

TP : station de comptage de trafic (jour 3)

1 Introduction

On réutilise le jeu de données `S21.csv` provenant d'une station de comptage de trafic routier (boucle magnétique), fournie par la société SISTeMA ITS. On cherche aujourd'hui à étudier le recalage des données.

Comme précédemment, on donne ci-dessous une liste de questions pour guider le travail. On demandera de répondre de manière brève mais aussi précise que possible en donnant :

- le code utilisé,
- le résultat (données, graphiques),
- éventuellement un commentaire et/ou des explications sur ce que vous avez fait.

On part là encore des mêmes données. On n'a pas pris les données reconstruites du TP2, mais cela ne change pas grand chose. Par contre, on retire les vendredis et les samedis, dont on a vu au TP précédent qu'ils sont très différents des autres jours.

```
> # dat0 contient les donnees brutes
> dat0=as.matrix(read.table("S21.txt", head=T))
> heures=c(0:239)/10
> # Vecteur des donnees manquantes par ligne
> na_count = rowSums(is.na(dat0))
> # Journees sans donnee manquante
> dat = dat0[na_count == 0,]
> # Les jours de la semaine (0=dimanche, ... 6=samedi)
> jours=as.numeric(format(as.Date(rownames(dat), format="%d/%m/%Y"), "%w"))
> # on se limite dimanche-jeudi
> dat = dat[jours <= 4,]
> # on adapte les jours
> jours = jours[jours <= 4]
> njours = nrow(dat)
> # Si on veut reduire les grandes valeurs
> #dat=sqrt(dat)
> require(fda)
```

2 Modélisation efficace de la dérivée du flux

On veut dans la partie suivante explorer les possibilités de recalage de courbes de débit. Comme les extrema locaux correspondent à l'annulation de la dérivée, on va commencer par faire un bonne modélisation de cette dérivée. On sait que les calculs de recalage sont longs et on cherche à réduire le nombre de fonctions de la base.

Question 1: *représenter de manière fonctionnelle (avec une base de B-splines) les débits de manière à pouvoir calculer leur dérivée première. Étudier l'effet de la diminution du nombre de points sur le SSE et choisir une taille de base raisonnable.*

3 Recalage des données

Les calculs de recalage étant longs, on pourra se limiter, notamment pour le recalage fonctionnel, à quelques courbes (au moins 5).

Question 2: *Faire un recalage par décalage uniforme de la dérivée du débit. On notera le temps de calcul (voir fonction `system.time()`)*

Question 3: *Faire un recalage fonctionnel de la dérivée du débit. On notera le temps de calcul et la proportion de variabilité due à la phase.*

Question 4: *Appliquer la méthode de Procuste. Est-ce que les résultats sont meilleurs ?*

Question 5: *Faire un recalage par point de repère de la dérivée du débit. Comment choisit-on un bon point de repère ? On notera le temps de calcul et la proportion de variabilité due à la phase. Commenter la qualité des résultats*

Pour trouver les points de repères, on pourra utiliser la fonction suivante (présente dans le fichier `points.repere.R`) qui demande interactivement à l'utilisateur de sélectionner le point de repère sur chaque courbe et renvoie un tableau de valeurs.

```
points.repere = fonction(dat) {  
  ncourbes = ncol(coef(dat))  
  pt.rep = rep(0, times=ncourbes)  
  
  for(i in 1:ncourbes) {  
    plot(dat[i], main=paste("Courbe", i))  
    pt.rep[i] = locator(1)$x  
  }  
  pt.rep  
}
```

Question 6: Appliquer le recalage fonctionnel aux données recalées par point de repère. Est-ce que le résultat est meilleur ?

Question 7: Quelle est la méthode la plus adaptée à l'analyse des données complètes ? À quel temps de calcul faut-il s'attendre ?