UNIVERSITÉ DE NANTES

Master I : Ingénierie mathématique

Outils informatiques pour la statistique

Exercices

Anne Philippe

Le logiciel utilisé pendant les séances de TP est le logiciel libre R.

14 septembre 2011
Bureau 118 bâtiment 10
Laboratoire de Mathématiques Jean Leray
anne.philippe@univ-nantes.fr
http://www.math.sciences.univ-nantes.fr/~philippe
http://www.math.sciences.univ-nantes.fr/~philippe/R_freeware.html

1. Les objets : vecteurs/matrices

Exercice 1.1. On définit

```
x=c(1,3,5,7,9)
y=c(2,3,5,7,11,13)
```

Donner le résultat des commandes R suivantes

1) x+1
2) y*2
3) length(x) et length(y)
4) x+y
5) sum(x>5) et sum(x[x>5])
6) sum(x> 5 | x< 3)
7) y[3]
8) y[-3]
9) y[x]
10) (y>7) et y[y>7]

Exercice 1.2.

1) Créer les vecteurs suivants :
   – y_0 est constitué de la suite des entiers de 0 à 8 par pas de 2.
   – y_1 contient tous les entiers impairs entre 1 et 18.
   – On pose $d = 4$, y_2 contient trois fois la valeur de d, puis trois fois celle de d^2, puis trois fois celle de $\sqrt{d}$,
   – y_2 est une suite arithmétique prenant ses valeurs entre 1 et 20 avec un pas de deux.
   – y_3 contient 10 chiffres compris entre 1 et 30 avec un intervalle constant.
2) Extraire de y_3
   – le 3 ème élément
   – tous les éléments sauf le 3 ème
   – tous les éléments inférieurs à 12
3) Comparer les commandes suivantes
   – matrix(y3,nrow=2)
   – matrix(y3,nrow=2,byrow=T)
4) Construire une matrice qui contient sur la première ligne y_0 et sur la seconde y_1.

Exercice 1.3.

1) Créer le vecteur qui contient les valeurs 2 5 4 10 8.

2) En utilisant les fonctions vectorielles
   – diviser par 2 chacune des valeurs
   – calculer la racine carré de chacune des valeurs

Exercice 1.4.

1) Construire la matrice Z suivante :

$$Z = \begin{pmatrix} 2 & 4 & 1 & 2 \\ 3 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 \\ 11 & 12 & 35 & 7 \end{pmatrix}$$

2) Afficher l'élément de Z contenu dans
   - la première ligne et troisième colonne
   - La première ligne de Z
   - La troisième colonne de Z
   - la sous-matrice après avoir enlevé la première ligne et la première colonne de Z

Exercice 1.5. Construire une liste qui contient les éléments Z, y_1 et y_3.

Exercice 1.6.

1) Sans faire de boucles et en utilisant uniquement des fonctions déjà existantes :
   a) construire la matrice symétrique M de dimension 10 définie par $M_{i,j} = 1/\sqrt{|i-j|+1}$.
   b) construire la matrice N de dimension 10 définie par $N_{i,j} = i/j^2$.
2) Calculer le déterminant de la matrice M.
3) Si M est inversible, calculer son inverse.
4) Calculer les valeurs propres et les vecteurs propres de la matrice M.

Exercice 1.7. On considère une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ à valeurs dans $\{1,2\}$. La loi de X_0 est la loi uniforme sur $\{1,2\}$.

Conditionnellement à l'évènement $X_n = 1$, la loi de X_{n+1} est

$$P(X_{n+1} = 1 | X_n = 1) = \frac{1}{3} = 1 - P(X_{n+1} = 2 | X_n = 1)$$

et conditionnellement à l'évènement $X_n = 2$, la loi de X_{n+1} est définie

$$P(X_{n+1} = 1 | X_n = 2) = \frac{3}{4} = 1 - P(X_{n+1} = 2 | X_n = 2)$$

1) Construire la matrice P définie par $P(i,j) = P(X_{n+1} = j | X_n = i)$ et la matrice ligne V contenant la loi de X_0
2) Donner l'expression de la loi de X_1 en fonction de P et V (Utiliser la formule des probabilités totales) .
3) En déduire l'expression de la loi de X_k en fonction de P et V
4) Applications numériques
   a) Calculer la loi de X_k pour $k = 1....10$.
   b) Refaire les calculs précédents en changeant la loi de X_0
   c) Commenter les résultats

Exercice 1.8. On considère deux variables aléatoires X et Y indépendantes. X est à valeurs dans $\{0, ..., n\}$ et Y est à valeurs dans $\{0, ..., m\}$.

1) Exprimer la loi de $Z = X + Y$, on note $p_Z(k) = P(Z = k)$
2) Programmer le calcul de la loi de Z en utilisant la fonction `convolve` avec l'option `type='open'`.
3) Tester votre fonction sur les exemples suivants
   a) X suit une binomiale $B(4, p)$ et Y suit une binomiale $B(5, p)$.
      – Quelle est la loi de $X + Y$.
      – Calculer les valeurs de $p_Z(k)$ en utilisant votre fonction.
      – Calculer les valeurs de $p_Z(k)$ en utilisant la fonction `dbinom` (voir l'aide). Comparer.
   b) $n = 2$ et $m = 3$.

k	0	1	2	3
P(X=k)	1/2	1/4	1/4	
P(Y =k)	1/6	1/6	1/6	1/2

Calculer les valeurs de $p_Z(k)$ en utilisant votre fonction. On peut vérifier facilement si les valeur de $P(Z = 0)$ et $P(Z = 5)$ sont correctes.

2. Graphiques

Exercice 2.1.

1) Construire une fonction qui calcule les valeurs de la fonction f définie par
$$f : x \mapsto \sin(x)^2 + \sqrt{|x - 3|}$$
2) Tracer la courbe représentative de la fonction f sur le domaine $[-6, 3]$.
3) Donner une valeur approchée de l'intégrale de la fonction f sur $[-6, 3]$.
4) Utiliser la fonction `ifelse` pour construire une fonction qui calcule la valeur de la fonction
$$g : x \mapsto \begin{cases} \sin(x)^2 \log(x) & x > 0 \\ \sin(x)^2 x & x \leq 0 \end{cases}$$
5) Tracer la courbe représentative de la fonction g sur $[-\pi, \pi]$

Exercice 2.2. On considère la fonction f définie sur $\mathbb{R}^2$ par
$$f(x, y) = \exp\{-10(x^2 - y)^2 - (y - 1/4)^4\}$$

-1- Écrire une fonction qui calcule les valeurs de f sur une grille définie par les vecteurs x et y. Indication : Utiliser la fonction `outer`

-2- Proposer différentes représentations graphiques de la fonction f à l'aide des fonctions `image` ou `persp`

3. Itérations

Exercice 3.1. Construire les fonctions suivantes de deux façons différentes

(a) En utilisant une boucle

(b) Sans utiliser de boucle

1) Construire une fonction qui compte le nombre d'occurrences d'un nombre donné dans un vecteur.
2) Construire une fonction qui retourne les indices des coordonnées d'un vecteur qui sont égales à un nombre donné.
3) Construire une fonction qui teste la présence d'un nombre donné dans un vecteur.

Exercice 3.2. Soit $N = (N_{i,j})$ une matrice. On note $N_{i.}$ la somme des termes de la i ème ligne, $N_{.i}$ la somme des termes de la i ème colonne et n la somme des termes de la matrice.

Sans utiliser de boucles, construire une fonction qui retourne la quantité suivante

$$\sum_i \sum_j \frac{\left(N_{i,j} - \dfrac{N_{i.}N_{.j}}{n}\right)^2}{\dfrac{N_{i.}N_{.j}}{n}}$$

Exercice 3.3. On construit une partition de l'intervalle $]0,1]$ en prenant $\bigsqcup_{i=1}^p A_i$ avec $A_1 =]0, a_1]$, $A_i =]a_{i-1}, a_i]$ pour $i = 2...p-1$ et et $A_p =]a_{p-1}, 1]$.

On note $a = (a_1, \ldots, a_{p-1})$ un vecteur sans répétition et ordonné par ordre croissant.

```
ind = function(x,a)
{
if (length(x)!= 1 |  x<= 0 | x>1) stop('x n est pas réel dans ]0,1]')
sum(x-a > 0) +1
}

find = function(x,a)
{
sapply(x,ind,a)
}
```

1) Expliquer le code et préciser
   - Quel objet retourne ind(x,a) où x un réel
   - Quel objet retourne find(x,a) où x un vecteur

4. Autour des lois de probabilités

Exercice 4.1. Soit X une variable aléatoire qui suit une loi normale de moyenne 0 et de variance 3 .

1) Calculer la probabilité des événements suivants $[X \leq 1]$; $[X \geq 2.6]$; $[0.5 < X \leq 1]$;
2) Calculer le quantile d'ordre $a = 0.75$, c'est à dire la valeur de x telle que
$$P(X \leq x) = a$$
3) Représenter graphiquement la densité et la fonction de répartition de la loi de X
4) Simuler un échantillon $(x_1; \ldots; x_{100})$ de taille $n = 100$ suivant la loi de X
5) Représenter le nuage de points $\{(x_i, x_{i+1}), i = 1, ..., n-1\}$.
6) Créer une liste qui contient la moyenne empirique $\frac{1}{100}\sum_{i=1}^{100} x_i$, le minimum de $(x_1; \ldots; x_{100})$ et le maximum de $(x_1; \ldots; x_{100})$.

Exercice 4.2.

1) Simuler une matrice aléatoire de dimension 2×2 dont les termes sont des variables aléatoires iid suivant la loi gaussienne standard $\mathcal{N}(0, 1)$
2) Calculer les valeurs propres de cette matrice
3) Recommencer N= 1000 fois les étapes précédentes et stocker uniquement les valeurs propres dans une matrice $N \times 2$
4) Calculer la proportion des matrices qui possédent deux valeurs propres réelles.
5) Construire deux matrices, l'une contenant les valeurs propres réelles et l'autre les valeurs propres imaginaires.
6) Sur le sous ensemble des matrices avec deux valeurs propres réelles, construire un vecteur $VPsup$ qui contient la plus grande valeur propre de ces matrices
7) Calculer la moyenne $m.vp$ et la variance $v.vp$ des valeurs du vecteur $VPsup$
8) Tracer l'histogramme des valeurs de $VPsup$.
9) Superposer la densité de la loi gaussienne de moyenne $m.vp$ et de variance $v.vp$. Commenter
10) Tracer l'histogramme du module et de l'argument dans le cas imaginaire.

Exercice 4.3.

1) Soit x un vecteur de dimension n. On supprime la i ème coordonnées de x et on note x_{-i} le vecteur obtenu. Écrire une fonction V qui calcule la norme euclidienne de x_{-i}, les paramètres d'entrée sont i et x.
2) Construire une fonction qui retourne le vecteur $W = (V(x_{-1}), \ldots, V(x_{-n}))$
   (a) En utilisant une boucle
   (b) Sans utiliser de boucle

3) Appliquer vos fonctions sur un vecteur x constitué de nombres aléatoires simulés suivant la loi uniforme sur $[0, 1]$ (voir `runif`).
4) Tracer l'histogramme de W. Commenter.

Exercice 4.4. Soit P_n polynôme de degré n à coefficient aléatoire,

$$P_n(z) = a_1 + a_2 z + ... + a_n z^{n-1} + z^n$$

avec $(a_1, ..., a_n)$ des variables aléatoires indépendantes et identiquement distribuées suivant la loi gaussienne standard. On note $Z = (Z_1, ..., Z_n)$ les racines du polynômes P_n.

On fixe $n = 10$.

1) Construire une fonction qui retourne une réalisation de $Z = (Z_1, ..., Z_n)$. Le paramètre d'entrée de la fonction est n. *Indication :* Utiliser la fonction `polyroot`
2) Représenter graphiquement les racines du polynôme. *Indication :* Si z est un vecteur de complexes $z = x + iy$ la fonction `plot(z)` représente le nuage de points $\{(x_i, y_i)\, i = 1...n\}$. C'est donc équivalent à `plot(x,y)`
3) A l'aide d'une boucle `for`, superposer au graphique précédent 10 autres réalisations de $Z = (Z_1, ..., Z_n)$ en utilisant une couleur différente pour chaque réalisation.
4) Ajouter le cercle de rayon 1 et de centre $(0, 0)$.
5) Refaire les questions précédentes avec $n = 40$
6) Commenter les résultats obtenus.

Exercice 4.5. On se donne $x = (x_1, \ldots, x_n)$ un vecteur de longueur n.

1) Sans utiliser de boucles, construire une fonction qui a pour paramètre d'entrée x et retourne le scalaire
$$S_n = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$
2) Construire une fonction qui a pour paramètre d'entrée x et retourne le vecteur $(S_1, \ldots, S_n)$ (Sans utiliser de boucles).
3) Représenter graphiquement $(S_1, \ldots, S_n)$ lorsque le vecteur x est constitué de nombres aléatoires iid suivant la loi uniforme sur $[0, 1]$
4) Construire une matrice de $n \times N$ colonnes constituée de variables aléatoires iid suivant la loi uniforme sur $(0, 1)$
5) Appliquer la fonction programmée à la question 2 sur chacune des colonnes de la matrice. Stocker les résultats dans une matrice $n \times N$ appelée MC
6) Tracer sur un même graphique toutes les colonnes de la matrice MC obtenue à la question 5.
7) Tracer l'histogramme de la n ème ligne de la matrice MC.

Exercice 4.6.

1) Créer une fonction qui pour un couple donné $(n,p) \in \mathbb{N} \times [0,1]$, évalue le maximum de l'erreur commise lorsque l'on approche la loi binomiale par la loi de Poisson
$$M_{n,p} = \max_{k=0,\ldots,n} |P(X_n = k) - P(Y_n = k)|$$
où X_n suit une loi binomiale de paramètres (n,p) et Y_n suit une loi de poisson de paramètre np.
2) Pour $p = 1/2$, représenter graphiquement l'erreur en fonction de n
3) Pour $n = 40$, représenter graphiquement l'erreur en fonction de p

Exercice 4.7. Loi discrète.

1) Générer un échantillon de taille $n = 100$ suivant la loi discrète définie par
$$P(X = j) = p_j$$
pour tout $j = 1, \ldots, 3$ et $p = (0.2,\ 0.6,\ 0.2)$.
   a) Pour tout $j = 1, \ldots, 3$, calculer la proportion d'observations égales à j dans l'échantillon simulé (voir la fonction `table`).
   b) Représenter sur un même graphique les proportions et les probabilités p_j ($j = 1, \ldots, 3$). Commenter.
   c) Calculer l'espérance de la loi et comparer avec la moyenne de l'échantillon simulé.
   d) Calculer la variance de la loi et comparer avec la variance de l'échantillon simulé.
2) Recommencer avec $n = 5000$ puis 10000. Commenter.