMA(p,d,q) (éventuellement, p et/ou d et/ou q peuvent être nul(s))
`ARMAacf`	calcul des fonctions théoriques acf d'un processus arma (stationnaire)
`polyroot`	extrait les racines du polynôme caractéristique
`diff`	créée une série différenciée (on peut choisir l'ordre "`differences`" et l'horizon "`lag`")
`ndiffs`	estime le nombre de différentiations nécessaires pour rendre une série stationnaire
`nsdiffs`	estime le nombre de différentiations saisonnières nécessaires pour rendre une série stationnaire
`lm`	ajuste un modèle de régression linéaire par moindres carrés généralisés
`gls`	ajuste un modèle de régression linéaire par moindres carrés ordinaires
`coef`	extrait les estimateurs des paramètres du modèle ajusté
`resid`	extrait les résidus d'un modèle ajusté
`confint`	fournit un IC pour les paramètres du modèle ajusté
`factor`	retourne une variable discrète ou catégorielle sous la forme d'une indicatrice
`acf`	renvoie le corrélogramme (simple)
`pacf`	renvoie le corrélogramme partiel
`eacf`	renvoie la matrice du corrélogramme étendu (méthode du coin du triangle)
`ar`	ajuste un modèle de type autorégressif
`arima`	ajuste un modèle de type ARIMA(p,d,q) aux données (éventuellement, p et/ou d et/ou q peuvent être nul(s))
`tsdiag`	fournit quelques diagnostics après utilisation de arima
`dwtest`	effectue un test d'autocorrélation de Durbin-Watson
`bptest`	effectue un test d'hétéroscédasticité de Breush-Pagan
`bdstest`	effectue un test de Brock-Dechert-Scheinkman d'indépendance
`Box.test`	effectue un test de Box-Pierce ou Ljung-Box
`ur.df`=`adf.test`	effectue un test de Dickey-Fuller
`pp.test` = `ur.pp`	effectue un test de Phillips-Perron
`ur.ers`	effectue un test de Elliott-Rothenberg-Stock
`kpss.test` = `ur.kpss`	effectue un test de Kwiatkowski-Phillips-Schmidt-Shin