Séries chronologiques
M1 Statistique et économétrie
TD 3 - Modélisation du processus stationnaire (modèle AR)
V. Monbet

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) sur la page du cours
(`http://perso.univ-rennes1.fr/valerie.monbet/ST_M1/ST_M1.html`)
Importer ces données en utilisant par exemple l'instruction `read.table` sous R ; on appelera la table `temperatures`. Puis créer une variable `x` de type *time series* qui contient les données à l'aide de la fonction `ts`.

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

2 Visualisation

Représenter graphiquement la série temporelle. Peut-on identifier une tendance et/ou une composante saisonnière ?
La composante déterministe a été modélisée au TD précédent.

3 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 donnée, 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 fonction `Box.test`. Rappeler les hypthèses de test associées à ce test. Conclure.

4 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) On peut envisager plusieurs approches pour valider le modèle.

1. La première consiste à regarder si le modèle ajusté vérifie bien les hypothèse qu'on a posées ; en particulier que le processus résidu $\{\epsilon_t\}_{t\in\mathbb{Z}}$ est bien un bruit blanc centré et qu'il suit une loi de Gauss si on a utilisé les estimateurs du maximum de vraisemblance.
   Pour vérifier l'hypothèse de bruit blanc, on peut utiliser le test de Ljung-Box (`Box.test(epsilon,lag=1,type="Ljung-Box")`). Ce test est basé sur l'hypothèse nulle suivante

   $$H_0 : \text{Les données sont indépendantes}$$

   (i.e. Les autocorrélation de la population dont l'échantillon est tiré sont nulles)
   Pour vérifier la normalité, on peut réaliser un test de Shapiro-Wilks (`shapiro.test()`). Il est toujours utile de prendre le temps de tracer un graphique des quantiles observés contre les quantiles théoriques (`qqnorm()`). Si les données sont bien issues d'un échantillon gaussien, les points sont (presque parfaitement) alignées sur la droite $y = x$.
2. La seconde méthode consiste à comparer les statistiques empiriques des observations aux statistiques théoriques. Ici, on pourra par exemple comparer la moyenne des observations à celle du processus et les fonctions d'autocorrélation. La comparaison peut se faire graphiquement (on trace les deux fonctions d'autocorrélation). On n'oubliera pas que les estimateurs empiriques ont une certaine variance et on tracera donc in intervalle de confiance autour de la fonction d'autocorrélation empirique.

5 Etude de la série de températures sous SAS

On modélise la tendance et la saison via une approche paramétrique et on considère un processus auto-régressif d'ordre 2. On utilise l'estimateur du maximum de vraisemblance.

1. Ecrire le modèle à ajuster.
2. Ecrire la vraisemblance.
3. Estimation. Avant d'appeler la procédure `autoreg`, créer une table contenant les température (y), le temps (t) et les deux premières fonctions de base de Fourier de période 1 an ($\cos(2\pi t/T)$ et $\sin(2\pi t/T)$) que vous pourrez nommer `cos_an` et `sin_an`.

```
proc autoreg data=temperature;
   model y = t cos_an  sin_an / nlag=2 method=ml;
run;
```