L2 Sciences pour la Santé

Statistiques

TD1 - Simulation de lois de probabilités

Introduction : Générer des nombres aléatoires avec R

Tout comme de nombreux langages de programmation, R offre la possibilité de choisir de générer des nombres aléatoires, ou plus précisément ‘pseudo-aléatoires’. On parle de nombres ‘pseudo-aléatoires’ car le comportement d’un ordinateur n’a justement rien d’aléatoire. Si vous vous posez la question de savoir comment un ordinateur déterministe permet de générer du hasard et bien sachez que c’est une très bonne question mais ce n’est malheureusement pas l’objet de ce TD. Pour votre information, il existe de nombreux algorithmes ‘générateurs’ de nombres pseudo-aléatoires et celui utilisé par défaut dans R est le générateur Marsenne-Twister (il s’agit de l’un des plus avancés). Afin de pouvoir répéter à l’identique des algorithmes contenant des nombres pseudo-aléatoires, les générateurs utilisent une amorce ou (graine) de telle sorte qu’une même graine renverra toujours le même résultat. Dans R, la graine est spécifiée via la fonction set.seed. Dans R, les caractéristiques de nombreuses lois de probabilité sont accessibles via les quatre fonctions suivantes :

1. dloi calcule la fonction de densité de probabilité (loi continue) ou la fonction de masse de probabilité (loi discrète)
2. ploi calcule la fonction de répartition
3. qloi calcule la fonction de quantile
4. rloi simule des observations de cette loi.

Nota bene : pour la loi uniforme, les fonctions sont dunif, punif,. . . et pour la loi normale dnorm, pnorm. . .

Exercice 1 : Simulation d’une loi normale centrée réduite

1. Simuler une valeur aléatoire selon une loi normale centrée réduite

   ```
   > x = rnorm(1,mean=0,sd=1)
   ```

   Refaire la même chose en augmentant le nombre de valeurs (10, 100, 1000, 1 000 000) puis tracer l’histogramme correspondant

   ```
   > x = rnorm(10,mean=0,sd=1)
   > hist(x)
   ```

2. Comparer les moyennes et variances empiriques du vecteur généré avec les valeurs théoriques (0 et 1 respectivement)

```
> mean(x) # espérance théorique = 0
> var(x) # variance théorique = 1
```

3. Comparer l'histogramme obtenu avec la densité théorique

```
> hist(x,freq=FALSE)
> nx <- seq(-100,100,length=5000)
> lines(nx,dnorm(nx,mean=0,sd=1),col="red")
```

4. Comparer les quantiles empiriques aux quantiles théoriques

```
> qqnorm(x, pch = 1, frame = FALSE)
> qqline(x, col = "steelblue", lwd = 2)
```

Exercice 2 : Simulation d'une loi du Khi deux

Soient $X_1, \ldots, X_k$ des variables aléatoires normales centrées réduites et indépendantes. Alors, la variable

$$X = \sum_{i=1}^{k} X_i^2$$

suit une loi du Khi deux à k degrés de liberté. On note $X \sim \chi^2(k)$
On a $E[X] = k$ et $Var[X] = 2k$

1. A partir de 3 variables générées selon une loi normale centrée réduite, simuler 1000 valeurs issues d'une loi du Khi deux à 3 degrés de liberté

```
> n = 1000

> x1 <- rnorm(n,mean=0,sd=1)
> x2 <- rnorm(n,mean=0,sd=1)
> x3 <- rnorm(n,mean=0,sd=1)

> x <- x1^2 + x2^2 + x3^2
```

2. Comparer les moyennes et variances empiriques du vecteur généré avec les valeurs théoriques

```
> mean(x) # espérance théorique = 3
> var(x) # variance théorique = 6
```

3. Comparer l'histogramme obtenu avec la densité théorique

```
> hist(x,freq=FALSE)
> nx <- seq(-100,100,length=5000)
> lines(nx,dchisq(nx,df=3),col="red")
```

4. Comparer les quantiles empiriques aux quantiles théoriques

```
> y <- qchisq(ppoints(length(x)),df=3)
> qqplot(x,y)
```

Exercice 3 : Simulation d'une loi de Student

Soient Z une variable normale centrée réduite et U une variable suivant une loi du Khi deux à k degrés de liberté. Supposons de plus que Z et U sont indépendantes. Alors, la variable

$$T = \frac{Z}{\sqrt{\frac{U}{k}}}$$

suit une loi de Student à k degrés de liberté. On note $T \sim T(k)$
On a $E[T] = 0$ et $Var[T] = \frac{k}{k-2}$
En utilisant la définition ci-dessus, répétez les étapes de l'exercice 2 en posant $k = 50$.

Exercice 4 : Simulation d'une loi de Fisher

Soient U_1 et U_2 deux variables aléatoires indépendantes suivant des lois du khi deux à respectivement k_1 et k_2 degrés de liberté.
Alors, la variable

$$X = \frac{U_1/d_1}{U_2/d_2}$$

suit une loi de Fisher à d_1 et d_2 degrés de liberté. On note $X \sim F(d_1, d_2)$
On a $E[X] = \frac{d_2}{d_2-2}$ et $Var[X] = \frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$
En utilisant la définition ci-dessus, répétez les étapes de l'exercice 2 en posant $d_1 = 10$ et $d_2 = 15$.