

## Description des données - Projet R

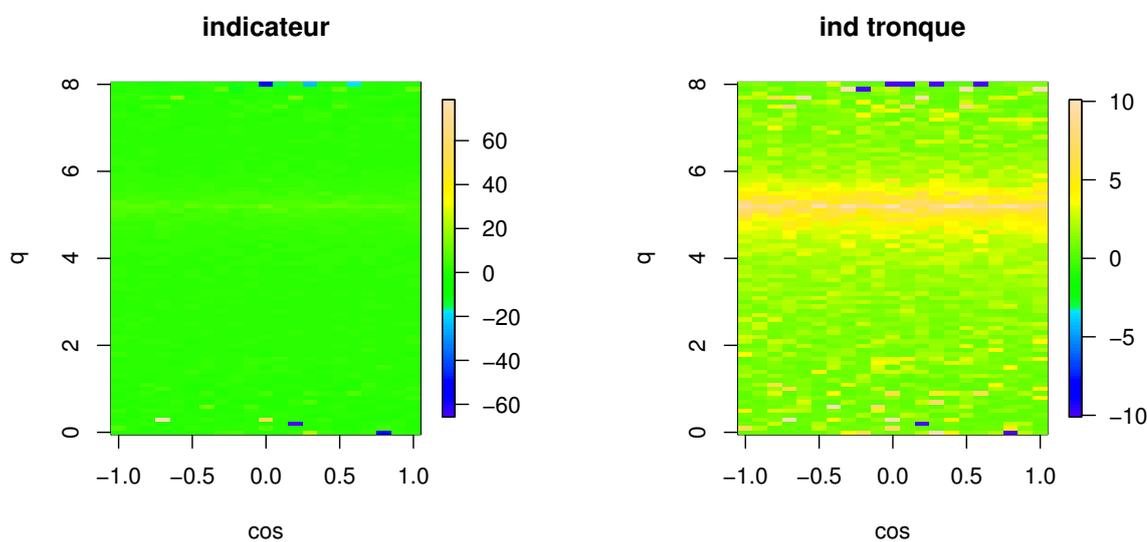
### M2 TIDE 2018-2019

Les jeux des données représentent des perturbations planétaires qui sont appliquées aux comètes du nuage d'Oort. Les perturbations sont associées aux positions dans l'espace  $K = \{(q, \cos i) \in [0, 15] \times [-1, 1]\}$  où  $q$  représente la distance périhélique et  $i$  représente l'angle d'inclinaison de la trajectoire elliptique d'une comète quelconque. Les valeurs  $q = 5.2$  et  $q = 9.6$  correspondent aux demi-grand axes de Jupiter et Saturne. Cela veut dire que la comète ayant un périhélie proche peut croiser l'orbite d'une de ces grandes planètes. Ainsi, on attend que les perturbations dans ces régions soit plus grandes ... statistiquement parlant ... Malgré le fait, que les perturbations soient uniformément réparties dans  $K$ .

On dispose deux jeux de données `data1` et `data2` qui sont des matrices de trois dimensions dont la première contient 10000 observations de perturbations et les deux dernières représentent respectivement  $q$  et  $\cos i$ . C'est à dire pour chaque point dans  $K$ ,  $(q, \cos i)$ , on a 10000 observations de perturbations. Pour discrétiser les positions, on a pris 0.1 comme la longueur des intervalles. Le premier jeu de données est localisé dans le domaine  $K_1 = [0, 8] \times [-1, 1]$ , alors que le deuxième est localisé dans  $K_2 = [6, 15] \times [-1, 1]$ .

Le programme ci dessous calcule les indicateurs de queue lourde pour `data1`. La formule de l'indicateur est  $\frac{\hat{z}_{0.99}}{\hat{n}_{0.99}} - 1$  où  $\hat{z}_{0.99}$  est le 0.99-quantile empirique et  $\hat{n}_{0.99}$  est le 0.99-quantile théorique de la loi normale avec la moyenne et l'écart-type estimés par  $\hat{z}_{0.5}$  et  $(\hat{z}_{0.84} - \hat{z}_{0.16})/2$ . Les résultats sont présentés dans les images suivants.

```
nrow = dim(data1)[2]; ncol = dim(data1)[3]
ind = matrix(0, nrow = nrow, ncol = ncol); indt = ind; q = 0.99
for(i in 1:nrow){
  for(j in 1:ncol){
    d = data1[,i,j]; qd = quantile(d, probs = q)
    moy = median(d); qec = quantile(d, probs = c(0.84,0.16))
    ec = (qec[1]-qec[2])/2; qn = qnorm(q, moy, ec)
    ind[i,j] = qd/qn-1
  }
}
indt = ind; indt[indt < -10] = -10; indt[indt > 10] = 10;
library(fields);
index.q = seq(from = 0, to = 8, by = 0.1); index.cos = seq(-1, 1, 0.1);
image.plot(index.cos, index.q, t(ind), xlab = "cos", ylab = "q",
col=topo.colors(100), main="indicateur")
image.plot(index.cos, index.q, t(indt), xlab = "cos", ylab = "q",
col=topo.colors(100), main="ind tronque")
```



Téléchargez les deux jeux des données disponibles à partir de la page suivante :

<https://filex.univ-paris1.fr/get?k=FxN8GoB9kW2CEQGFXgB>

**A l'attention de l'étudiant :** Envoyez moi **avant le 31 mai 2019 minuit**, un compte-rendu contenant les solutions aux trois exercices du projet. Après cette date tout compte-rendu sera irrecevable. Vous pouvez travailler en binôme. Les solutions doivent être soigneusement rédigées. Veillez à inclure les graphiques demandés ainsi que les codes (ou fragments de code) qui vous semblent les plus appropriés et qui argumentent au mieux vos réponses.

Vous pouvez me contacter directement dans mon bureau (C.20.07) ou bien par email (Shuyan.Liu@univ-paris1.fr).