



# Big Data

## TP 1 : Illustration et évaluations

### BUT 3

Guillaume Metzler et Antoine Rolland  
Institut de Communication (ICOM)  
Université de Lyon, Université Lumière Lyon 2  
Laboratoire ERIC UR 3083, Lyon, France

[guillaume.metzler@univ-lyon2.fr](mailto:guillaume.metzler@univ-lyon2.fr); [antoine.rolland@univ-lyon2.fr](mailto:antoine.rolland@univ-lyon2.fr)

Dans le présent TP, on se propose d'illustrer les différents points vus en cours ainsi que certains problèmes réalisés dans les premières fiches de TD.

Dans la suite, nous emploierons le langage  pour la résolution de ces derniers.

## Exercice 1 : Retour sur la distance entre individus en grandes dimensions

On se propose d'étudier les distance entre individus en fonction de la dimension du jeu de données.

Une étude a déjà été faite lors du premier TD et on souhaite simplement représenter les résultats graphiquement. Pour cela, on va considérer des échantillons de taille  $n = 100$  et des espaces avec les dimensions suivantes :

- $p = 2$
- $p = 10$
- $p = 100$
- $p = 1000$

Les données seront générées selon une loi normale centrée et réduite.

On pourra utiliser les fonctions *dist* pour le calcul des distances entre individus et *hist* ou encore *lines* de  pour représenter les histogrammes des distances.

## Exercice 2 : Retour sur la suite de Fibonacci

On revient sur les approches récursives et itératives permettant de calculer le  $n$ -ème terme de la suite de Fibonacci  $F_n$ .

1. En reprenant la fiche de TD 4, implémenter la procédure récursive et représenter le temps de calcul en fonction de  $n$  (prendre des de valeurs de  $n \leq 35$ ).
2. Calculer le temps nécessaire au calcul de  $Fibo(30)$  avec la précédente procédure. Puis, en utilisant le fait (on ne le montrera pas ici) que, pour tout  $m > n$ , le temps de calcul de  $Fibo(m)$  est environ égal au temps de calcul de  $Fibo(n) * \left(\frac{1 + \sqrt{5}}{2}\right)^{m-n}$ , estimer le temps nécessaire pour calculer :
  - (a)  $Fibo(50)$ , en heure.
  - (b)  $Fibo(500)$ , en années.
3. Refaire la même chose en implémentant la procédure itérative, de complexité linéaire en temps et constante en mémoire. Mettre en évidence la complexité linéaire en temps.