Séries chronologiques
sept. 2010
TD 1

Modélisation des moyennes mensuelles de la température en France

On considère une série temporelle qui décrit les moyennes mensuelles de la température en France entre 1901 et 2000.

1. Importation des données

Télécharger les données (au format ascii) à l'adresse :
`http://www-labsticc.univ-ubs.fr/~monbet/doc/data_temp.dat`
Importer ces données sous `R` puis créer une variable `x` de type *time series* qui contient les données à l'aide de la fonction `ts`.

```
x=ts(as.numeric(temp),frequency = 12, start = 1901)
```

2. Visualisation

Représenter graphiquement la série temporelle. Peut-on identifier une tendance et/ou une composante saisonnière ?

3. Etude de la tendance

(a) Moyennes mobiles. On propose dans cette question d'estimer la tendance par la méthode des moyennes mobiles. On rappelle que si $\{x_i\}_{i\in\{1,\cdots,n\}}$ est une série temporelle, alors la moyenne mobile d'ordre p associée est la série temporelle définie pour $t \in \{p+1,\cdots,n-p\}$ par

$$\hat{x}_t = \sum_{i=t-p}^{p} x_i$$

avec $2p+1$ la largeur de la fenêtre.

1. Calculer les séries temporelles obtenues par la méthode des moyennes mobiles avec des largeurs de fenêtre de 1 an, 2 ans, 5 ans puis 10 ans à l'aide de la fonction `filter`. Pourquoi choisit-on des largeurs de fenêtre qui sont des multiples de la période de la composante saisonnière (1 an ici) pour estimer la tendance?

2. Tracer sur une même figure la série temporelle initiale et ces moyennes mobiles. Discuter les résultats obtenus : quelle largeur de fenêtre vous semble être la mieux adaptée? Peut-on identifier une tendance dans la série temporelle résiduelle?
3. A t'on modélisé une tendance additive ou multiplicative? Ecrire le modèle en notant X_t le processus moyenne mensuelle de la température et T_t la tendance.

(b) Ajustement d'un modèle paramétrique. On propose dans cette question d'utiliser un modèle de tendance polynomiale.

1. Quel degré de polynôme la question précédente suggère t'elle d'utiliser?
2. Ajuster le modèle à l'aide de la fonction `lm`.
3. Tracer sur une même figure la série temporelle et la tendance estimée. Discuter la qualité de l'ajustement. Donner une expression analytique du modèle en définissant rigoureusement vos notations.
4. Quelle est le degré de signification (p-value) du test associé à l'hypothèse selon laquelle il n'y a pas de tendance dans la série temporelle? Sur quelles hypothèses reposent la validité de ce test? Ces hypothèses sont-elle vérifiées?
5. Calculer la série temporelle des moyennes annuelles de la température. Représenter graphiquement cette série et ajuster le modèle de tendance polynomiale. Que devient le degré de signification du test? Discuter les résultats.

(c) Différenciation. On propose dans cette question d'utiliser la méthode de la différenciation pour éliminer la tendance.

1. Calculer la série temporelle `d` définie par $\nabla x_t = x_t - x_{t-1}$ pour $t \in \{2, \cdots, n\}$ à l'aide de la fonction `diff`
2. Représenter graphiquement cette série temporelle. Discuter.
3. Ajuster un modèle de tendance linéaire à la série ∇x_t. Discuter.

(d) Comparaison des différentes méthodes. Discuter les avantages et inconvénients respectifs des différentes méthodes vues dans les questions précédentes. Quelle méthode vous semble être la mieux adaptée ? On notera dans la suite `xdet` la série sans tendance correspondante.

4. Etude de la composante saisonnière

(a) Calculer les coefficients saisonniers (i.e. moyennes mensuelles) de la série `xdet`. On notera `xstat` la série obtenue après avoir retiré les coefficients saisonniers.
(b) Représenter graphiquement la série `xstat`. Cette série vous paraît-elle être stationnaire ?

5. Etude simultanée de la tendance et de la composante saisonnière par un modèle paramétrique

Une méthode classique pour modéliser les composantes non-stationnaires dans les séries temporelles météorologiques consiste à utiliser le modèle paramétrique suivant :

$$x_t = at + b + \alpha \cos(\omega t) + \beta \sin(\omega t) + \epsilon_t$$

avec a, b, α et β des paramètres inconnus et ω fixé de telle manière que la période de la fonction $t \longmapsto \cos(\omega t)$ soit égale à 1 an.

(a) Ce modèle vous paraît-il adapté? Comment y sont modélisées la tendance et les composantes saisonnières? Que représente t'il?
(b) Ajuster ce modèle avec la fonction `lm` et tracer la série résiduelle correspondante. Discuter les résultats et proposer des généralisations de ce modèle.

6. Utilisation de la fonction stl

Taper les commandes ci-dessous et interpréter les résultats obtenus

```
> dog=stl(x,'periodic')
> plot(dog)
```

7. Prédiction

Peut-on prévoir la température moyenne en Janvier 2100 et Juillet 2100 ? Si oui, donner les valeurs numériques et éventuellement un intervalle de confiance.

8. Etude de la série stationnaire.

(a) Rappeler la (ou les) définitions d'un processus stationnaire. Donner un exemple de processus stationnaire d'ordre 2 puis de processus non stationnaire. Simuler une trajectoire de chacun de ces processus et tracer ces trajectoires. Commenter les graphiques.

(b) Rappeler les définitions de la fonction d'autocorrélation et de la fonction d'autocorrélation empirique. Que représentent ces fonctions? Tracer la fonction d'autocorrélation empirique de la série initiale x à l'aide de la fonction `acf`. Discuter les résultats.

(c) Tracer la fonction d'autocorrélation empirique de la série *xstat*. Discuter. Peut-on supposer que les températures moyennes de deux mois successifs sont indépendantes? Peut-on affirmer que s'il fait plus chaud que la normale un mois donné, alors le mois suivant sera plutôt froid? Pourquoi?

(d) Rappeler la définition de la fonction d'autocorrélation partielle. Que représente cette fonction? Tracer la fonction d'autocorrélation partielle empirique de la série `xstat` à l'aide de la fonction `pacf`. Discuter.

(e) Peut-on supposer que la série *xstat* est un bruit blanc ? On répondra à l'aide la fonction `Box.test`. Rappeler les hypthèses de test associées à ce test. Conclure.

9. Modélisation de la série stationnaire.

(a) Quel modèle peut-on proposer pour la série stationnaire d'après les résultats de la section précédente? Justifier la réponse. Donner une expression analytique pour le modèle. Quels en sont les paramètres?

(b) Inférence - Quels estimateurs peut-on utiliser pour estimer les paramètres du modèle? Quel estimateur préconisez vous de choisir? Pourquoi? Utiliser la fonction `ar` pour estimer les paramètres du modèle.

(c) Proposer une méthode de validation du modèle. Mettez la en oeuvre.