

TD1 - Fouille

September 12, 2021

1 Manipulations manuelles simples

Exercice 1 On considère la série *exo1dat* sur le site Moodle. Calculer la moyenne simple et la médiane.

Le code R pour générer ces données est le suivant:

```
round(100*rexp(20))
```

Exercice 2 On considère la série *exo2dat* sur le site Moodle. Calculer la moyenne mobile pour un intervalle de durée 4.

Le code R pour générer ces données est le suivant:

```
round(100*runif(10))
```

Exercice 3 Sans calculer les moyennes, dire quelle distribution a la plus grande moyenne

- – 3 5 5 5 8 11 11 11 13
– 3 5 5 5 8 11 11 11 20
- – 20 0 0 0 15 25 30 30
– 40 0 0 0 15 25 30 30
- – 0 2 4 6 8 10
– 20 22 24 26 28 30
- – 100 200 300 400 500
– 0 50 300 550 600

Exercice 4 On considère la série *exo3dat* sur le site Moodle. Il s'agit de données qualitatives. Calculer le mode et le nombre d'occurrence de chaque valeurs.

Le code R pour générer ces données est le suivant:

```
colo = c("jaune","vert","rouge","noir","blanc","violet")  
sample(colo,10,replace=TRUE)
```

2 Codes en C

Exercice 5 Ecrire un programme C qui calcule la trimmed mean avec $k=1$ pour une série de longueur n avec une complexité linéaire en n (donc sans faire de tri...). Les données (des réels) seront dans un fichier texte avec la longueur de la série lue au clavier. Le fichier est sur le site Moodle (*exo4dat*)

Exercice 6 Ecrire un programme *C* qui lit un fichier texte de p individus et n variables. Le fichier est sur le site Moodle (exo5dat). La table est notée $T(i, v)$. Les individus sont en ligne (1 individu par ligne). Vous pouvez supposer que p et n sont lus au clavier avant de lire le fichier. Les variables sont qualitatives.

- Construire une matrice de distance entre individus (donc une matrice $p \times p$) en utilisant la distance de Hamming entre $i1$ et $i2$

$$d(i1, i2) = \frac{\sum_{v=1}^n 1_{T(i1,v) \neq T(i2,v)}}{n}$$

- Construire une matrice de distance entre variable (donc une matrice $n \times n$) en utilisant la distance de Hamming entre $v1$ et $v2$

$$d(v1, v2) = \frac{\sum_{i=1}^p 1_{T(i,v1) \neq T(i,v2)}}{p}$$

3 R

Exercice 7 Installer *R* (le langage et les bibliothèques) et *Rstudio* (l'interface graphique) sur votre machine.

R : <https://cran.rstudio.com>

Rstudio : <https://www.rstudio.com/products/rstudio/download/>

Exercice 8 En *R*, lire le fichier *exo6dat* (toujours sur Moodle) et le mettre dans une table. La première ligne contient les titres des variables. Faire imprimer les statistiques descriptives des variables et commentez.

4 Math

Exercice 9 On suppose que l'on a des entiers aléatoires strictement positifs. On va montrer que sur ces entiers la moyenne géométrique est inférieure ou égale à la moyenne arithmétique.

1. On rappelle qu'une fonction dérivable deux fois est concave si sa dérivée seconde est négative ou nulle. Montrez que le log est une fonction concave.
2. On admet les inégalités de Jensen. Si f est une fonction convexe, alors pour tout vecteur x_1, \dots, x_n et toute pondération t_1, \dots, t_n telle que $\sum_i t_i = 1$ et $t_i \geq 0$.

$$f\left(\sum_i t_i x_i\right) \leq \sum_i t_i f(x_i)$$

3. en utilisant que si une fonction f est concave alors $-f$ est convexe et que la fonction logarithme est croissante, démontrez que la moyenne géométrique est inférieure ou égale à la moyenne arithmétique.

Exercice 10 On suppose que les données sont des réels strictement positifs. Montrer que la moyenne arithmétique est égale à la moyenne géométrique si et seulement si toutes les données sont identiques.

Exercice 11 Montrez que la distance entre individus codée dans l'exercice 6 est bien une distance.

5 Expérience

Exercice 12 1. *Etape 1 Les deux premiers rangs vont essayer de deviner combien il y a de pièces de 1c dans la bouteille (figure ci dessous).*

2. *Les étudiants de 2 premiers rangs, vous écrivez chacun votre proposition sur une feuille.*
3. *Etape 2 Je ramasse, j'écris au tableau les valeurs, vous les rentrez dans R et vous calculez la moyenne et la médiane*
4. *Etape 3 On refait la même expérience avec les rangs 3 et 4.*
5. *Que constatez vous ?*

