

Modélisation des séries temporelles

Séance 1

Conseils : créez un script R par séance (ou exercice). N'hésitez pas à faire des copier/coller depuis le fichier pdf, et à commenter vos scripts pour garder une trace de vos réflexions.

Rappels : `setwd` permet de changer de répertoire de travail (attention aux ' / ' dans les chemins), `getwd` de vérifier le répertoire courant, `help` pour l'aide (exemple : `help(setwd)`).

Les commentaires commencent par #.

Le symbole d'affectation est `<-`. Le symbole `=` peut être utilisé dans la console mais attention aux confusions avec le `=` utilisé dans les appels de fonctions.

Notions abordées : prise en main de R, chargement et représentation graphique des chroniques, décomposition par lissage, périodogramme.

1 Exercice 1 : chronique AirPassengers

La chronique `AirPassengers` représente le trafic aérien en milliers de passagers par mois, de janvier 1949 à décembre 1960.

Il s'agit d'une chronique du package `stat` chargé par défaut par R. La liste des chroniques disponibles peut être obtenue par :

```
data()
```

Les chroniques sont représentées par des objets de la classe `ts` (time series), qui ont des méthodes dédiées. Exécutez successivement :

```
AirPassengers
start(AirPassengers)
end(AirPassengers)
frequency(AirPassengers)
```

On voit que la chronique contient déjà l'information de la date. Que comprendre de la sortie de `frequency` ?

Pour une représentation graphique de la chronique :

```
plot(Airpassengers)
```

ou de manière plus lisible :

```
plot(AirPassengers, ylab="Passengers (1000s)", type="o", pch =20)
grid()
```

Quelle information tirer de ce graphe ?

Le graphe suggère-t-il un modèle additif ou multiplicatif ? Tracez le graphe du logarithme de la chronique.

Éliminez la composante saisonnière par un filtre simple (regardez la documentation de `filter`), et observez le graphique de la chronique filtrée superposé à la chronique de départ (utilisez `lines`). On rappelle que le filtre utile ici est représenté par le vecteur `R` :

```
c(1/24, rep(1/12, 11), 1/24)
```

Quelle information apparaît de manière plus visible sur la chronique filtrée ?

Retranchez cette chronique filtrée à la chronique de départ, pour observer une composante « pseudo-saisonnière » que vous représenterez graphiquement.

Différents algorithmes de décomposition par filtrage existent, l'un d'eux est `decompose`. Observez la sortie de :

```
decompAP <- decompose(AirPassengers, type="multiplicative")
plot(decompAP)
```

Affichez les graphiques de la chronique à laquelle on a retirée la composante irrégulière, de la composante saisonnière, et de la chronique corrigée des variations saisonnières. On utilisera les sorties de `decompose` : `decompAP$trend`, `decompAP$seasonal`, et `decompAP$random`. Quelle information qualitative tire-t-on ?

Que dire de la décomposition selon un modèle additif ?

2 Exercice 2 : chronique IPI

La chronique IPI (cf page web du cours) donne l'indice trimestriel de la production industrielle de la France (hors BTP) du premier trimestre 1963 au quatrième trimestre 1982 (source INSEE). Faites le même travail que dans l'exercice 1. Quels événements « voit-on » dans la composante résiduelle et la chronique CVS ?

Pour charger et préparer le fichier `IPI.txt` :

```
IPI <- read.table(file="ipi.txt", header=TRUE)
IPI <- ts(IPI, start = c(1963, 1), frequency=4)
plot(IPI, ylab="Indice production industrielle", type="o", pch =20)
```

On lit le fichier (que vous pouvez aussi ouvrir avec un éditeur de texte quelconque) comme une table dont la première ligne comprend l'entête. Puis `IPI` est transformé en objet `ts`, dont il faut préciser date initiale (ici, janvier 1963) et « fréquence » (ici, 4 données par année).

3 Exercice 3 : chronique sunspot.month

Que représente la chronique `sunspot.month` ? (`help(sunspot.month)`)

Le périodogramme représente la contribution des différentes composantes fréquentielles de la chronique. La fonction `periodogram` est fournie par la bibliothèque `TSA`.

```
install.packages(TSA) # télécharge/installe TSA si ce n'est déjà fait
library(TSA) # active TSA
```

Rappel : on installe un package (`install.packages`) une seule fois, mais on l'active (`library`) à chaque nouvelle session R.

Observez le résultat de :

```
plot(sunspot.month)
p <- periodogram(sunspot.month)
```

Tout signal « régulier » a un spectre décroissant avec la fréquence. Un pic dans le périodogramme est signe d'une composante périodique contribuant de manière importante. Les pics permettent donc d'identifier les composantes périodiques sous-jacentes dans le signal. Il ne faut pas tenir compte des basses fréquences dont les coefficients de Fourier associés auront toujours une valeur assez importante. Que fait le code suivant ?

```
dd <- data.frame(freq=p$freq, spec=p$spec, period=1/p$freq)
dd <- dd[-c(1:4),]
order <- dd[order(-dd$spec),]
top5 <- head(order, 5)
top5
```

Quel est le rapport avec : https://fr.wikipedia.org/wiki/Cycle_solaire?

À l'aide du périodogramme, retrouvez les périodicités des chroniques des exercices 1, 2, 3.

4 Exercice 4 : taux de chômage aux États-Unis (pour les plus rapides)

Chargez les données relatives au taux de chômage américain ici : <https://data.bls.gov/timeseries/LNU04000000> (vous pouvez sélectionner un intervalle de temps assez grand, par exemple une quarantaine d'années).

Dans le fichier Excel, supprimez les premières lignes pour garder uniquement les données et les entêtes des colonnes (à partir de la ligne 12 du tableau généré par le site web). Dans cet exemple, le fichier s'appelle `unemployment.xlsx`, et la date de départ est janvier 1973.

```
install.packages("xlsx")
library("xlsx")
unemploy <- read.xlsx("Chemin/vers/fichier/unemployment.xlsx", 1)
unemploy.ts <- ts(as.vector(t(unemploy[, -1])), frequency=12, start=c(1973, 1))
```

Ici, on construit l'objet `ts` à partir du vecteur ligne des données : il a fallu supprimer la première colonne et transposer la matrice pour former le vecteur des valeurs successives de la chronique, mois après mois.

En cas de difficulté avec l'import de fichiers Excel, utilisez le fichier sur la page web du cours (cf exercice 2).

Observez les variations saisonnières et la chronique corrigée des variations saisonnières.

Remarque : en cas de problème Java à l'activation de `xlsx` : mettez à jour la Java JDK sur votre machine, sinon essayez le package `readxl`. Une autre possibilité est de convertir le fichier `xlsx` en `csv` sous Excel.

5 Exercice 5 : chronique Airmiles (pour les plus rapides)

À titre de complément, décomposez selon la même méthode que dans l'exercice 1 la série `airmiles` de la bibliothèque TSA, sur la période janvier 1996 – mai 2007. Justifiez qu'un modèle additif convient. Pour charger la chronique (il faut télécharger la bibliothèque TSA auparavant) :

```
library(TSA) # active TSA si ce n'est déjà fait
data(airmiles)
plot(airmiles, type="o", pch =20)
```

Si `data(airmiles)` ne vous indique pas la bonne chronique (dates incorrectes), faites :

```
data(airmiles, package="TSA")
```