TD 1 : Statistique descriptive à une variable

1 Calcul matriciel avec R

La documentation en ligne R se trouve à l'adresse suivante : `http://stat.ethz.ch/R-manual/`.

Équivalences sous R :

vecteur colonne $u = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$	`u = c(a,b,c,d)`
matrice nulle de taille $m \times n$	`A = matrix(0, m, n)`
$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$	`A = matrix(c(a,b,c,d),2,2)`
accès à la valeur i, j : $(A)_{i,j}$	`A[i,j]`
matrice transposée : A^T	`t(A)`
somme : $A + B$	`A + B`
produit matriciel : $A \cdot B$	`A %*% B`
déterminant de $\det(A)$	`det(A)`
matrice inverse : A^{-1}	`solve(A)`
valeurs et vecteurs propres de A	`eigen(A)$values; eigen(A)$vectors`

Exercice 1 : Une matrice A est inversible si et seulement si $\det(A) \neq 0$.
Calculez le déterminant et la matrice inverse (si elle existe) des matrices suivantes puis vérifiez vos résultats en utilisant R.

- $A = \begin{pmatrix} -1 & 3 \\ 2 & 5 \end{pmatrix}$
- $B = \begin{pmatrix} -1 & 0 & 2 \\ -2 & 3 & 4 \\ 0 & 1 & 1 \end{pmatrix}$

<u>Indications :</u>

- Pour le calcul du déterminant, on pourra utiliser la formule de Laplace :
  Pour $A \in \mathcal{M}_n(\mathbb{R})$, on choisit une ligne i (resp. une colonne).

$$\det(A) = \sum_{j=1}^{n} (A)_{i,j} \mathrm{Cof}_{i,j} \tag{1}$$

  Avec $\mathrm{Cof}_{i,j} = (-1)^{i+j} D_{i,j}$ où $D_{i,j}$ est le déterminant de la matrice A sans la ligne i ni la colonne j.
- Pour calculer l'inverse d'une matrice, on pourra utiliser la transformation de Gauss-Jordan :
  La méthode consiste à créer une nouvelle matrice $(A|I)$ puis à calculer $(I|A^{-1})$ en appliquant le pivot de Gauss. (I correspond à la matrice identité de même dimension que A).

Exercice 2 : Ecrivez les systèmes d'équations linéaires suivants sous forme de produit matriciel puis résolvez les avec R :

$$\begin{cases} 2x - y + 4z + u &= 2 \\ -x + 3y - z + 2u &= 0 \\ 5x - 2z + 6u &= -1 \\ 4x - 2y + z &= 3 \end{cases} \qquad \begin{cases} x + 3y + z &= -1 \\ x + y - z &= 3 \\ 2x + 4y + 3z &= 4 \end{cases}$$

Exercice 3 : Calculez les valeurs et vecteurs propres des matrices suivantes. Vérifiez avec R que $A = U\Lambda U^{-1}$, U étant la matrice des vecteurs propres u_i et Λ la matrice diagonale des λ_i.

- $C = \begin{pmatrix} 1 & -3 \\ -2 & -5 \end{pmatrix}$
- $D = \begin{pmatrix} 2 & -5 \\ 0 & 1 \end{pmatrix}$

Indications :
- Les valeurs propres λ_i sont solutions du système $\det(A - \lambda I) = 0$,
- les vecteurs propres u_i vérifient $A \cdot u_i = \lambda_i u_i$.

2 Statistique descriptive

2.1 Chargement de données et visualisation

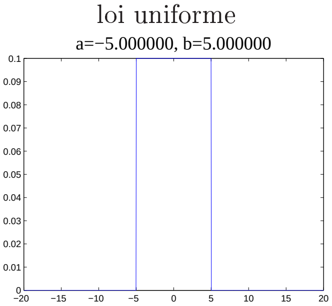
Le chargement et la sauvegarde d'un fichier csv[1] contenant une variable x s'effectue à l'aide des fonctions suivantes : `x = read.csv('fichier.csv')` ou `write.csv(x, 'fichier.csv')`

Exercice 4 : Récupérez l'archive située à l'adresse `http://www.labri.fr/~fourer/Ens/1112/stats/data1.tar.gz` et récupérez les données dans les vecteurs x, y et z puis affichez leur contenu à l'aide de la fonction `plot()`.

2.2 Loi de distribution d'un échantillon

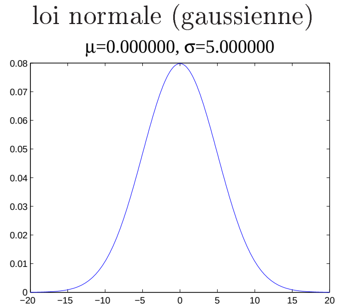
Quelques exemples de lois mathématiques :

loi uniforme



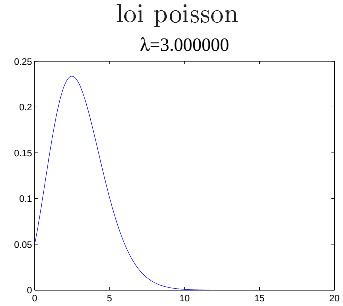
$$U_{a,b}(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

loi normale (gaussienne)



$$\mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

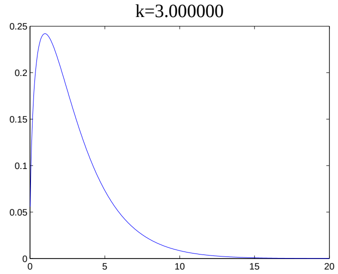
loi poisson



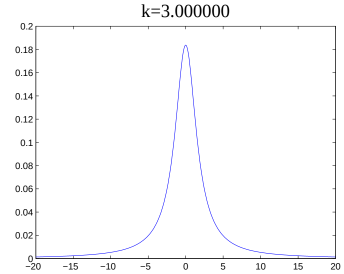
$$P_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

[1] fichier texte dont les champs sont séparés par une virgule

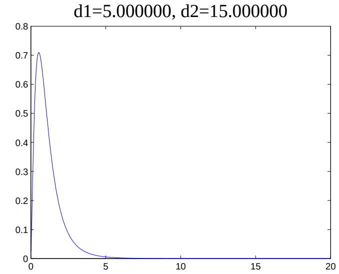
loi χ^2



loi Student



loi Fischer



$$\chi_k(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \qquad \mathcal{S}_k(x) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{(1+\frac{x^2}{k})^{\frac{k+1}{2}}} \qquad F_{d_1,d_2}(x) = \frac{\left(\frac{d_1\, x}{d_1\, x+d_2}\right)^{d_1/2} \left(1-\frac{d_1\, x}{d_1\, x+d_2}\right)^{d_2/2}}{x\ \mathrm{B}(d_1/2,d_2/2)}$$

Avec $\Gamma(x) = \displaystyle\int_0^{+\infty} t^{x-1}\, e^{-t}\, \mathrm{d}t$ et $B(x,y) = \displaystyle\int_0^1 t^{x-1}(1-t)^{y-1}\, dt$.

Exercice 5 : Calculez la moyenne, la variance, l'écart type et les valeurs minimales et maximales pour chacun des vecteurs x, y et z.

Exercice 6 : Affichez l'histogramme des vecteurs x, y et z à l'aide de la fonction `hist()`. Estimez le mode (valeur prise le plus fréquemment) pour chacun des vecteur. Quelle(s) hypothèse(s) peut on faire concernant leur loi de distribution ?

2.3 Utilisation des fonctions

La déclaration d'une nouvelle fonction sous R se fait comme suit :

```
maFonction <- function(parametre1, parametre2)
{
 ...
 resultat = calcul(parametre1, parametre2)
 ...
 return(resultat)
}
```

Pour gagner en clarté vous pourrez déclarer chaque nouvelle fonction dans un nouveau fichier `mafonction.r` . Vous pourrez alors initialiser votre fonction via la fonction `source('mafonction.r')` , afin de la rendre utilisable depuis la console ou depuis une autre fonction.

Exercice 7 : A partir des paramètres estimés et de vos hypothèse, tracez la distribution théorique correspondante à chaque vecteur. Vous aurez alors besoin de déclarer la fonction qui calcule pour un valeur donnée sa probabilité d'apparition pour une loi de distribution donnée. (Votre fonction pourra prendre en entrée un vecteur)

Exercice 8 : Donnez une approximation du 1er et le 3eme quartile respectivement pour les lois de distribution théoriques. Le 1er et le 3eme quartile sont respectivement les valeurs x_1 et x_3 telles que $F(x_1) = 1/4$ et $F(x_3) = 3/4$ où F est la fonction de répartition de la distribution f.
$F(x) = \displaystyle\int_{-\infty}^{x} f(t)dt.$

<u>Indication :</u>
Vous pourrez vous aider de la fonction `cumsum()`