

# TP : station de comptage de trafic (jour 1)

## 1 Introduction

On dispose d'un jeu de données `S21.csv` provenant d'une station de comptage de trafic routier (boucle magnétique), fournie par la société SISTeMA ITS. On cherche à étudier la structure des données fournies. Les mesures ont été faites du 01/01/2010 au 31/01/2012.

Le script `sistema.r` a permis de transformer cette liste de mesures en une table `S21.txt` dont chaque ligne donne pour une journée donnée le nombre de véhicules comptés par tranches de 6 minutes. Certaines données de comptage sont manquantes (NA).

On demande pour ce TP de fournir un petit rapport d'analyse préliminaire des données. Pour cela on s'appuiera sur

- le logiciel R installé sur les machines, ainsi que le paquet « fda » ;
- le fichier de données `S21.txt`, dont les lignes représentent des journées entières et les colonnes les mesures toutes les 6 minutes. Au total la table comporte 761 lignes et 240 colonnes.

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.

## 2 Examen rapide des données

En fait, contrairement à ce que mentionne la documentation du paquet `fda`, les valeurs manquantes ne sont pas gérées correctement. On va donc se limiter aux lignes dont les données sont complètes dans un premier temps. On charge les données et on retire les individus ayant des données manquantes de la manière suivante :

```
> # Il est preferable de conserver les donnees dans une matrice
> dat0 = as.matrix(read.table("S21.txt", head=T))
> na_count = rowSums(is.na(dat0))
> dat = dat0[na_count == 0,]
> # La variable heures contient le temps dans la journee en heures
> # toutes les 6 minutes (240 points).
> heures = c(0:239)/10
```

Dans cette section et la suivante on utilisera le tableau `dat`.

**Question 1:** *vérifier la répartition des valeurs pour les débits. Expliquer pourquoi on peut appliquer la méthode de stabilisation de la variance pour processus de Poisson. Est-ce que cela améliore la queue de distribution ?*

**Question 2:** *Tracer des courbes pour quelques journées (au moins 5)*

**Question 3:** *Tracer la courbe du débit moyen au cours de la journée. Commenter la forme générale de la courbe.*

## 3 Lissage de données

**Question 4:** *Lissage de la moyenne : essayer avec plusieurs valeurs de  $\lambda$  pour en trouver une qui lisse correctement la moyenne tout en gardant un SSE correct. Tracer la moyenne correspondante*

**Question 5:** *Lissage d'une courbe : choisir un des individus et calculer sa version lissée (on expliquera comment on choisit  $\lambda$ ). On fera un graphique joint des données brutes et des données bruitées.*

**Question 6:** *Lissage de plusieurs courbes : lisser les mêmes courbes qui ont été tracées dans la section ?? et faire un tracé joint des courbes lissées (pour une valeur raisonnable de  $\lambda$ )*