TP 4 : Méthode de Box-Muller, risque quadratique

Ce TP n°4 aborde la méthode de Box-Muller vu en TD (exercice 1) et la notion de risque quadratique associé à un estimateur (exercice 2).

Exercice 1 : Méthode de Box-Muller en pratique, et digression. Nous rappelons ici la méthode de Box-Muller vue en TD. L'idée est de générer deux variables aléatoires X, Y, i.i.d. de loi $\mathcal{N}(0,1)$, de la forme $X = r\cos\theta, Y = r\sin\theta$, où r^2 suit une loi exponentielle de paramètre $1/2$, et θ suit la loi uniforme sur $[0, 2\pi]$.

1. *En utilisant uniquement des lois uniformes sur* $[0,1]$, simulez $n = 10^5$ réalisations de X et Y à l'aide de la méthode de Box-Muller. *Indication : pensez à utiliser l'inversion de la fonction de répartition, comme dans le TP2 !.*

2. Simuler le même nombre de réalisations de $Z \sim \mathcal{N}(0,1)$ à l'aide de `rnorm`. Tracer les histogrammes (séparés, avec des couleurs différentes) pour les réalisations de X, Y et Z. Qu'observe-t-on ?

3. Pour cette question, on besoin du package `rgl`, permettant d'afficher des données en 3 dimensions et de les visualiser. On l'installe avec la commande

`install.packages("rgl")`.

Une fois l'installation effectuée, on importe le package, avec la commande

```
library(rgl)
```

Nous donnons à présent le code suivant. Que fait ce code ? Exécutez-le et observer le résultat.

```
n=1000
x = rnorm(n)
y = rnorm(n)
z = rnorm(n)
normes = sqrt(x^2+y^2+z^2)
plot3d(x/normes,y/normes,z/normes,type="s",radius=0.05,col="red")
```

Que peut-on conjecturer sur la loi du vecteur

$$\left(\frac{X_1}{\sqrt{X_1^2+X_2^2+X_3^2}}, \frac{X_2}{\sqrt{X_1^2+X_2^2+X_3^2}}, \frac{X_3}{\sqrt{X_1^2+X_2^2+X_3^2}}\right),$$

lorsque X_1, X_2, X_3 sont trois gaussiennes centrées réduites indépendantes ?

Exercice 2 : Estimation et risque quadratique Dans cet exercice, on travaille avec le modèle paramétrique suivant : X suit une loi uniforme sur l'intervalle $[0, a]$ où a est un paramètre > 0. On cherche à estimer a. On suppose que l'on dispose d'un échantillon de copies i.i.d. de X, noté $X_1, \ldots, X_n$. On notera comme dans le cours

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{et } \tilde{S}_n^2 := \frac{n}{n-1} S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$

les estimateurs de la moyenne et de la variance sans biais. *Pour l'exercice, on rappelle qu'en R, comme vu en TP3, la commande* `var` *pour un échantillon calcule directement* $\tilde{S}_n^2$ *(et fait donc par défaut une estimation sans biais).*

1. Donner deux estimateurs de a : l'un utilisant $\bar{X}_n$, que l'on notera $\hat{a}_1$, et l'autre utilisant $\tilde{S}_n^2$, noté $\hat{a}_2$.

2. On cherche à représenter le risque quadratique pour un a fixé ($\mathbb{E}\left[(\hat{a}-a)^2\right]$ pour rappel), pour les deux estimateurs. Pour cela, on définit une fonction $R1$ qui revoie le biais estimé, la variance estimée et enfin le risque quadratique estimé de $\hat{a}_1$ en simulant un nombre N d'échantillons indépendants de taille n.

```
R1 <- function(a,n,N){
  estims = replicate(N,0) # on initialise
  for (i in 1:N){
    ### A COMPLETER
    # X =  # on simule un échantillon de taille n de loi U([0,a])
    # estims[i] =  # on calcule hat(a)_1
  }
  ### A COMPLETER
  # biais =
  # variance
  # risque =
  c(biais, variance, risque) # on retourne le vecteur des trois valeurs
}
```

Compléter cette fonction, et la tester. *Indication : on pourra prendre* $n = 2000$ *et* $N = 5000$.

Redéfinir cette fonction (et la renommant par exemples $R12$) pour qu'elle renvoie aussi le biais estimé, la variance estimée et enfin le risque quadratique estimé de $\hat{a}_2$. *Indication : on pourra encapsuler les trois valeurs à l'aide de* `c(...)` *puis utiliser* `rbind` *pour renvoyer "l'union des deux vecteurs".*

La tester. Quel semble être le meilleur estimateur au sens du risque quadratique entre $\hat{a}_1$ et $\hat{a}_2$?

3. Quelle relation y a-t-il entre biais, variance et risque quadratique pour un estimateur ?

4. On introduit à présent un troisième estimateur $\hat{a}_3$ définit par

$$\hat{a}_3 := \max_{1\leq i\leq n} X_i.$$

Adapter la fonction précédente (la renommer par exemple $R123$) pour qu'elle renvoie aussi le triplet biais/variance/risque associé à $\hat{a}_3$.

La tester à nouveau. Commenter les résultats à la lumière de la question 3.

5. Pour a variant de 0.5 à 8 (*indication : on utilisera* `seq`), compléter le code ci-dessous pour représenter sur un même graphique les risques quadratiques (estimés) en échelle logarithmique pour les trois estimateurs.

Remarquer la syntaxe pour superposer des plots, mettre une légende, des couleurs différentes, etc.

```
### A COMPLETER
# x = # on définit les abscisses (valeurs de a)
taille = length(x)
risques1 = replicate(taille,0) # on initialise
risques2 = replicate(taille,0) # on initialise
risques3 = replicate(taille,0) # on initialise

for (i in 1:taille){
```

```
  ### A COMPLETER
  # res = # on fait appel à R123 avec a=x[i],n=2*10^3, N=5*10^3
  ## on stocke les variables qui nous intéressent :
  # risques1[i] =
  # risques2[i] =
  # risques3[i] =
}

### A DECOMMENTER (sélectionner puis Ctrl+C)
## puis, on réalise le plot
# plot(x,log(risques1),type="o",col="red",xlab = "Valeur de a",
#      ylab = "log-risque quadratique",main = "log-risques des trois estimateurs",
#      lty = 2, ylim=c(-16.2,-4.8))
# points(x, log(risques2), col="blue",pch="o")
# lines(x, log(risques2), col="blue", lty = 3)
# points(x, log(risques3), col="green",pch="o")
# lines(x, log(risques3), col="green", lty = 4)
# legend(0.6,-4.9,legend=c("Estimateur 1","Estimateur 2","Estimateur 3"),
#        col=c("red", "blue","green"),lty=2:4)
```

Commenter le résultat obtenu.

Pour celleux qui veulent aller plus loin : Monte-Carlo On cherche à estimer l'intégrale

$$I = \int_0^{+\infty} \frac{e^{-x}}{1+x^7}$$

à l'aide de simulations (on ne sait pas la calculer en primitivant cette horreur !).

Comment faire ?