M1 MIASHS
2020-2021

TP

1 Rappels : Types de variables sous R

La principale difficulté dans l'utilisation de R est de bien identifier les types d'objets manipulés.

```
# Vecteur
x = 1:10 # définition d'une séquence
x
y = 2*x + 3
y[5] ; y[1:3] ; y[-3] # composants d'un vecteur

# Matrice
A = matrix(1:15,ncol=5); A
B = matrix(1:15,nc=5,byrow=TRUE) ; B
A[1,3] ; A[,2] ; A[2,] ; A[1:3,1:3] # composants

# Liste
x=list(mat=A, texte="testliste",vec=y)
x[[2]] ; x$vec # composants

# Base de donnée ou data frame
# Tableau contenant des vecteurs de types
# éventuellement différents
taille = c(147, 132, 156, 167, 156, 140)
poids = c( 50, 46, 47, 62, 58, 45)
sexe = c("M","F","F","M","M","F")
H = data.frame(taille,poids,sexe)
H
summary(H)
plot(H$poids,H$taille)
```

2 Simulation et représentation graphique

Dans cette section, nous allons illustrer, par des simulations, les propriétés des estimateurs élémentaires (moyenne, écart-type, histogramme).

2.1 Loi normale $\mathcal{N}(\mu = 80, \sigma^2 = 30)$

- Utiliser la fonction `rnorm` pour générer n valeurs aléatoires d'une variable Y selon une loi normale de moyenne m et d'écart-type s avec : $n = 10$, $\mu = 80$, $\sigma^2 = 30$.
- Utiliser les fonctions `summary()` et `boxplot()` pour donner les quartiles et les représenter.
- Tracer un histogramme de la **distribution statistique** avec la fonction `hist()`.
  Que fait le paramètre `breaks` de la fonction `hist()` ? et le paramètre `proba` ?
- Tracer la courbe de la **densité de probabilité**. (Indication : utiliser la fonction `dnorm()`)
- Tracer à nouveau l'histogramme et superposer la courbe de densité avec la fonction `lines()`.

2.2 Une loi discrète

Tracer le graphe (diagramme en bâtons) représentant la loi de probabilité suivante :

v	-10	20	100
$P(X = v)$	$\frac{7}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

2.3 Loi Binomiale $\mathcal{B}in(n = 10, p = 0.7)$

- construire le tableau de la loi de probabilité à l'aide de la fonction *dbinom*
- faire la représentation graphique
- comparer pour différentes valeurs de p

2.4 Loi Géométrique $\mathcal{G}eom(p = 0.25)$

- faire la représentation graphique à l'aide de la fonction *dgeom*
- comparer pour différentes valeurs de p

3 Lois limites

3.1 Loi des grands nombres

Une moyenne et un écart-type sont la réalisation d'une variable aléatoire appelée **"estimateur"** ; ce sont des **estimations**.

1) Refaire les calculs et graphiques effectués Section 2.1 en posant $n = 100$; $n = 1000$; $n = 10000$;
2) Comparer les résultats obtenus, notamment les estimations des indicateurs par rapport aux valeurs théoriques.
3) Comparer leur comportement en fonction de la taille n de l'échantillon.
4) Donner l'expression d'un "bon" estimateur de la moyenne de la variable Y. Quelles sont ses propriétés ?
5) Donner l'expression d'un "bon" estimateur de la variance de la variable Y. Quelles sont ses propriétés ?
6) Calculer les estimations de la moyenne et de la variance de Y pour $n = 10$; $n = 100$; $n = 1000$; $n = 10000$. Que remarquez-vous ?

3.2 Théorème central limite

Cette section a pour objectif d'illustrer le résultat fondamental du théorème de la limite centrale : la variable aléatoire moyenne (de variables indépendantes et de même loi) centrée réduite converge vers une variable aléatoire de loi gaussienne centrée réduite.

Nous allons l'illustrer en prenant l'exemple d'un loi de Poisson d'espérance $\lambda = 5$.

1) Ecrire un programme qui exécute les opérations suivantes :
   - Initialisation des dimensions `K` et `n` pour un tirage de `K` échantillons, chacun de taille `n`.
   - Initialisation par des 0 d'un vecteur de taille N pour stocker les valeurs observées de la moyenne centrée réduite sur les `K` échantillons

- Définition des valeurs de l'espérance `mu` et de la variance `sigma2` de la loi de Poisson d'espérance $\lambda = 5$.
- Ecrire une boucle pour calculer la valeur observée de la moyenne centrée réduite sur chacun des `K` échantillons
- Représenter par un histogramme la distribution empirique de la variable aléatoire moyenne centrée réduite à partir des `K` échantillons.
- Comparer la distribution empirique avec la densité loi théorique limite qui est, d'après le TCL, la loi gaussienne centrée réduite.

2) Faire varier `K` $= 5\,; 20\,; 50\,; 100\,; 500\,; 1000$.
   Que concluez-vous ?

3) Faire varier `n` $= 5\,; 10\,; 20\,; 50\,; 100$.
   Que concluez-vous ?