

## TP 2 : Loi des grands nombres, Théorème central limite

Ce TP n°2 continue de nous familiariser avec la simulation sous R, en abordant la Loi des grands nombres (LGN) et le Théorème central limite (TCL).

### Exercice 1 : Loi des grands nombres et Théorème central limite

1. La LGN affirme que pour toute suite  $(X_n)_n$  de variables i.i.d. d'espérance finie  $\mu$ , on a la convergence presque sûre

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mu$$

Le TCL donne un "premier ordre" dans la LGN, sous les mêmes hypothèses, en supposant en plus que la loi est de variance finie  $\sigma^2$ . Dans ce cas, on a la convergence en loi suivante :

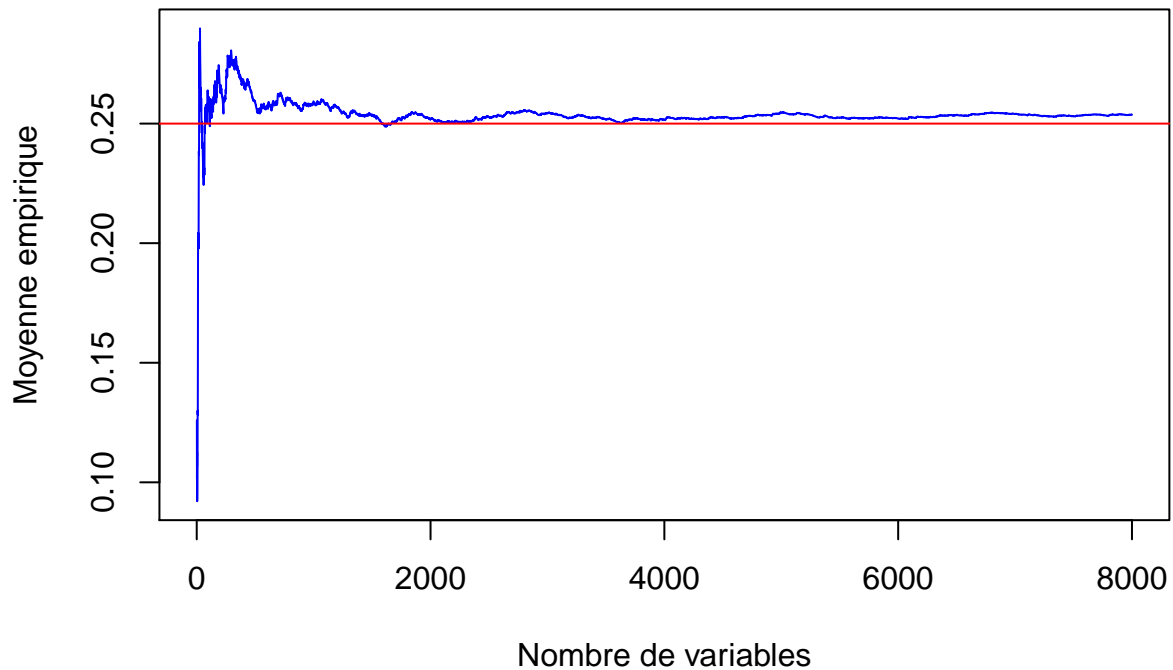
$$\sqrt{n} \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sigma} \rightarrow \mathcal{N}(0, 1).$$

2. On propose un premier code avec une boucle for:

```
n = 8000
X = rexp(n,rate=4) # on tire nos variables
Y = replicate(n,0) # on initialise le vecteur Y de taille n avec des zéros
for (k in 1:n){
  Y[k] = sum(X[1:k])/k # chaque terme est la somme partielle des k premières
  # valeurs de X, divisés par k
}

plot(1:n,Y,type="l",col="blue",main="Loi des grands nombres",
     xlab="Nombre de variables",ylab="Moyenne empirique")
abline(a=0.25,b=0,col="red") #code pour tracer la droite y = 0.25
```

## Loi des grands nombres

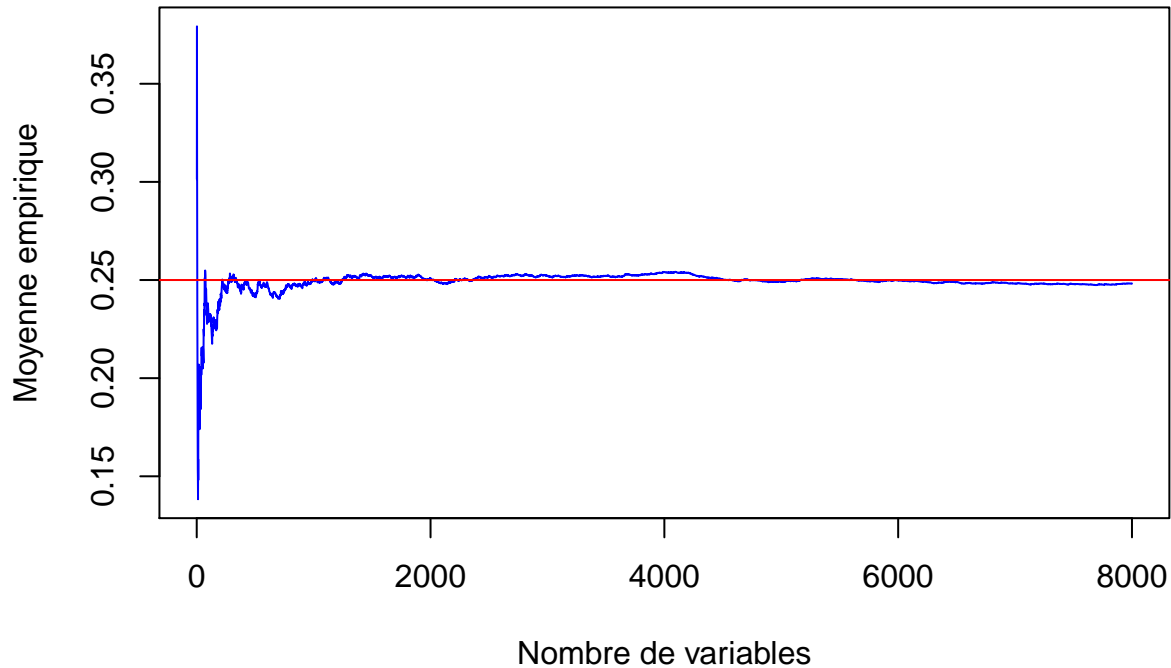


Voici un code plus optimisé qui n'utilise pas de boucle for, grâce à R qui comprend très bien la division de deux vecteurs.

```
n = 8000
X = rexp(n,rate=4)
Y = cumsum(X)/1:n # Y_k sera ainsi la somme partielle des X de rang k,
# divisé par le terme de rang k de 1:n, c'est à dire k !

plot(1:n,Y,type="l",col="blue",main="Loi des grands nombres",
     xlab="Nombre de variables",ylab="Moyenne empirique",)
abline(a=0.25,b=0,col="red")
```

## Loi des grands nombres



Réponse : dans tous les cas, la suite des  $Y_k$  converge vers l'espérance de  $X$  : 0.25. C'est la loi des grands nombres.

3. Rappelons que la variance d'une loi exponentielle de paramètre  $\lambda$  est  $1/\lambda^2$  qui vaut ici  $1/16$ . Donc  $\sigma = 1/4$ .

On a bien envie de superposer la courbe de la densité gaussienne.

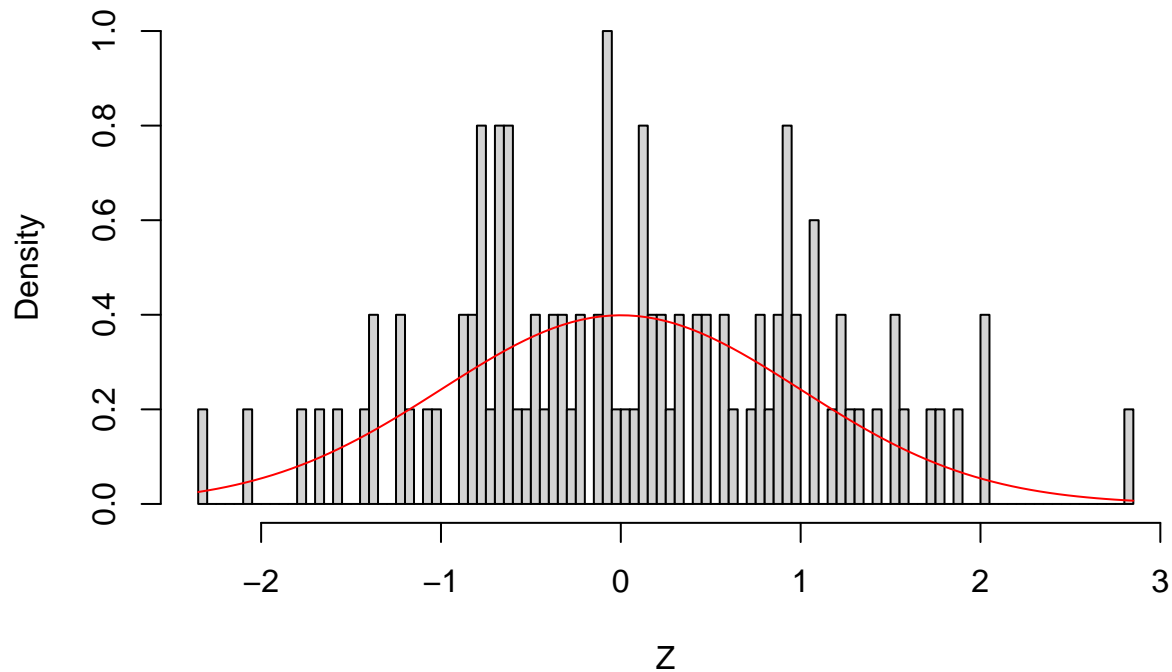
Voici un premier code qui simule une matrice tous les  $X_{i,j}$  puis qui utilise une boucle for pour remplir le vecteur  $Z$ .

```
N = 100
n = 100
mu = 0.25
sigma = 0.25
X = matrix(rexp(n*N,rate=4),nrow=N,ncol=n) # On simule directement tous les  $X_{ij}$ , et
# on les staocke dans une matrices de taille  $N*n$ 
Z = replicate(N,0) # on intialise Z avec que des zéros, vecteur de taille N

for (i in 1:N){
  Z[i] = sqrt(n)*(sum(X[i,])/n - mu)/sigma # on calcule  $Z_i$  avec la ligne i de la matrice X
}

hist(Z,freq=F,breaks = 100, main="Théorème central limite")
curve(dnorm(x,mean=0,sd=1),add=T, col="red") # on superpose à l'histogramme
```

## Théorème central limite



```
# la densité d'une gaussienne centrée réduite
```

On donne une deuxième façon de faire : on remplit  $Z$  en simulant à chaque fois un nouvel échantillon de taille  $n$ .

```
N = 100 # le nombre de variables  $Z_i$  que l'on veut
```

```
n = 100 # la taille de chaque échantillon
```

```
mu = 0.25
```

```
sigma = 0.25
```

```
Z = replicate(N,0) # on initialise Z avec que des zéros, vecteur de taille N
```

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

```
  echantillon = rexp(n,rate=4)
```

```
  Z[i] = sqrt(n)*(sum(echantillon)/n - mu)/sigma # on calcule  $Z_i$  avec
```

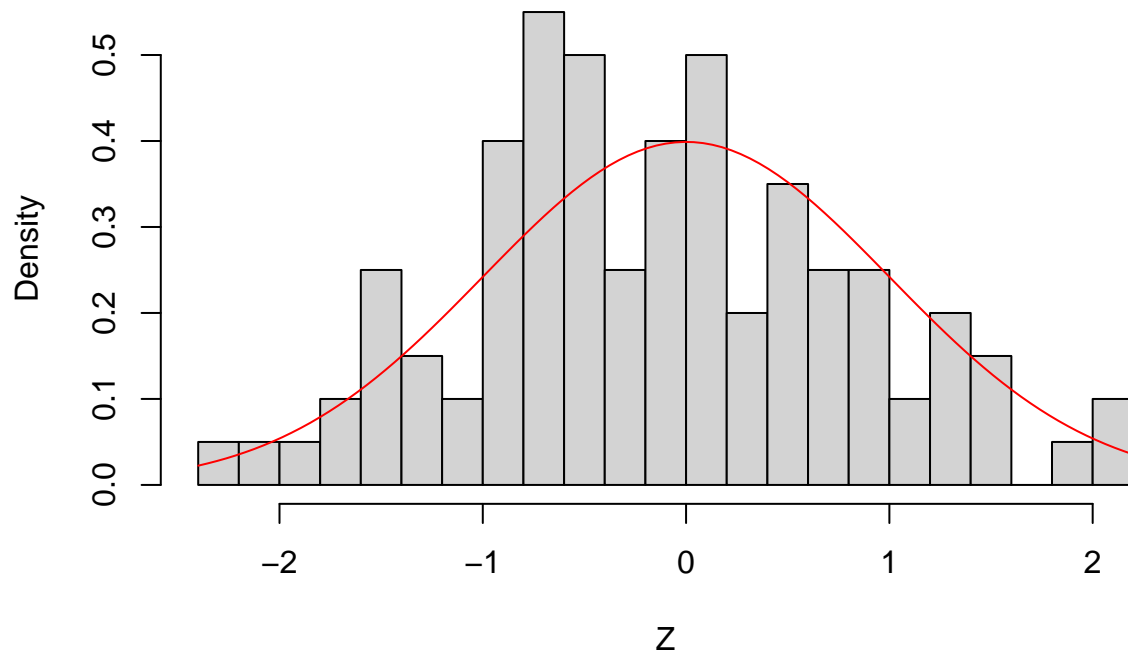
```
  # une nouvelle simu de n variables
```

```
}
```

```
hist(Z,freq=FALSE,breaks = 25, main="Théorème central limite")
```

```
curve(dnorm(x,mean=0,sd=1),add=TRUE, col="red") # on superpose à l'histogramme
```

## Théorème central limite



```
# la densité d'une gaussienne centrée réduite
```

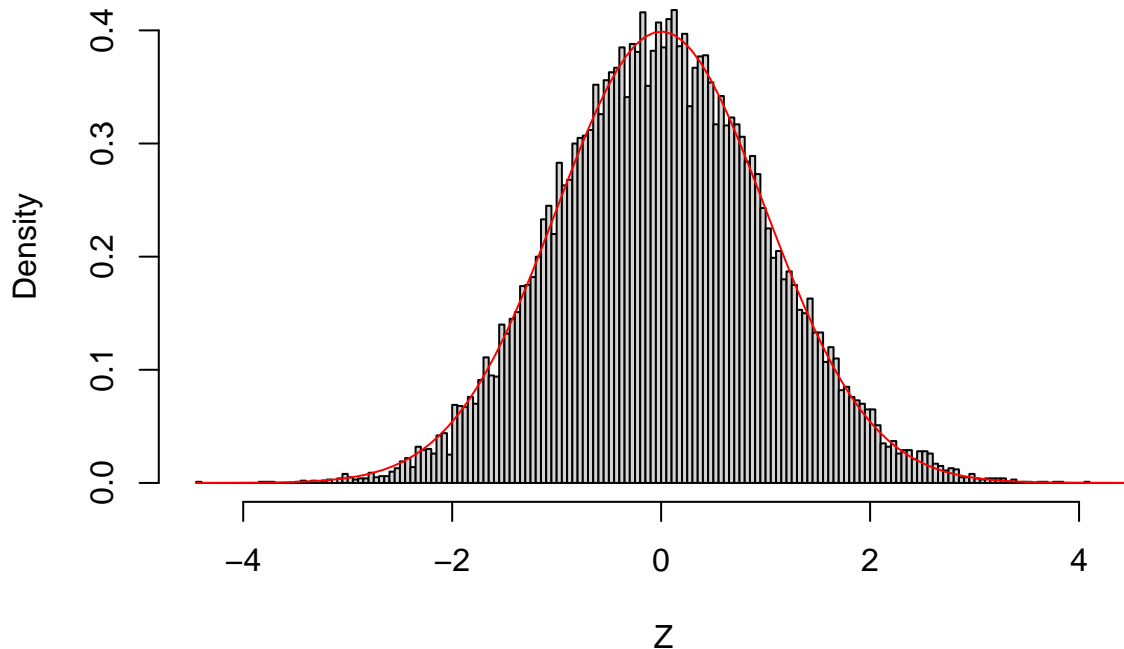
Dans tous les cas, on ne voit pas grand chose... Refaisons l'expérience en augmentant  $N$  et  $n$ :

```
N = 20000 # nb de variables Z_i
n = 1000 # taille de chaque échantillon
mu = 0.25
sigma = 0.25
X = matrix(rexp(n*N,rate=4),nrow=N,ncol=n) # On simule directement tous les Xij,
# et on les staocke dans une matrices de taille N*n
Z = replicate(N,0) # on intialise Z avec que des zéros, vecteur de taille N

for (i in 1:N){
  Z[i] = sqrt(n)*(sum(X[i,])/n - mu)/sigma # on calcule Z_i avec la ligne i
  # de la matrice X
}

hist(Z,freq=FALSE,breaks = 150, main="Théorème central limite")
curve(dnorm(x,mean=0,sd=1),add=TRUE, col="red") # on superpose à l'histogramme
```

## Théorème central limite



# la densité d'une gaussienne centrée réduite

Ici, on voit nettement mieux l'illustration du théorème central limite !

4. Pour calculer la constante, on utilise le fait qu'une densité est d'intégrale 1, on doit avoir  $1 = C \int_1^{+\infty} \frac{1}{x^2} dx = C \left[-\frac{1}{x}\right]_1^{+\infty} = C(0 + 1)$ , ce qui donne  $C = 1$ .

5. La fonction de répartition  $F$  est définie pour tout  $x \in \mathbb{R}$  par

$$F(x) = \int_1^x \frac{1}{t^2} dt = 1 - 1/x.$$

Nous connaissons la densité de la variable que l'on veut simuler (ainsi que sa fonction de répartition), mais nous ne savons pas directement simuler une telle variable sous R. Nous sommes ici typiquement dans le cas où nous pouvons (et devons) utiliser la *méthode de l'inversion de la fonction de répartition*. Faisons un rappel du cours:

**(Rappel du cours)** Pour simuler une variable aléatoire réelle  $X$  de fonction de répartition  $F$ , si  $F$  est inversible, alors:

- Nous pouvons d'abord simuler une variable aléatoire  $U$  uniforme sur  $[0, 1]$  (facile sous R),
- Et poser  $X = F^{-1}(U)$ . En effet  $F^{-1}(U)$  aura la loi de  $X$ .

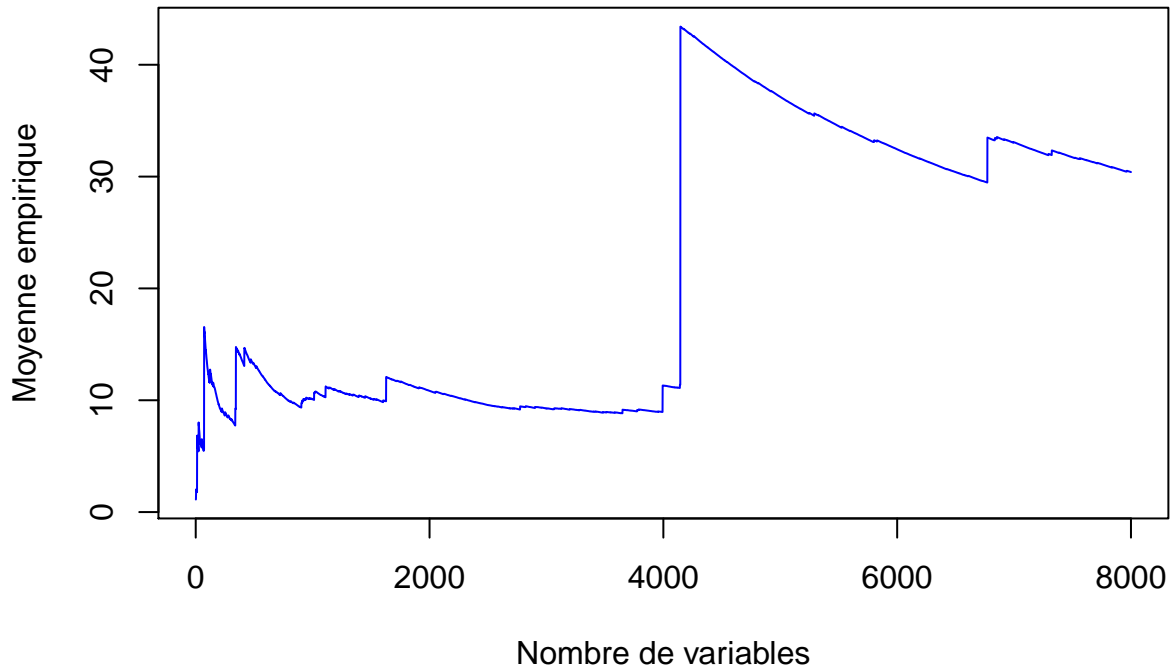
On calcule  $F^{-1}$  rapidement (en résolvant l'équation  $y = 1 - 1/x$ , on trouve  $x = 1/(1-y)$ , donc  $F^{-1}(u) = \frac{1}{1-u}$ ). Cela étant fait, on peut simuler un échantillon de  $X$  en deux lignes

```
n = 8000
U = runif(n) #on simule n variables uniformes classiques
X = 1/(1-U) # on simule X grâce à U en appliquant F^{-1}
```

Ensuite, le code est le même que pour la question 2.

```
Y = cumsum(X)/1:n
plot(1:n,Y,type="l",col="blue",main="Loi des grands nombres ?",
     xlab="Nombre de variables",ylab="Moyenne empirique",)
```

## Loi des grands nombres ?



On remarque que la suite ne semble pas vraiment converger. La loi des grands nombres ne s'applique pas, pourquoi ? Car la variable  $X$  n'est pas d'espérance finie. En effet, l'intégrale  $\int_1^{+\infty} xf(x)dx = \int_1^{+\infty} \frac{1}{x} dx$  diverge !

**Cette question montre que l'hypothèse d'espérance finie dans la LGN est cruciale.**

6. Pour la constante, on utilise le même raisonnement que pour la question 4 : on doit avoir  $1 = D \int_1^{+\infty} \frac{1}{x^3} dx = D \left[ -\frac{1}{2x^2} \right]_1^{+\infty} = D(0 + 1/2)$ , ce qui donne  $D/2 = 1$ , donc  $D = 2$ .

7. On a pour tout réel  $x$ ,

$$G(x) = \int_1^x \frac{2}{t^3} dt = 1 - 1/x^2.$$

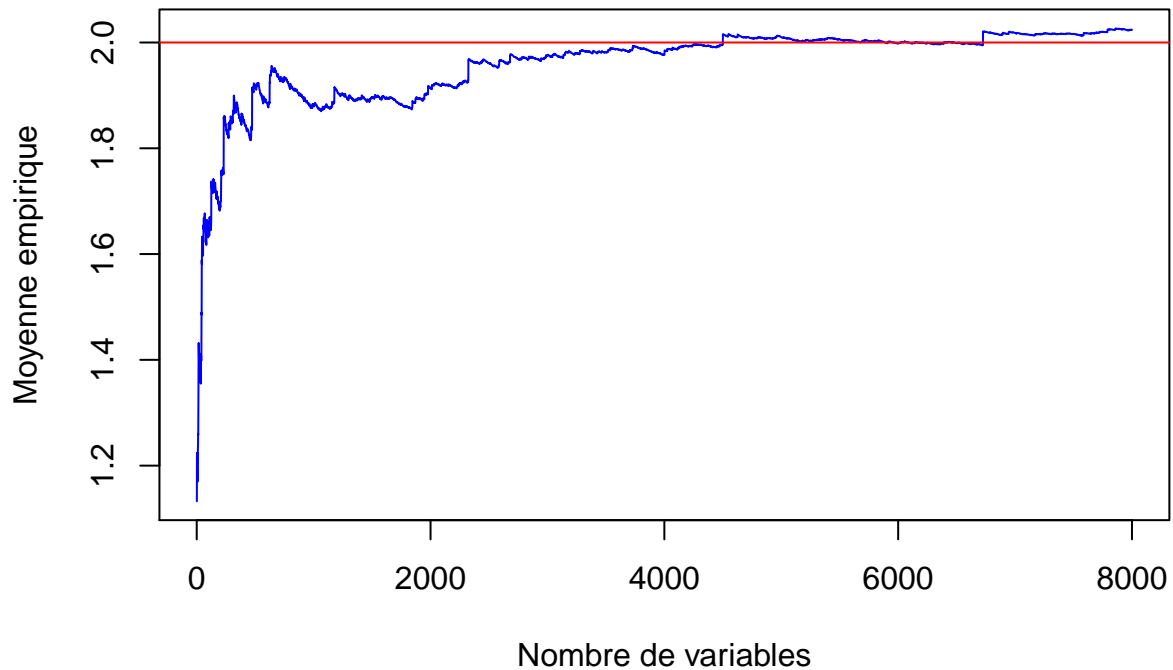
On peut de même "inverser" cette fonction en se restreignant aux valeurs positives de  $x$  : on trouve  $G^{-1}(u) = \frac{1}{\sqrt{1-u}}$ . On utilise à nouveau la méthode de l'inversion de la fonction de répartition pour simuler  $X$ :

```
n = 8000
U = runif(n)
X = 1/sqrt(1-U) # on simule X en appliquant G^{-1}
```

Ici la moyenne de  $X$  se calcule bien et vaut  $\int_1^{+\infty} x \frac{2}{x^3} dx = 2$ . Le code est le même qu'à la question 2. :

```
Y = cumsum(X)/1:n
plot(1:n,Y,type="l",col="blue",main="Loi des grands nombres",
     xlab="Nombre de variables",ylab="Moyenne empirique")
abline(a=2,b=0,col="red")
```

## Loi des grands nombres



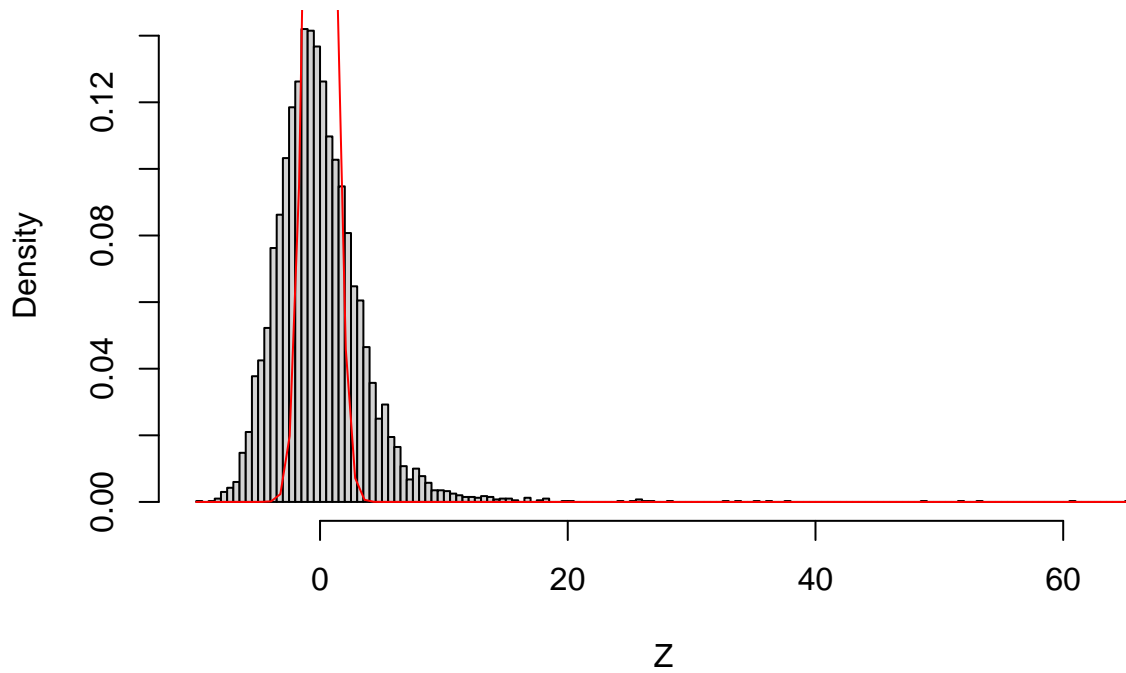
La loi est d'espérance finie, la LGN s'applique donc.

8. Même code qu'en question 3.

```
N = 8000
n = 8000
mu = 2
Z = replicate(n,0) # Z est initialisé avec que des zéros, de taille N
for (i in 1:N){
  X = 1/sqrt(1-runif(n)) # on simule X avec l'inversion
  # de la fonction de répartition
  Z[i] = sqrt(n)*(sum(X)/n - 2) # on remplit Z au rang i
}
hist(Z,freq=F,breaks = 150, main="Théorème central limite ?")
curve(dnorm(x,mean=0,sd=1),add=T, col="red") # on superpose à l'histogramme
```



## Théorème central limite ?



*# la densité d'une gaussienne centrée réduite*

L'histogramme n'est pas proche d'une densité gaussienne : en effet il ne semble pas être centré (trop de valeurs à droite) : cela est le signe que le TCL ne s'applique pas. En fait, ce n'est pas surprenant, car dans notre cas la variable  $X$  a un moment d'ordre 2 infini.

**Cette question montre que l'hypothèse de variance finie dans le TCL est cruciale.**

9. (non corrigé)