Master 1 Mathématique
Université d'Angers

Année 2011-2012

TP 1 : initiation à R

R est un logiciel de statistique distribué gratuitement par le CRAN (Comprehensive R Archive Network) à l'adresse suivante http ://cran.r-project.org/. Après seulement une dizaine d'années d'existence, le logiciel R est devenu un outil incontournable de statistique et de visualisation de données, tant dans le monde universitaire que dans le monde de l'entreprise. Ce développement exceptionnel s'explique par ses trois principales qualités :

il est gratuit, très complet, et en essort permanent.

Exercice préliminaire

Lancer R, et exécuter les commandes suivantes une à une, en tentant à chaque fois de comprendre ce qu'elle fait. A tout moment, vous pouvez quitter la session en tapant `q()`. Pour obtenir de l'aide sur une commande, taper help(*commande*). Pour une aide générale, taper `help.start()`.

Manipulations élémentaires de vecteurs, matrices, ...

```
a=c(1,6,-2)
a
a[2]
b=a[c(1,3)]
2*a
log(a)
cos(a)
1/a^2
a/a
is.numeric(a)
a>0
a[a<0]
ab<-c(a,b)
ab
length(ab)
ab[-2]
max(ab)
which.max(ab)
ab=max(ab)
sort(ab)
order(ab)
a[order(a)]
A=seq(0,1,length=4)
A
B=c(1:3,7)
A*B
A0=matrix(c(1:6),ncol=2)
A0
A=matrix(c(1:6),ncol=2,byrow=TRUE)
A[1,2]
```

```
A[,1]
dim(A)
t(A)
C=diag(c(1,2))
B=t(A)%*%A
B
B*B
B[-1,]
B[,-2]
cbind(B,B)
rbind(B,B)
solve(B)
eigen(B)
y=c(1,2)
solve(B)%*%y
x=solve(B,y)
B%*%x
```

Générer de l'aléatoire...

```
X=runif(100)
X[1:6]
mean(X)
x=rnorm(20,1,2)
y=rnorm(20,1,2)
x
y
dnorm(0)
pnorm(1.96)
qnorm(0.95)
punif(0.5)
qunif(0.5)
help.search("normality")
shapiro.test(x)
qqnorm(x)
x <- rnorm(50)
y <- rnorm(50)
plot(x,y)
plot(x,y,xlab="variable x",ylab="variable y",main="mon graphe")
lines(x,x)
hist(x)
plot(density(x))
```

Manipulations de jeux de données, fonctions avancées

```
summary(faithful)
hist(faithful~eruptions)
attach(faithful)
plot(eruptions,waiting)
x <-rnorm(100)
y <- 0.2+0.3*x+0.1*rnorm(x)
linmod <- lm(y ~ x)
summary(linmod)
plot(linmod)
library(datasets)
help(ChickWeight)
attach(ChickWeight)
```

```
linmod2 <- lm(weight ~ Time+Diet)
summary(linmod2)
```

Exercice 1 : Simulations

1. Ecrire une fonction "pileouface(n,p)" qui prend comme entrées un entier n et $0 < p < 1$, et qui donne en sortie un vecteur de taille n dont les coefficients simulent des variables aléatoires indépendantes de Bernoulli de paramètre p.
   Indication : on utilisera le début suivant :

   ```
   pileouface<-function(n,p){
   x=(runif(n)<p)
   ...}
   ```

   et la commande "as.numeric".
   Visualiser dans trois fenêtres graphiques différentes, pour $n = 1000$, $p = 0.5$, $p = 0.2$, et $p = 0.8$, un échantillon de taille 1000 de loi de Bernoulli sur un diagramme en barre.
2. Ecrire une fonction "uniforme(n,a,b)" qui donne un vecteur de taille n dont chaque coefficient simule des variables aléatoires indépendantes de loi uniforme sur l'intervalle $[a, b]$. Visualiser un échantillon de taille 10000 de loi uniforme sur $[a, b]$ sur un histogramme.
3. Ecrire une fonction "de(n)" qui simule n tirages indépendants d'un dé non-pipé.
4. Ecrire une fonction "tir(n,D)" qui donne, pour n entier positif et $D > 0$, une série de n tirs indépendants sur un mur à distance D avec un angle choisi au hasard uniformément sur l'intervalle $]-\frac{\pi}{2}, \frac{\pi}{2}[$.
5. Simuler, toujours à partir de la fonction "runif()", n tirages indépendants de loi normale bidimensionnelle.

Exercice 2 : Loi des Grands Nombres

On veut étudier le comportement asymptotique de $\bar{X}_n$ lorsque $(X_1, \ldots, X_n)$ est un échantillon de loi de Bernoulli de paramètre p.

1. Ecrire une fonction "lgn(n,p)" qui contient l'évolution de $\bar{X}_n$ en fonction de n lorsqu'on dispose de n réalisations de Bernoulli de paramètre p. Tracer la courbe correspondante.
2. Tracer sur une même fenêtre graphique, 10 courbes de l'évolution de $\bar{X}_n$ en fonction de n pour $n = 1000$ et $p = 0.5$.

Exercice 3 : Théorème Limite Central

On veut à présent illustrer le théorème limite central, qui s'intéresse à la loi de probabilité asymptotique de $\bar{X}_n$. On se restreint au cas où $(X_1, \ldots, X_n)$ est un échantillon indépendant de loi de Bernoulli de paramètre p.

1. Ecrire une fonction "tlc(m,n,p)" qui trace l'histogramme de m réalisations de la variable $\sqrt{n}(\bar{X}_n - p)$, auquel on superpose la densité de la loi normale centrée de variance bien choisie.
2. tester la fonction pour plusieurs valeurs de p, puis de n et de m.