Simulation des Lois de Probabilités avec R

Exposé par:

Samah MOUHTADI

Ouakou Perrine Messie

Encadrées par:

Prof. M. El Merouani

Lois usuelles et génération de données aléatoires :
Le logiciel R permet d'effectuer des calculs avec toutes les lois de probabilité usuelles, et aussi de simuler des échantillons issus de ces lois. Le tableau suivant résume les différentes lois implémentées dans R.

Loi	appelation R	Arguments
bêta	beta	forme 1, forme 2
binomiale	binom	size, prob
chi deux	chisq	df (degrés de liberté)
uniforme	unif	min, max
exponentielle	exp	rate
Fisher	f	df1, df2
gamma	gamma	forme, echelle
géometrique	geom	prob
hypergéometrique	hyper	m, n, k (taille échantillon)
binomiale négative	nbinom	size, prob
normale	norm	mean, sd
Poisson	pois	lambda
Student	t	df
Weibull	weibull	forme, echelle

Pour certaines lois, les paramètres ont des valeurs par défaut : parmi les plus utilisées, la loi uniforme unif porte par défaut sur l'intervalle $[0,1]$, et la loi normale norm est centrée réduite par défaut.

Pour effectuer un calcul avec une de ces lois, il suffit d'utiliser comme fonction l'une des appellations R ci-dessus avec le préfixe d pour une densité, p pour une fonction de répartition, qpour une fonction quantile et r pour un tirage aléatoire. Voici quelques exemples :

```
x <- rnorm(100)        # 100 tirages, loi N(0,1)
w <- rexp(1000, rate=.1) # 1000 tirages, loi exponentielle
dpois(0:2, lambda=4)     # probablités de 0,1,2 pour la loi Poisson(4)
pnorm(12,mean=10,sd=2)   # P(X < 12) pour la loi N(10,4)
qnorm(.75,mean=10,sd=2)  # 3ème quartile de  la loi N(10,4)
```

Exemple d'application

Tracés de densités et de fonctions de répartitions

1-Des tracés de densités de probabilité ou de fonctions de répartition de lois diverses peuvent s'obtenir à l'aide de la function plot(). Ainsi par exemple pour tracer la densité (répartition des masses) d'une loi binomiale avec $n = 10$ and $p = .25$, reproduite dans la figure 13, on exécute dans R :

```
x <- 0:10
y <- dbinom(x, size=10, prob=.25)    # évalue les probas
{plot(x, y, type = "h", lwd = 30,
    main = "Densité Binomiale avec \n n
    = 10, p = .25", ylab = "p(x)", lend ="square")}
```

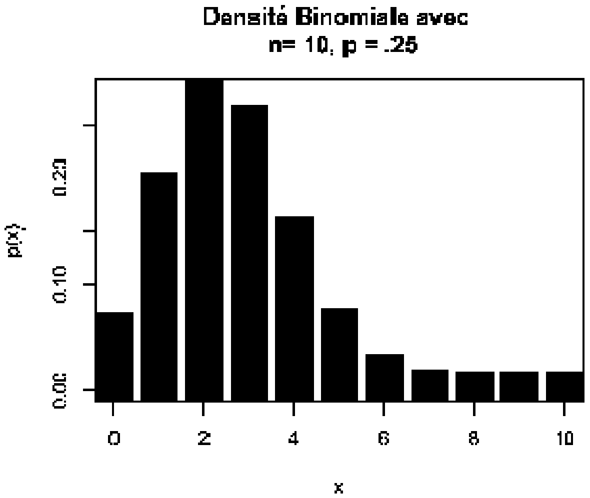

Figure 13: Loi binomiale $B(10, 0.25)$.

Pour cet exemple, nous avons d'abord créé le vecteur x contenant les entiers allant de 0 à 10. Nous avons ensuite calculé les probabilités qu'une variable de loi binomiale $B(10, 0.25)$ prenne chacune de ces valeurs, par dbinom. Le type de tracé est spécifié avec l'option type=h (lignes verticales d'un diagramme en bâtons), épaissies grâce à l'option lwd=30. L'optionlend ="square" permet de tracer des barres rectangulaires. Nous avons ensuite ajouté des légendes.

2-Pour tracer des densités de lois absolument continues ou des fonctions de répartitions de celles-ci, on peut utiliser la fonction curve(). Par exemple, le tracé de la densité d'une loi Gaussienne centrée réduite sur l'intervalle $[-3, 3]$ s'obtient avec la commande curve(dnorm(x), from = -3, to = 3) qui réalise la figure 14.

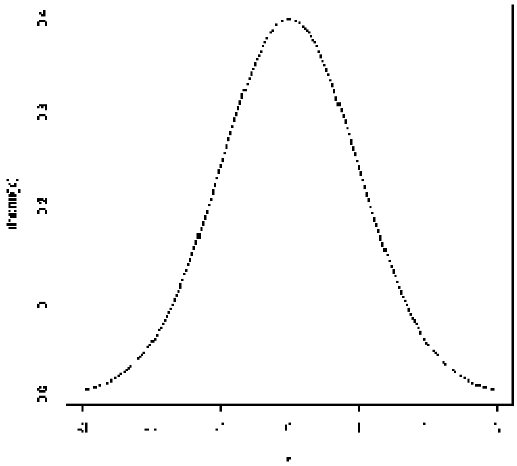

Figure 14: Densité de la loi $\mathcal{N}(0, 1)$

3-alors que la commande curve(pnorm(x, mean=10, sd=2), from = 4, to = 16) produit le tracé de la fonction de répartition d'une loi $\mathcal{N}(10, 4)$ sur l'intervalle $[4, 16]$:

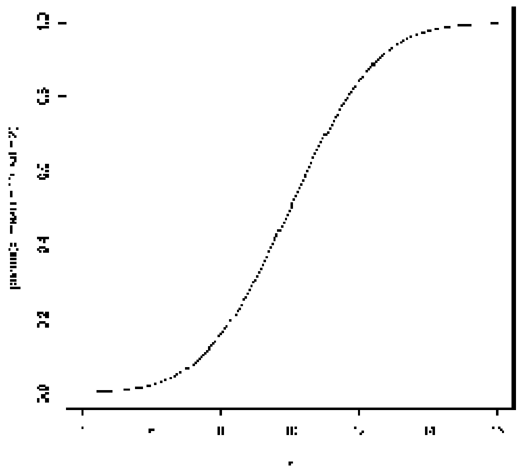

Figure 15: Fonction de répartition de la loi $\mathcal{N}(10, 4)$

Notons que la fonction curve() permet également de superposer une courbe sur un autre tracé (dans ce cas il est inutile de spécifier from et to).

4-Essayons par exemple de comparer l'histogramme des fréquences des valeurs obtenues par un tirage de 1000 nombres selon la loi $\mathcal{N}(0,1)$ avec la densité de la loi $\mathcal{N}(0,1)$:

```
simu <- rnorm(1000)
{hist(simu, prob=T, breaks="FD",
   main="Histogramme de 1000 tirages N(0,1)")}
curve(dnorm(x), add=T)
```

qui donne la figure 16.

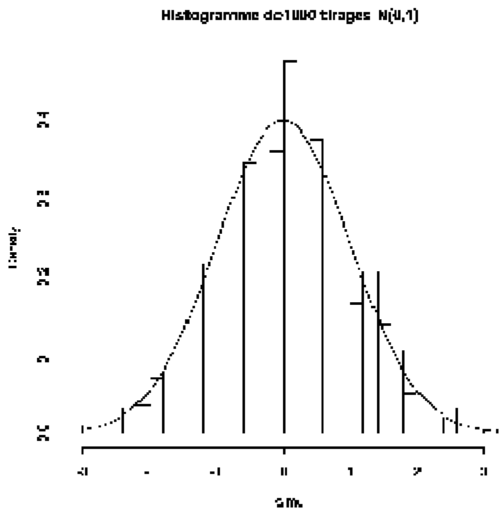

Figure 16: Histogramme d'un échantillon de taille 1000 de la loi $\mathcal{N}(0,1)$.

La loi forte des grands nombres permet d'affirmer que pour une loi admettant une espérance μ *et une variance* σ^2*, la moyenne empirique* $\bar{X}_n$ *converge vers* μ *alors que le théorème de la limite centrale dit que la loi de la moyenne empirique* $\bar{X}_n$ *peut être approchée de mieux en mieux quand la taille de l'échantillon tend vers l'infini par une loi normale d'espérance* μ *et de variance* σ^2/n.

5-On peut illustrer ces propriétés par des expériences aléatoires appropriés, programmées avec R de la manière suivante.

Pour un échantillon de la loi uniforme sur $[0,1]$ de taille $n=1000$, calculons les moyennes empiriques successives et traçons la moyenne empirique en fonction de la taille de l'échantillon. On observe que la moyenne empirique converge bien vers la valeur $\mu=0.5$ représentée par une droite horizontale rouge sur le graphique 17.

```
n<-1000
X<-runif(n)
Y<-cumsum(X)    #somme cummulé
N<-seq(1,n, by=1)
plot(N, Y/N)
abline(h=0.5,col="red")
```

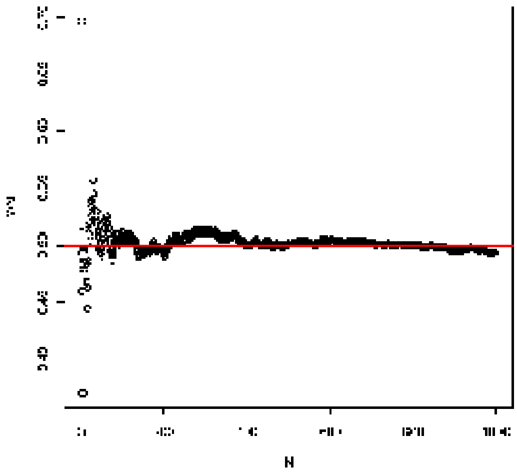

Figure 17: Illustration de la loi des grands nombres