

Master M.A.E.F. 2016 – 2017

TP de Séries Temporelles n^0 2 :

Tendance et saisonnalité d'une série réelle

L'objectif de ce TP est de mettre en pratique différentes procédures de régression permettant d'estimer les tendances et saisonnalités éventuelles d'une série chronologique. Cette fois-ci, et contrairement à ce que l'on avait fait avec les séries simulées du TP1, on travaille avec une série issue de données réelles et dont on ne sait rien d'une possible modélisation. Cette série représente le nombre de déménagements par mois aux Etats-Unis (en milliers).

Importation d'un fichier de données extérieur

```
hstart=read.table("C:/Donnees/hstart.txt")
```

On "importe" donc dans R les données écrites dans un fichier texte *hstart.txt*, se trouvant dans le répertoire *C : /Donnees/*.
On appelle *hstart* la table ainsi construite. Afficher *hstart* et déterminer quel type d'objet R est *hstart*.

```
hstart$V1
```

Affiche la première variable de *hstart*.

```
t=hstart$V1
```

On appellera *t* la variable temps,

```
nbdem=hstart$V2
```

et *nbdem* le nombre de déménagements correspondants.
Transformer *nbdem* en une série chronologique *nbdem2* de temps *t2*.
Afficher le graphe des variations de *nbdem2*.

Estimation de la tendance seule par régression

On commence par estimer l'éventuelle tendance de la série chronologique par des régressions linéaires de degrés de plus en plus élevé.

Régression linéaire simple

```
reg1=lm(nbdem ~ t)
```

Régression linéaire de x par tp .

```
summary(reg1)
```

Résumé des résultats de la régression.

```
plot(t,nbdem)
```

Visualisation du nuage de points

```
segments(t,reg1$fit,t,nbdem)
```

Visualisation des résidus.

```
par(mfrow = c(2,2))
```

On divise cette fenêtre graphique en (2, 2).

```
plot(reg1)
```

Trace 4 graphiques permettant d'analyser le résultat de la régression.

Régression polynomiale (degré 2)

```
reg2=lm(nbdem ~ poly(t,2))
```

Ajustement polynomial (degré 2).

```
plot.ts(nbdem)
```

Graphique de la série chronologique.

```
lines(t,reg2$fit)
```

Graphe du polynôme associé.

```
summary(reg2)
```

Analyse détaillée de la régression.

```
plot(t,nbdem)
```

Visualisation du nuage de points

```
segments(t,reg2$fit,t,nbdem)
```

Visualisation des résidus.

Régression polynomiale (degré 7)

```
reg7=lm(nbdem ~ poly(t,7))
```

On utilise maintenant un ajustement polynomial de degré 7
Refaire les mêmes démarches que précédemment.

Choix du degré du polynôme avec le critère BIC	
library(MASS)	On charge le package MASS
long=length(t)	Nombre de données de la série chronologique
kmax=10	Degré maximum du polynôme
Z=matrix(nrow=long,ncol=kmax+1)	On définit une matrice Z
for (i in 1:long)	On va définir la matrice Z élément par élément
for (j in 1:(kmax+1))	
Z[i,j]=i^(j-1)	
ZZ=as.data.frame(Z)	On transforme Z en objet "data.frame"
x.lm=lm(ndem~.,data=ZZ)	
x.AIC=stepAIC(x.lm,k=2)	On utilise une commande qui minimise le critère AIC
x.BIC=stepAIC(x.lm,k=log(long))	Même chose mais avec le critère BIC
	Commentaires sur les résultats?

Régressions locales pour estimer la tendance

Pour obtenir une meilleure adaptation de la régression aux variations locales de la série, on peut procéder avec des régressions locales (loess et lowess : "lo" comme local, "ss" comme scatterplot smoothing) qui fonctionnent avec une fenêtre (dont on peut directement choisir la longueur dans le cas lowess), "coulissant" successivement le long des temps de la série. Dans chaque fenêtre, on fait une régression polynomiale de degré en général 0, 1 ou 2, par moindres carrés pondérés, sur les données de la série, et on obtient ainsi un point lissé pris en général au milieu de la fenêtre. Les poids de la pondération sont obtenus suivant la règle suivante: les poids sont plus élevés plus on est proche du point à lisser. Dans le cas de la régression lowess, on opère de manière itérative, une nouvelle pondération (d'où le "w" pour weighted), en fonction des écarts aux valeurs initiales.

Régression loess	
regl1=loess(nbdem ~ t)	régression "locale" standard.
names(regl1)	Différents résultats de la régression loess.
ts.plot(nbdem)	On trace à nouveau les variations de la série.
lines(t,regl1\$fit)	Graphique de la régression locale associée.
points(t,nbdem)	Représentation des points.
title("Résultats de loess(nbdem ~ t)")	On met une légende au graphique.
R2=1-var(regl1\$res)/var(nbdem)	Calcul direct du R2. Conclusion?

Régression par lissage lowess	
regl2=lowess(t,nbdem,f=0.5)	Lissage par lowess de la série <i>nbdem</i> , avec une taille de fenêtre fixée à $0.5 \times$ (longueur de <i>nbdem</i>).
names(regl2)	Différents résultats de la régression lowess. Noter que lowess n'a pas la syntaxe et les options identiques aux autres régressions.
low1=ts(regl2\$y,freq=12,1966)	On forme à nouveau une série chronologique.
ts.plot(nbdem,low1)	Tracé de la série <i>nbdem</i> et de la série lissée.
regl3=lowess(t,nbdem,f=0.1)	Lissage par lowess de la série <i>nbdem</i> , avec une taille de fenêtre fixée à $0.1 \times$ (longueur de <i>nbdem</i>).
low2=ts(regl3\$y,freq=12,1966)	On forme la série chronologique.
ts.plot(nbdem,low2)	Tracé de la série <i>nbdem</i> et de la série lissée.
ts.plot(low1,low2)	Pour comparer les deux lissages; expliquer la différence entre les deux lissages. Changer à nouveau la taille de la fenêtre afin d'obtenir un saisonnier + bruit satisfaisant (qu'est-ce ?..) Quelle est la taille de la fenêtre choisie ?

Estimation du saisonnier

Première estimation du saisonnier

<code>ts.plot(nbdem2)</code>	On travaille toujours avec la même série chronologique représentant le nombre de le nombre de déménagements par mois aux Etats-Unis.
<code>monthplot(nbdem2)</code>	Tracé de la série chronologique. Moyenne par mois avec visualisation des évolutions suivant les années. On obtient ainsi une première évaluation du saisonnier.
<code>monthplot(nbdem2,type="h")</code>	Même chose avec une autre représentation de l'évolution. Il est à noter que malgré le nom, la commande <code>monthplot</code> est utilisable avec des données trimestrielles ou d'autres périodes (voir <code>help</code>). Remarquez la similitude des évolutions vis-à-vis de la saisonnalité. D'après vous, qu'est-ce qui explique cette similitude?

Autres estimations du saisonnier

<code>s=stl(nbdem2,s.win="perio")</code>	Permet de récupérer le saisonnier obtenu dans <code>monthplot</code> . Donc dans ce cas, on a une estimation du saisonnier qui est une moyenne par mois.
<code>names(s)</code>	Donne les noms attachés à la commande <code>stl</code> .
<code>s\$time</code>	Affiche les différents résultats de la commande <code>stl</code> .
<code>monthplot(s)</code>	Tracé des résultats. Notez que le saisonnier est bien de somme nulle.
<code>s\$time[, "seasonal"]</code>	Que représente cette série ? On peut aussi vouloir "lisser" le saisonnier pour éviter les petites fluctuations qui ne semblent essentielles, donc non liées au saisonnier lui-même, mais plutôt à son estimation.
<code>s=stl(nbdem2,s.window=60)</code>	Donne plus de continuité pour le saisonnier. Comparer les résultats obtenus en faisant varier le paramètre <code>s.win</code> . Lequel choisir ?

Exercice

A partir de la même série `hstart`, poursuivre le travail en estimant la tendance (expliciter votre choix d'estimateur), quand est retiré à la série son saisonnier. Mémoriser le résultat. Dans un deuxième temps, faire la démarche inverse, en commençant par estimer la tendance à partir de `hstart`, puis sur la série détendancialisée en estimant le saisonnier. Comparer ainsi les deux résultats obtenus.

Estimation conjointe de la tendance et du saisonnier

En fait, on va proposer deux méthodes pour estimer conjointement la tendance et le saisonnier: la commande `stl` permet une estimation à la fois du saisonnier et de la tendance (fondée sur une méthode d'estimation par la régression `loess`).

<code>s=stl(nbdem2,s.win="perio",t.win="20")</code>	On demande que la commande <code>stl</code> permette également l'estimation de la tendance, avec une fenêtre locale de taille <code>t.win</code> .
<code>plot(s)</code>	Visualisation du résultat.
<code>s\$time[, "remainder"]</code>	Que représente cette série ? Modifier les paramètres de la commande <code>stl</code> jusqu'à obtenir une double estimation qui vous paraît satisfaisante, ce que l'on peut juger à partir de la distribution des résidus, que l'on souhaiterait les plus proches possibles d'un bruit blanc.

Exercice

Utiliser une régression pour estimer conjointement la tendance et la saisonnalité, comme cela a été proposé dans le cours. Dans un premier temps, on travaillera en fixant le degré du polynôme, par exemple 1. Ensuite, on utilise le critère BIC pour choisir le degré du polynôme.