

TP 3

Estimation et élimination de la tendance et de la saisonnalité

Une série temporelle $(X_t)_{t \in \{1, \dots, n\}}$ est l'observation des n premières réalisations d'un processus stochastique $(X_t)_t$ indexé par le temps. C'est ce processus que l'on cherche désormais à modéliser. Pour cela, la démarche suivante doit être adoptée :

- représenter graphiquement la série afin de repérer les tendances et saisonnalités,
- estimer et supprimer les tendances et saisonnalités (partie déterministe du processus stochastique),
- choisir un modèle pour les résidus (partie aléatoire du processus stochastique) et l'estimer,
- prédire les réalisations futures à l'aide de ce modèle.

L'objectif de ce TP est de tester quelques méthodes pour estimer et supprimer les tendances et saisonnalités. Dans la suite, on reviendra sur la modélisation de processus stationnaires.

Quelques fonctions utiles pour la mise en oeuvre sous R

La fonction `decompose` permet d'extraire d'une série temporelle (via la méthode de la moyenne mobile) :

```
serie_decomp<-decompose(serie,type=c("additive","multiplicative"))
```

- la composante saisonnière : `serie_decomp$seasonal`, que l'on suppose additive ou multiplicative dans l'option `type`,
- la tendance : `serie_decomp$trend`,
- le partie aléatoire stationnaire de la série : `serie_decomp$random`.

La fonction `diff.ts(serie,lag=T,difference=k)` permet d'appliquer l'opérateur de différentiation k fois.

La fonction `Box.test(serie,lag=H)` examine l'hypothèse nulle de nullité des H premières autocorrélations, à l'aide du test du Portemanteau. Par défaut H est fixé à 1, et seule la nullité de l'autocorrélation d'ordre 1 est testée.

Pour tester si la série peut-être apparentée à un bruit blanc, nous fixerons arbitrairement un H de l'ordre de 20 (nous considérerons abusivement que si les 20 premières autocorrélations sont nulles, la série est indépendante).

Exercice 1 : Décomposition d'une série temporelle

Dans un modèle additif, une série temporelle Y_t s'écrit comme la somme de trois composantes :

$$Y_t = Tr_t + S_t + e_t,$$

où

- Tr_t est la partie tendance,
- S_t est la composante saisonnière,
- e_t est le terme aléatoire supposé stationnaire.

Le fichier `nybirths.dat` contient le nombre de naissances par mois dans la ville de New York entre janvier 1946 et décembre 1959.

1. En utilisant la fonction `scan`, enregistrer le contenu du fichier dans un vecteur `births`.
2. Créer l'objet série temporelle `birthsts` avec la commande :

```
birthsts<-ts(births,start = c(1946,1),frequency = 12)
```

-
3. Représenter la série avec la fonction `plot.ts`. Il est clair au vu du graphique qu'il y a une composante saisonnière avec un pic en été et un minimum en hiver.
 4. En utilisant la fonction `SMA` dans le package `TTR`, on va estimer la partie tendance. La fonction `SMA` utilise une moyenne mobile pour estimer la partie tendance.

```
birthstssma<-SMA(birthsts,n=12)
plot.ts(birthstssma)
```
 5. On va maintenant utiliser la fonction `decompose` pour obtenir les trois composantes de la série. Créer l'objet `birthsc` résultat de l'application de la fonction `decompose` à l'objet `birthsts`.
 6. Représenter les différents éléments de `birthsc` avec la commande :

```
plot(birthsc)
```
 7. Calculer et représenter la série sans la partie saisonnière, nommée `birthstssa`. Vous pouvez remarquer que la partie saisonnière de la série a été retirée.

Exercice 2 : Données AirPassengers

Nous étudions la série chronologique du nombre de passagers par mois (en milliers) dans les transports aériens, de 1949 à 1960. Cette série est disponible sous R (`AirPassengers`).

1. Estimation paramétrique de la tendance :
 - (a) Représenter graphiquement la série. Ce processus vous semble-t-il stationnaire? Présente-t-il des tendances et saisonnalités?
 - (b) Estimer les paramètres d'une tendance linéaire $at + b$.
 - (c) Supprimer cette tendance et représenter graphiquement la série ainsi obtenue. Vérifier que la série des résidus est de moyenne nulle.
 - (d) Calculer et représenter les autocorrélations de la série des résidus.
2. Méthode des différences :
 - (a) Appliquer la méthode des différences pour enlever la tendance et la saisonnalité. Préciser la période de la saisonnalité, le degré du polynôme de tendance.
 - (b) La série obtenue semble-t-elle stationnaire?
3. Méthode des moyennes mobiles :
 - (a) Appliquer la méthode des moyennes mobiles pour enlever la tendance et la saisonnalité.
 - (b) La série obtenue semble-t-elle stationnaire? Pourrait-on la modéliser par un bruit blanc?

Remarque : la fonction `Box.test` permet de tester la décorrélation d'une suite de variables aléatoires.

Exercice 3 : Données simulées

Récupérer la série temporelle contenue dans le fichier `simulation.dat`. Cette série a été simulée à partir d'un processus relativement simple. Essayer de le retrouver.

1. Analyser qualitativement cette série temporelle.
2. Pouvez-vous proposer une modélisation pour cette série (c'est-à-dire définir le processus X_t qui a généré cette série)?

Indication : la partie saisonnière est de la forme $a \cos(t\pi/b)$ ou $a \sin(t\pi/b) \dots$ à vous de deviner!