TPSTAT2_CHEN_GUANGYUE_GROUPE_3

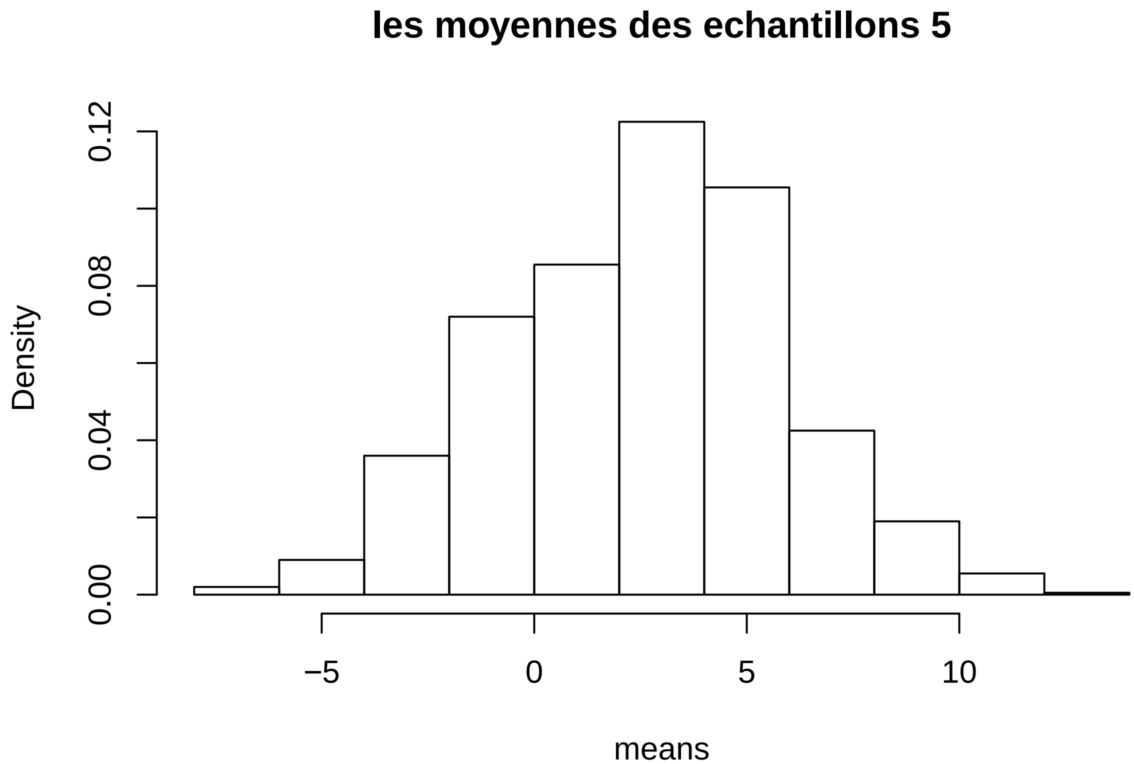
1 Echantillon, Théorème Central Limite, Estimation Monte Carlo

E1.1

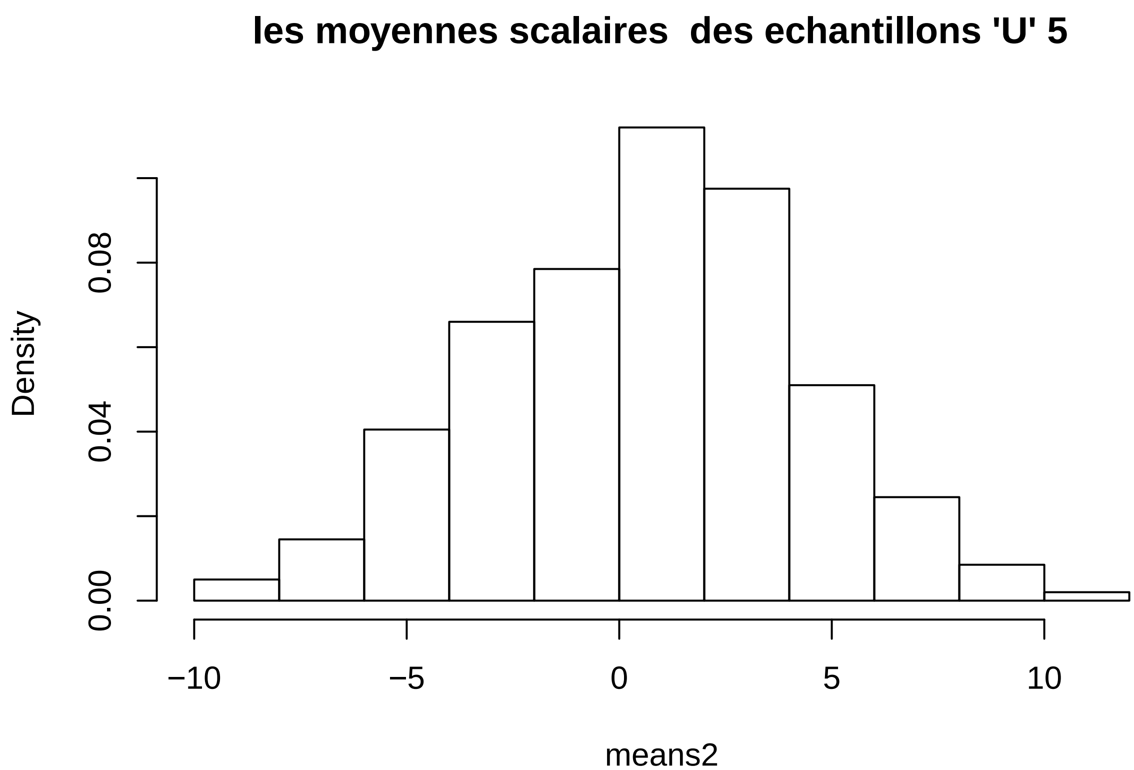
pour simuler les echantillons avec les taille defferentes

```
#mean et var
mymean<-function(echant) {
  s <- sum(echant)
  n <- length(echant)-1
  return (s / n)
}

myvar<-function(echant) {
  s <- 0
  n <- length(echant)-1
  m <- mymean(echant)
  for (x in echant) {
    s <- s + (x-m)^2
  }
  return (s / n)
}





#cette fonction c'est pour representer les histgramme et calcule les moyennes et les variances , et fai
show_echan <-function (n){
  echans<-matrix(rnorm(n*1000,2,6),nrow = n)
  means=vector(length=1000)
  vars=vector(length=1000)
  for (i in 1:1000) {
     means[i]=mymean(echans[,i])
     vars[i]=myvar(echans[,i])
  }
  str<-paste("les moyennes des echantillons",n)
  hist(means,main = str,freq = FALSE)

  cat("\n",n,"la moyenne empirique ",mean(means))

  #apres faire la renormalisation
  myfun=function(x)((x-2)/(2/sqrt(n)))
  means2=sapply(means,myfun)
  #means2=scale(means, center=T,scale=T)

  str<-paste("les moyennes scalaires  des echantillons 'U'",n)
  hist(means2,main = str,freq = FALSE)
```

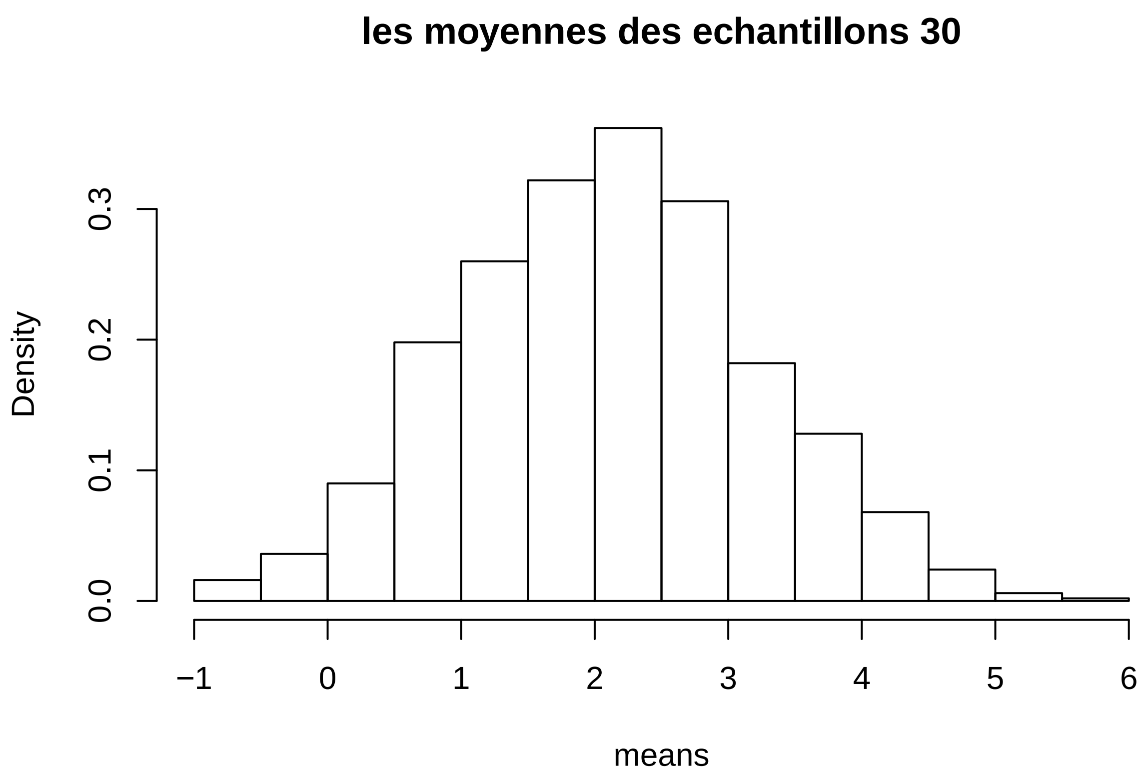
```
}
show_echan(5)
```




```
##
##  5 la moyenne empirique  2.555711
```

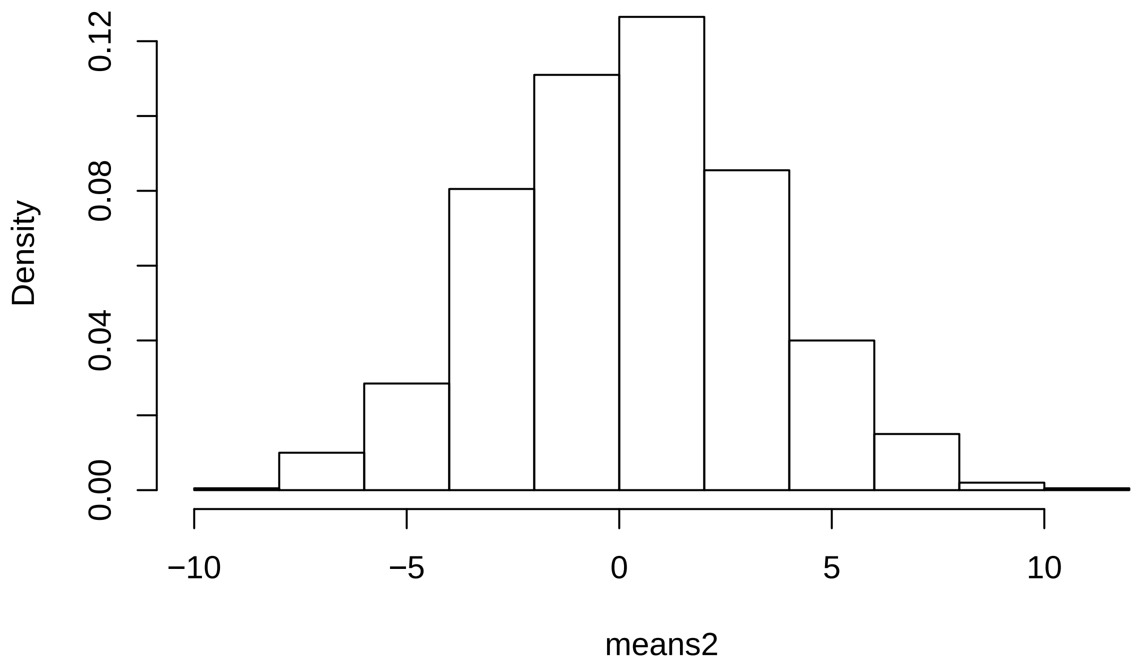



```r
show_echan(30)
```



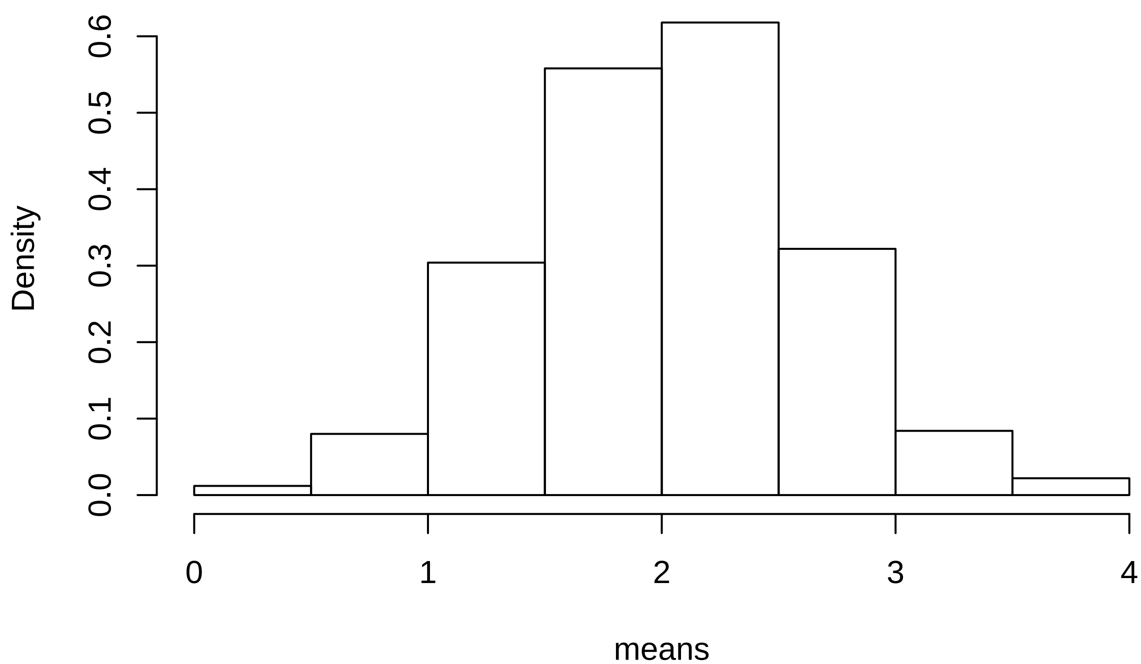

```
##
##  30 la moyenne empirique  2.103172
```

les moyennes scalaires des echantillons 'U' 30

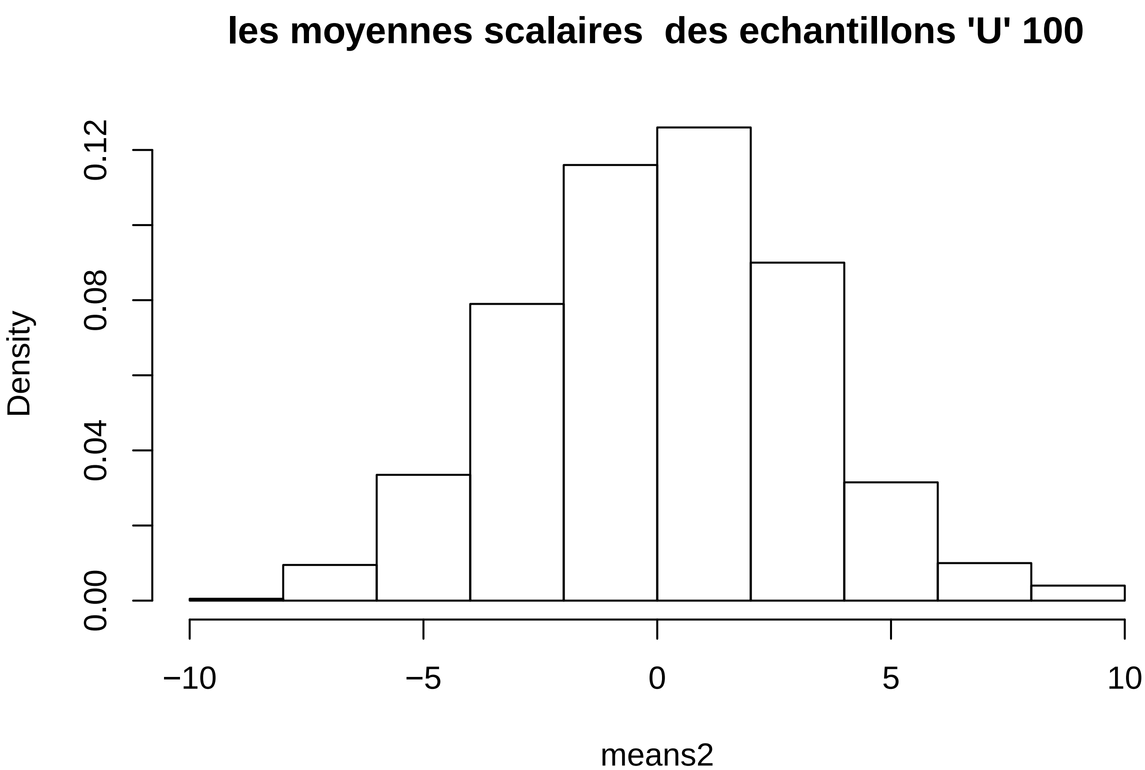



```r
show_echan(100)
```

les moyennes des echantillons 100




```
##
##  100 la moyenne empirique  2.025317
```



Quel est la loi de la moyenne empirique?

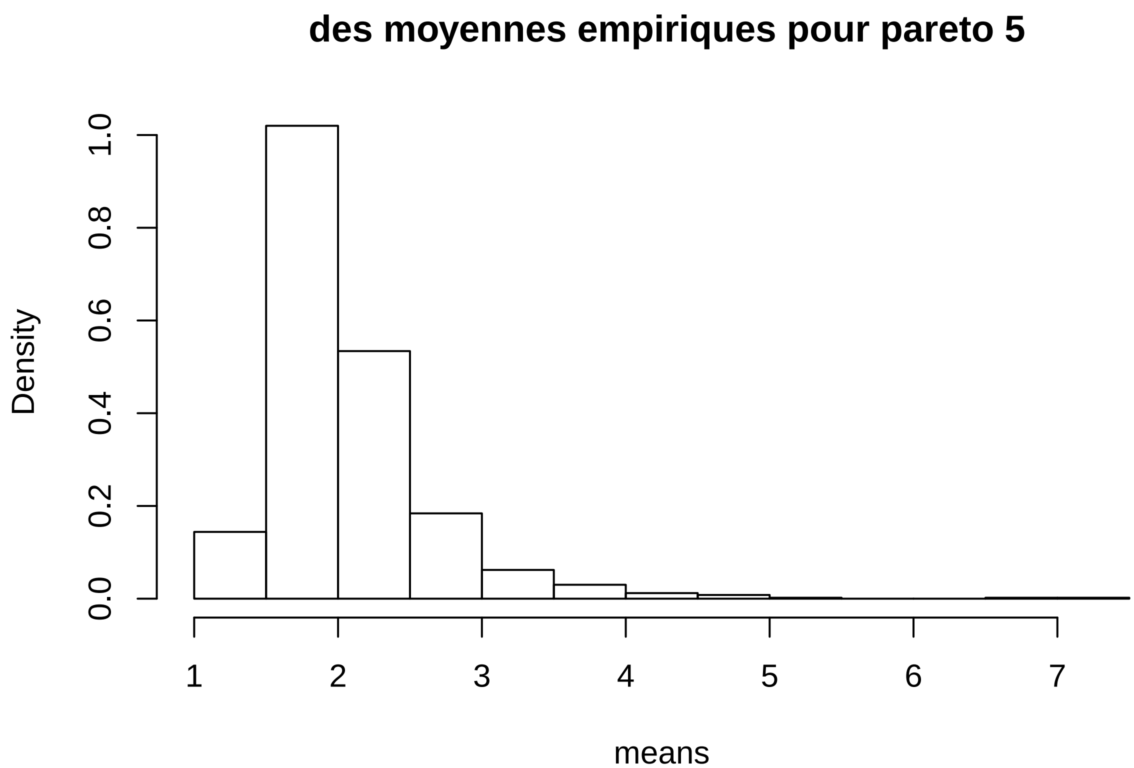
La moyenne empirique d'un échantillon est la somme de ses éléments divisée par leur nombre.La loi théorique des moyennes est la loi gaussienne de paramètres N(1, 2).

Quel est l'influence de la taille de l'echantillons? On remarque que quand N est puls grand , les moyennes sont plus proche de 2 ,et la loi moyenne empirique est plus renormalisé semble suivre une loi N(0, 1).

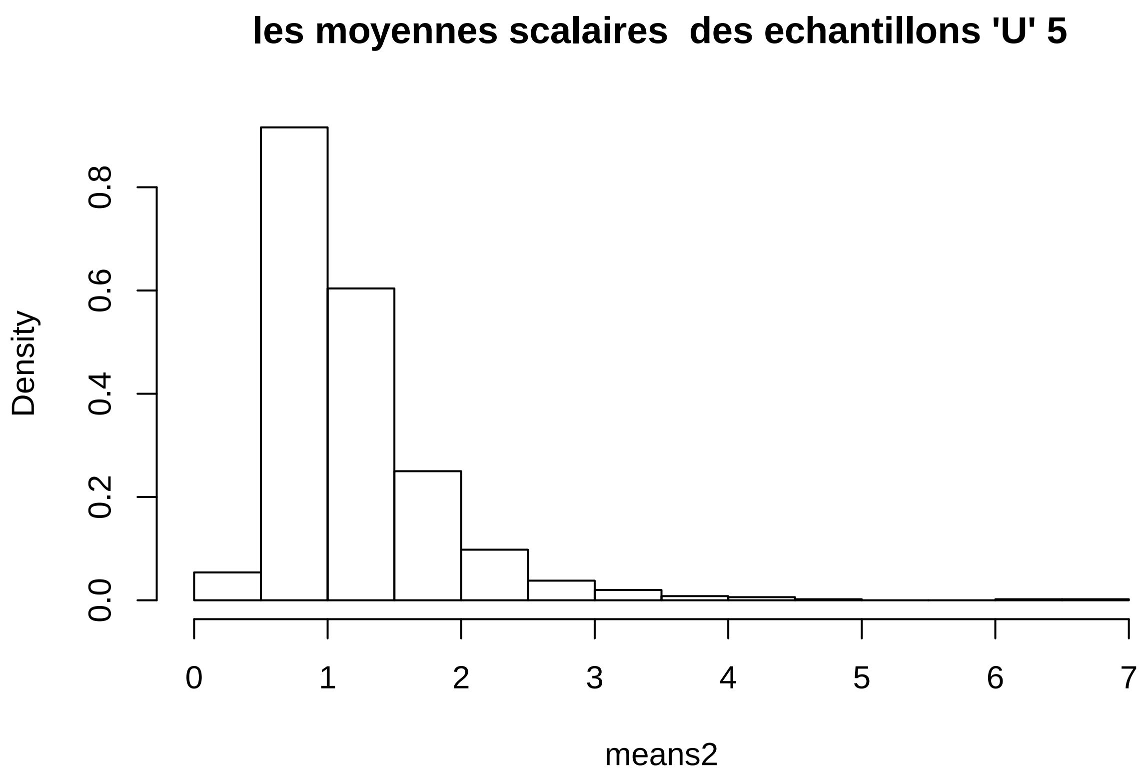
E1.2

```
show_echan_de_pareto <-function (n){
  a <- 1.0
  alpha <- 2.5
  echans<-matrix(rpareto(n*1000, a, alpha),nrow = n)
  means=vector(length=1000)
  vars=vector(length=1000)
  for (i in 1:1000) {
     means[i]=mymean(echans[,i])
     vars[i]=myvar(echans[,i])
  }
  str<-paste(" des moyennes empiriques pour pareto",n)
  hist(means,main = str,freq = FALSE)
cat("\nla moyenne empirique ",n,mean(means))

  #apres faire la renormalisation
  myfun=function(x)((x-1)/(2/sqrt(n)))
  means2=sapply(means,myfun)
```

```
  str<-paste("les moyennes scalaires  des echantillons 'U'",n)
  hist(means2,main = str,freq = FALSE)




}

show_echan_de_pareto(5)
```

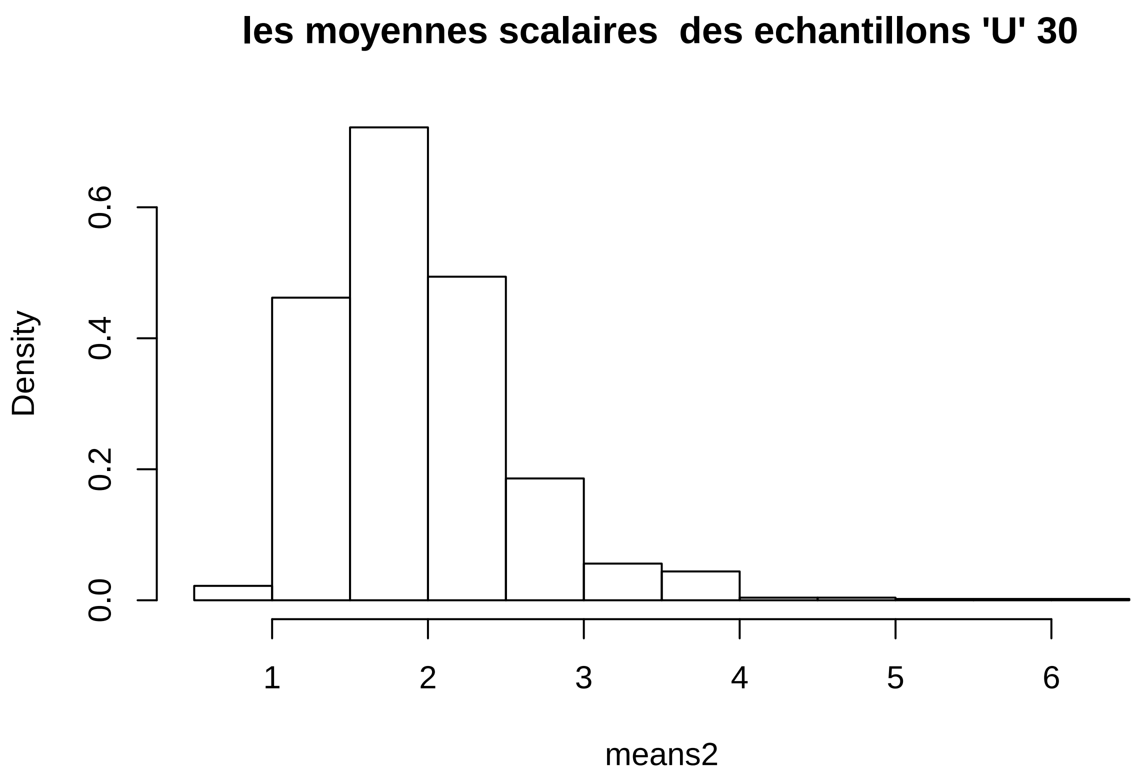



```
##
## la moyenne empirique  5 2.055211
```

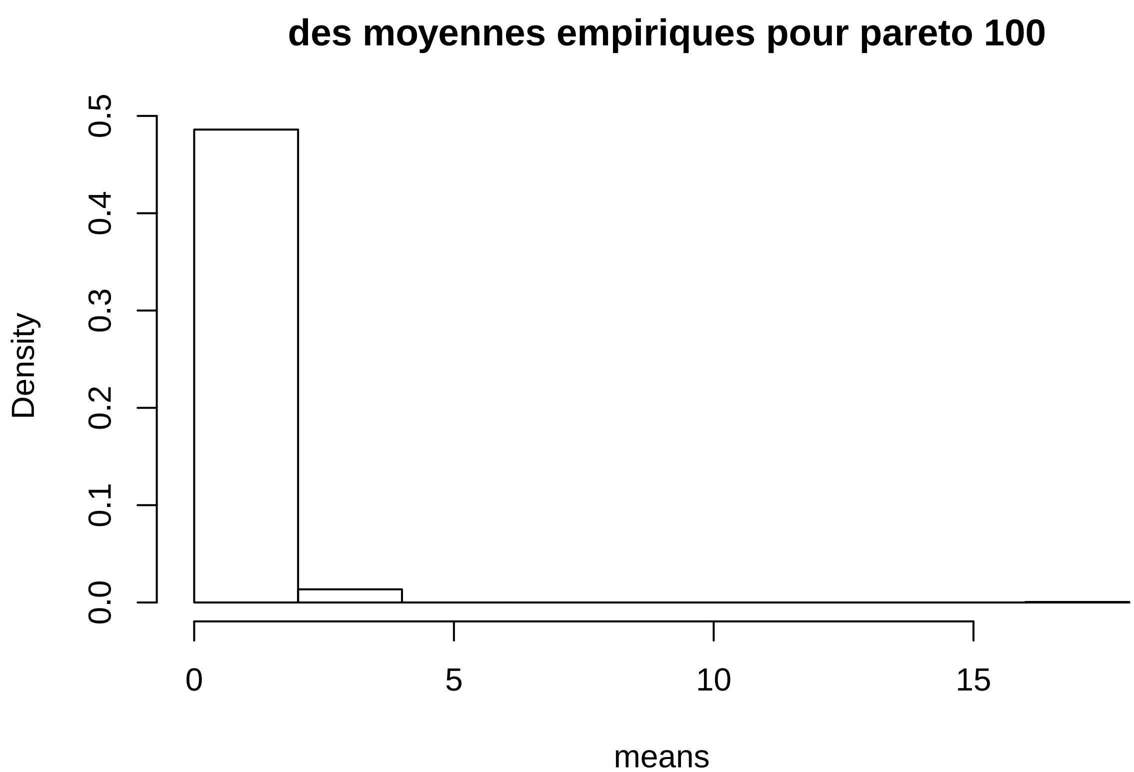



```
show_echan_de_pareto(30)
```

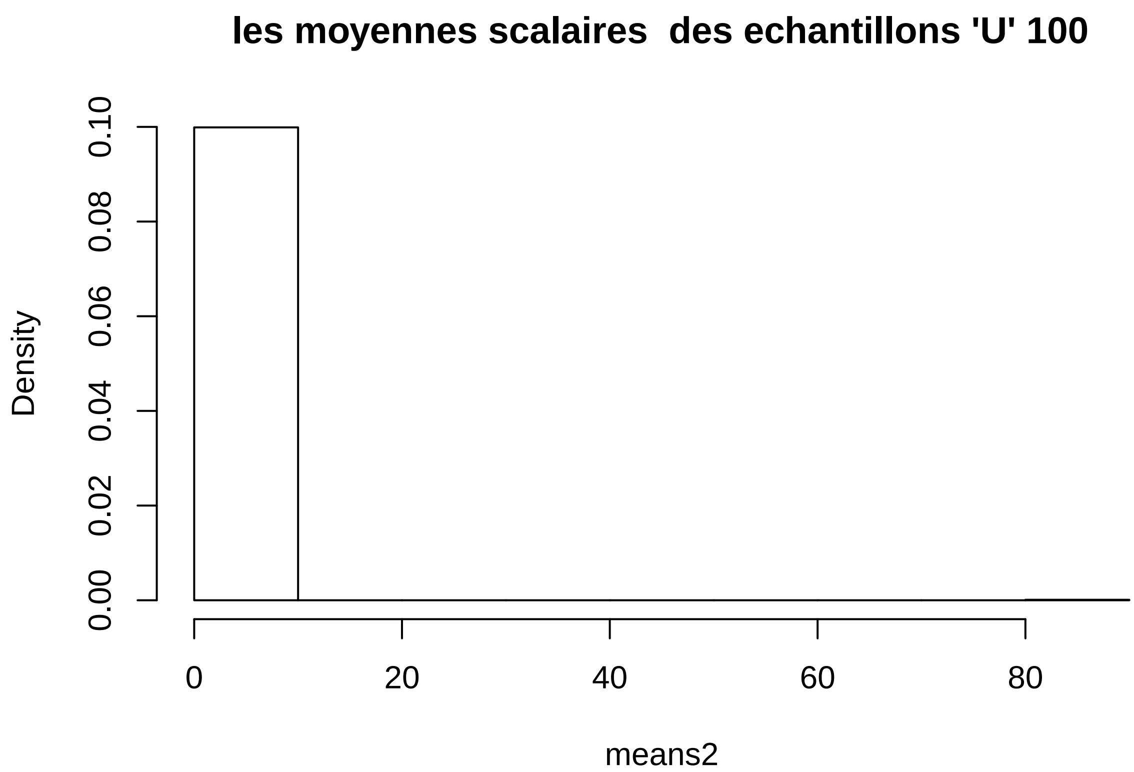

```
##
## la moyenne empirique  30 1.714003
```




```
show_echan_de_pareto(100)
```




```
##
## la moyenne empirique  100 1.69691
```



Quel est la loi theorique de la moyenne empirique?

`E = (alpha*a)/(alpha - 1)`

Var = ((a**2)*alpha)/(((alpha - 1)**2)*(alpha - 2))

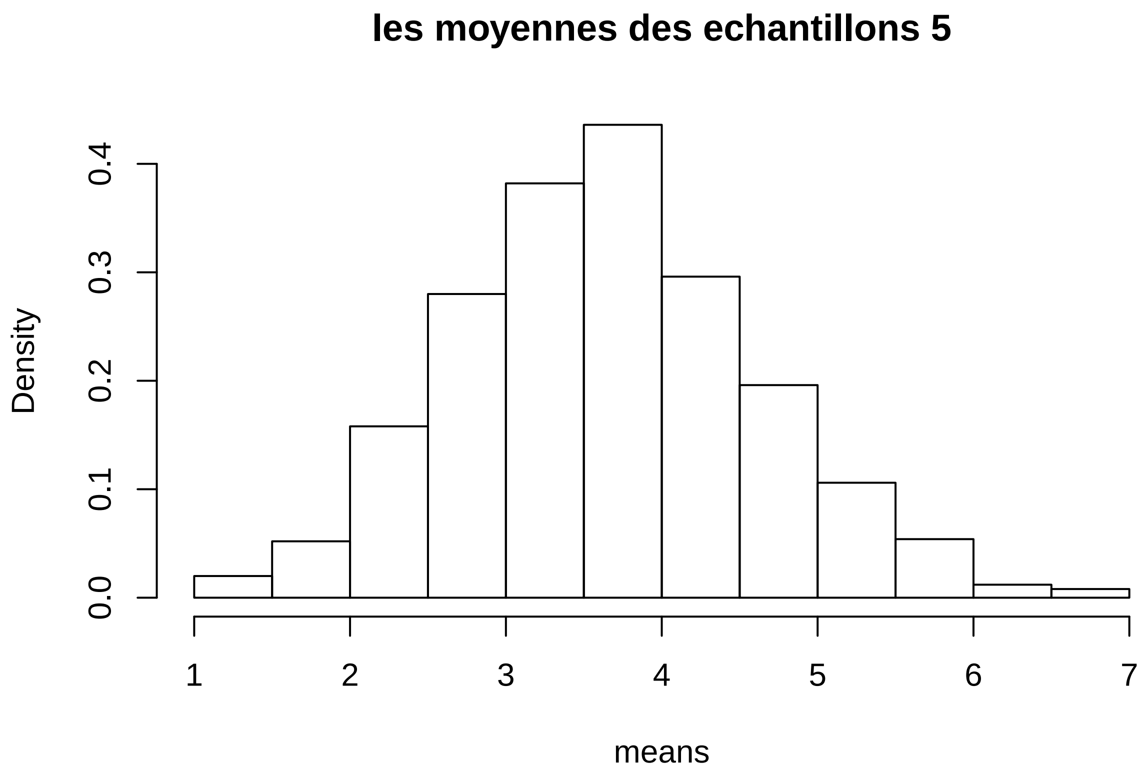
La loi théorique des moyennes est la loi gaussienne de paramètres N(Esperance, Variance) avec Esp l'espérance de Pa et Var la variance.

Quel est l'influence de la taille de l'echantillons? On remarque que quand N est puls grand , la loi moyenne empirique est plus renormalisé semble suivre une loi $N(0, 1)$.

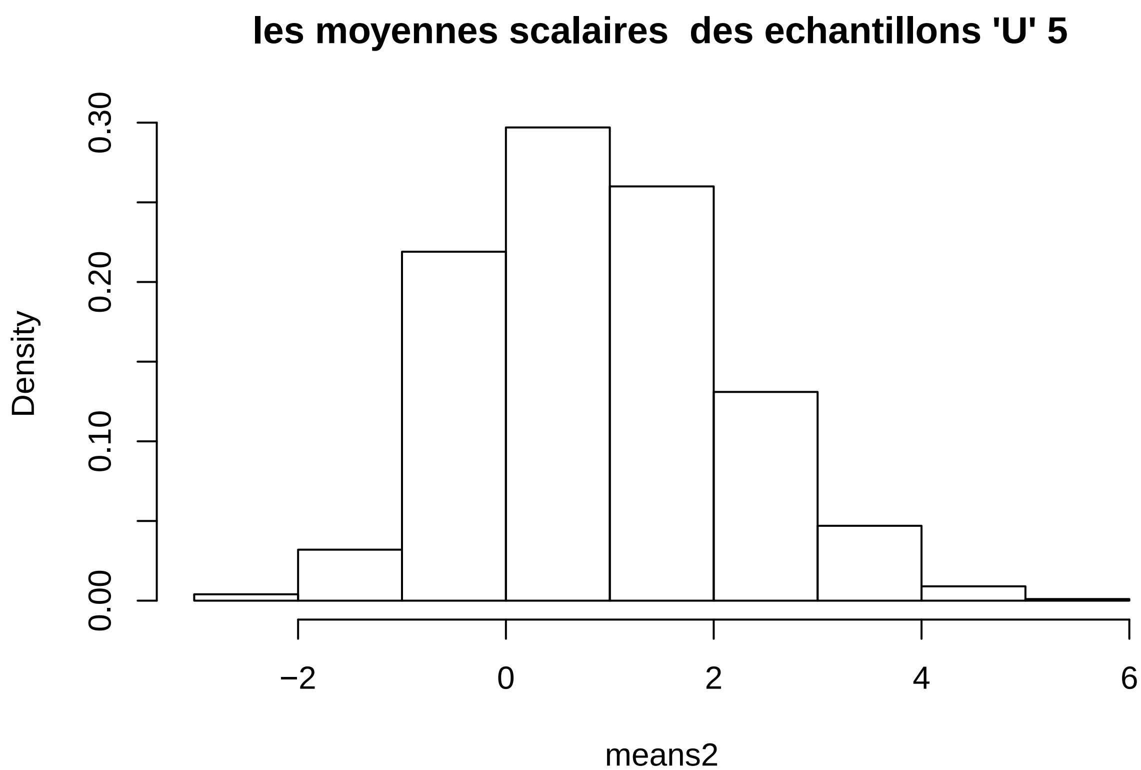
E1.3

```
show_echan_de_Poisson <-function (n){
  echans<-matrix(rpois(n*1000,3),nrow = n)
  means=vector(length=1000)
  vars=vector(length=1000)
  for (i in 1:1000) {
     means[i]=mymean(echans[,i])
     vars[i]=myvar(echans[,i])
  }
  str<-paste("les moyennes des echantillons",n)
  hist(means,main = str,freq = FALSE)

   #apres faire la renormalisation
  myfun=function(x)((x-3)/(sqrt(3)/sqrt(n)))
  means2=sapply(means,myfun)
  #means2=scale(means, center=T,scale=T)
 cat("\nla moyenne empirique ",n,mean(means))
```

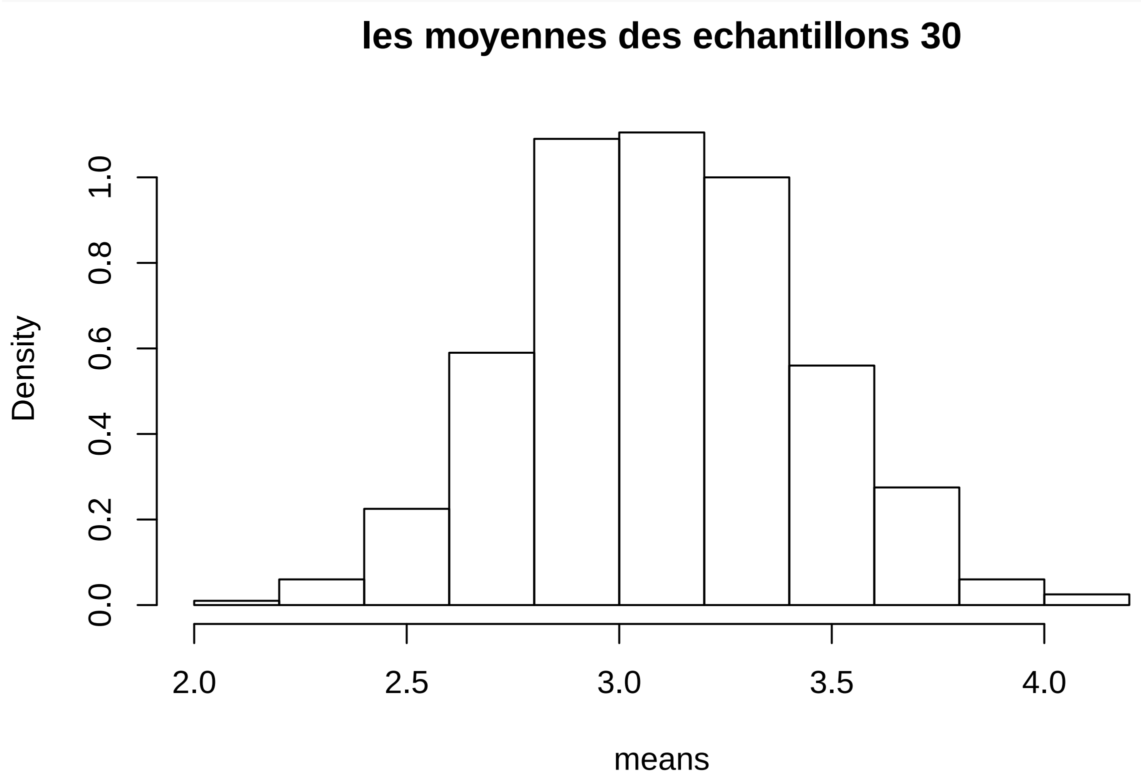
```
  str<-paste("les moyennes scalaires  des echantillons 'U'",n)
  hist(means2,main = str,freq = FALSE)




}
show_echan_de_Poisson(5)
```



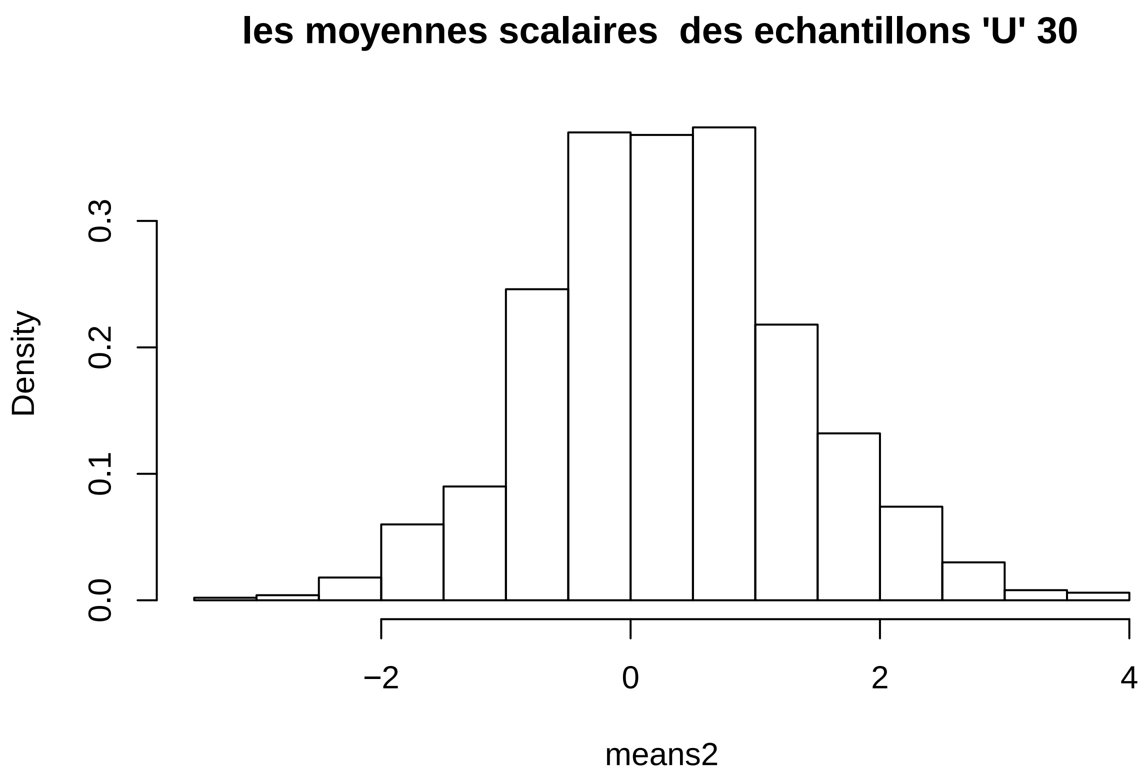

```
##
## la moyenne empirique  5 3.7815
```



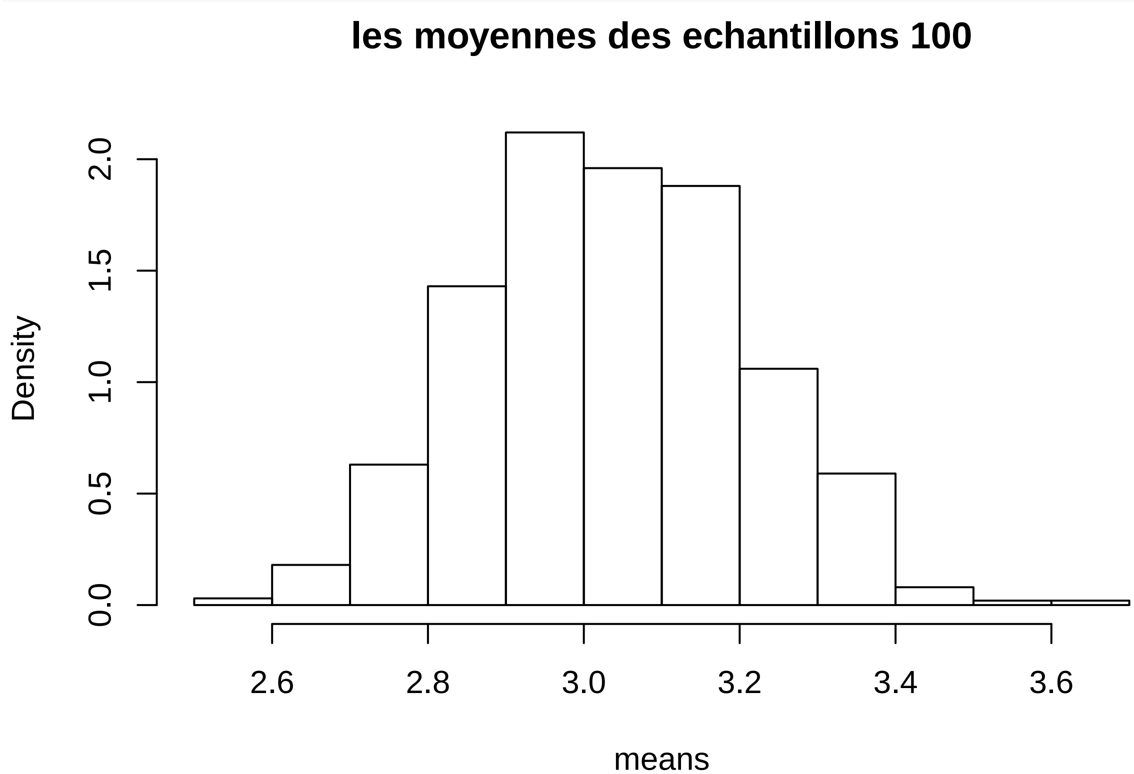

```r
show_echan_de_Poisson(30)
```



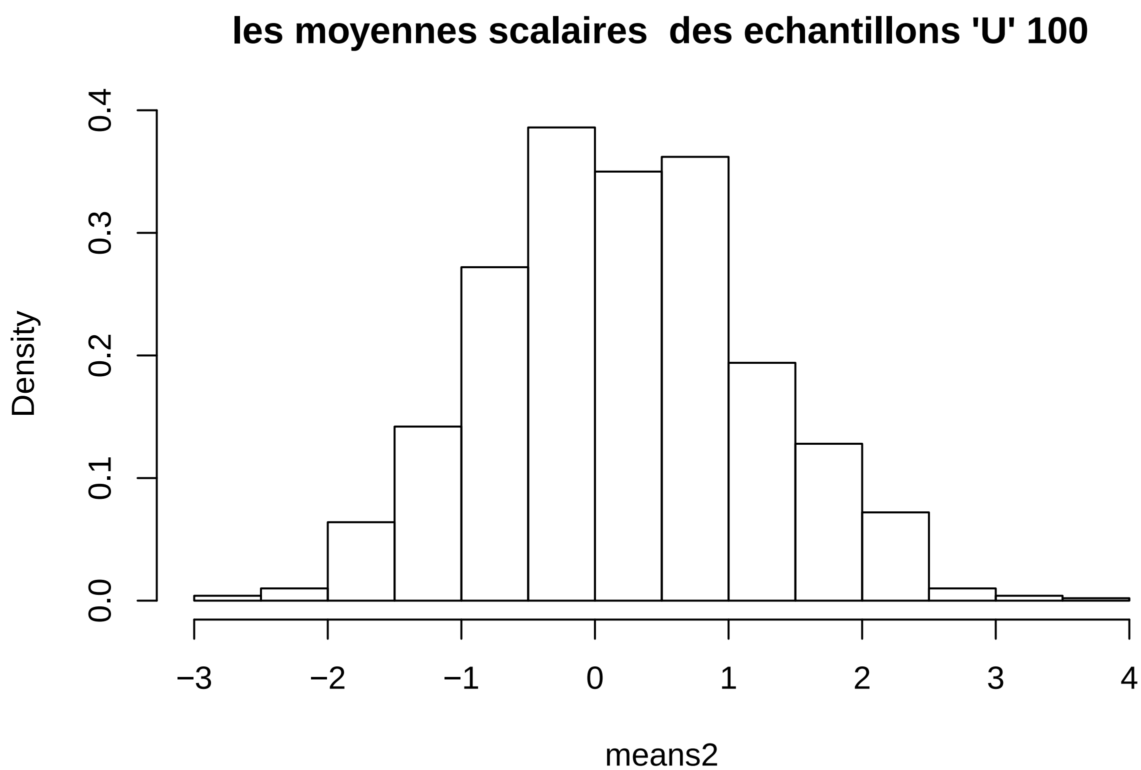

```
##
## la moyenne empirique  30 3.103483
```




```r
show_echan_de_Poisson(100)
```




```
##
## la moyenne empirique  100 3.036778
```



Quel est la loi de la moyenne empirique? La moyenne théorique d'une loi de Poisson ayant pour paramètre lambda est de lamba, et sa variance est aussi lambda.

Quel est l'influence de la taille de l'echantillons? On remarque que quand N est puls grand , la loi moyenne empirique est plus renormalisé semble suivre une loi N(0, 1).

E1.4

comment N influence la qualite de cette approximation?

La moyenne empirique est une bonne estimation de l'espérance quand N tends vers l'infini.

2 Moyenne et Dispersion

E2.1Bienaymé Chebychef dans les cas Gaussien et Poisson

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \alpha) \leq \frac{Var[X]}{\alpha^2}$$

Pour une loi Gaussienne $N(\mu, \sigma^2)$, on a :

$$\mathbb{E}[X] = \mu$$

$$Var[X] = \sigma^2$$

Pour une loi de Poisson :

$$\mathbb{E}[X] = \lambda$$

$$Var[X] = \lambda$$

E2.2(a)

$$\mathbb{P}(|X-\mu| \geq \delta) = \mathbb{E}[\nVdash_{\{|X-\mu|\geq\delta\}}] = \mathbb{E}[Z]$$

, ici

$$Z = \nVdash_{\{|X-\mu|\geq\delta\}}$$

E2.2(b)

```
estimation_norm <-function (n,delta,a,b){
  echans<-rnorm(n*n,a,b)
  myfonc=function(x) {
                              if (abs(x-a) >= delta) {
                                return (1)
                              }
                              return (0)
                            }
  echans<-sapply(echans,myfonc)
  y=sum(echans)/n/n
  y<-paste("Gauss:",y)
  print(y)
  return(echans)
}
P_gauss<-estimation_norm(1000,delta=1,0,1)
```

```
## [1] "Gauss: 0.316851"
```

```
estimation_poisson <-function (n,delta,b){
  echans<-rpois(n*n,b)
  myfonc=function(x) {
                              if (abs(x-b) >= delta) {
                                return (1)
                              }
                              return (0)
                            }
  echans<-sapply(echans,myfonc)
  y=sum(echans)/n/n
    y<-paste("POIsson:",y)
  print(y)
   return(echans)
}
P_poisson<-estimation_poisson(1000,delta=1,1)
```

```
## [1] "POIsson: 0.632224"
```

```
estimation_pareto <-function (n,delta,a,alpha){
  echans<-rpareto(n*n, a, alpha)
  b<-alpha*a/(alpha - 1)
   myfonc=function(x) {
                              if (abs(x-b) >= delta) {
                                return (1)
                              }
                              return (0)
```

```
                            }
  echans<-sapply(echans,myfonc)
  y=sum(echans)/n/n
    y<-paste("pareto:",y)
  print(y)
   return(echans)
}
P_pareto<-estimation_pareto(1000,delta=1,1,2.5)
```

```
## [1] "pareto: 0.086049"
```

Pouver vous determiner la precision de cette estimation?

$$\mathbb{P}[|\bar{Z}_N - \mathbb{E}[Z]| \geq \epsilon] \leq \frac{\mathbb{V}[Z]}{\epsilon\sqrt{N}}$$

#E2.2(c)

```
#pour la precision d'estimation
precision<- function(X, e) {
  return (var(X) / (e * sqrt(1000)));
}
print("la precision de gauss")
```

```
## [1] "la precision de gauss"
```

```
print(precision(P_gauss,0.05))
```

```
## [1] 0.1368992
```

```
print("la precision de poisson")
```

```
## [1] "la precision de poisson"
```

```
print(precision(P_poisson,0.05))
```

```
## [1] 0.1470567
```

```
print("la precision de perato")
```

```
## [1] "la precision de perato"
```

```
print(precision(P_pareto,0.05))
```

```
## [1] 0.04973924
```

E2.2(d)

Dans l'inégalité de Chernoff pour le cas Gaussien, en remplaçant les termes on a :

$$\mathbb{P}(X \geq \delta) \leq exp\left(-\frac{\delta^2}{2\sigma^2}\right)$$

$$\mathbb{P}(X \geq \delta) \leq exp\left(-\frac{\delta^2}{2\sigma^2}\right)$$

```
  BorneChernoff_G <-function(delta,V) exp(- delta**2/(2*(V**2)))
  cat("Estimation de Monte-Carlo du terme =", mean(P_gauss))
```

```
## Estimation de Monte-Carlo du terme = 0.316851
```

```r
cat("\nBorne de Chernoff =", BorneChernoff_G(1,1))
```

```
##
## Borne de Chernoff = 0.6065307
```

```r
cat("\nDifférence =", BorneChernoff_G(1,1)-mean(P_gauss))
```

```
##
## Différence = 0.2896797
```

Dans l'inégalité de Chernoff pour le cas Poisson, en remplaçant les termes on a :

$$\mathbb{P}(X \geq \delta) \leq exp\left(-\lambda\left[\left(1+\frac{\delta}{\lambda}\right)log\left(1+\frac{\delta}{\lambda}\right)-\frac{\delta}{\lambda}\right]\right)$$

```r
BorneChernoff_P <-  function(delta,l) exp(-l*((1 + (delta / l))*log(1 + (delta / l)) - (delta / l)))
  cat("Estimation de Monte-Carlo du terme =", mean(P_poisson))
```

```
## Estimation de Monte-Carlo du terme = 0.632224
```

```r
cat("\nBorne de Chernoff =", BorneChernoff_P(1,1))
```

```
##
## Borne de Chernoff = 0.6795705
```

```r
cat("\nDifférence =", BorneChernoff_P(1,1) - mean(P_poisson))
```

```
##
## Différence = 0.04734646
```

La borne donnée par l'inégalité de Chernoff est bien plus précise que celle donnée par l'inégalité de Bienaymé-Tchebychev.

E3 (a)

```r
chernoff_norm <-function (n,a,b){
  #n fois rnorm(n,a,b)
  echans<-matrix(rnorm(n*n,a,b),nrow = n)
  bornes=vector(length=n)
  vars=vector(length=n)
  count=0
  for (i in 1:n) {
    vars[i]=var(echans[,i])
    bornes[i]=BorneChernoff_G(1,sqrt(vars[i]))
  }
return(mymean(bornes))
}
B_gausee<-chernoff_norm(20,0,1)
cat("\nBorne de Chernoff de gausse N(0,1) 20=",mean(B_gausee))
```

```
##
## Borne de Chernoff de gausse N(0,1) 20= 0.5790724
```

```r
B_gausee<-chernoff_norm(100,0,1)
cat("\nBorne de Chernoff de gausse N(0,1) 100=",mean(B_gausee))
```

```
##
## Borne de Chernoff de gausse N(0,1) 100= 0.6076926
```

```r
B_gausee<-chernoff_norm(1000,0,1)
cat("\nBorne de Chernoff de gausse N(0,1) 1000=",mean(B_gausee))
```

```
##
## Borne de Chernoff de gausse N(0,1) 1000= 0.606907
```

```r
chernoff_poisson <-function (n,b){
  echans<-matrix(rpois(n*n,b),nrow = n)
  bornes=vector(length=n)
  vars=vector(length=n)
  count=0
  for (i in 1:n) {
    vars[i]=var(echans[,i])
    bornes[i]=BorneChernoff_P(1,vars[i])
  }
return(mymean(bornes))
}
B_poisson<-chernoff_poisson(20,1)
cat("\nBorne de Chernoff de poisson 20 =",mean(B_poisson))
```

```
##
## Borne de Chernoff de poisson 20 = 0.6946389
```

```r
B_poisson<-chernoff_poisson(100,1)
cat("\nBorne de Chernoff de poisson 100=",mean(B_poisson))
```

```
##
## Borne de Chernoff de poisson 100= 0.6800147
```

```r
B_poisson<-chernoff_poisson(1000,1)
cat("\nBorne de Chernoff de poisson 1000=",mean(B_poisson))
```

```
##
## Borne de Chernoff de poisson 1000= 0.679914
```

(b)en deduire un estimateur de mu et lameda ###E4 #(a)

```r
show_echan_de_Cauchy <-function (n){

  c = rcauchy(n,location=0,scale=1)
  return ( mymean(c))

}
y1<-show_echan_de_Cauchy(20)
y2<-show_echan_de_Cauchy(100)
y3<-show_echan_de_Cauchy(1000)
y4<-show_echan_de_Cauchy(10000)

cat("\nla moyenne empirique 20",y1)
```

```
##
## la moyenne empirique 20 -0.3098972
```

```r
cat("\nla moyenne empirique 100",y1)
```

```
##
```

```
## la moyenne empirique 100 -0.3098972
```

```r
cat("\nla moyenne empirique 1000",y1)
```

```
##
## la moyenne empirique 1000 -0.3098972
```

```r
cat("\nla moyenne empirique 10000",y1)
```

```
##
## la moyenne empirique 10000 -0.3098972
```

(b)expliquer ce compotement

Une v.a X suivant une loi de Cauchy(θ) n'admet pas d'espérance:

$f_X(x, \theta) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}$, et quand $x \to +\infty$, $x f_X(x, \theta) \sim \frac{1}{x}$, donc:

$\mathbb{E}[X] = \int_{-\infty}^{+\infty} |x f_X(x, \theta)| dx$ diverge.

Il n'y a pas d'espérance ,donc le théorème central limite ne s'applique pas, donc la moyenne empirique ne converge pas.

Ceci s'explique par le fait que la probabilité d'obtenir une valeur éloigné de la médiane est trop elévé pour que la moyenne converge.

(c)Quel est la mediance d'une loi de Cauchy

La médiane d'une loi de Cauchy de paramètres

$$(a, x_0)$$

est

$$x_0$$

```r
mediance<-function(n){

  echans<-matrix(rcauchy(1000*n,0,1),nrow = n)
  mediances<-vector(length=1000)
   for (i in 1:1000) {
     mediances[i]=median(echans[,i], na.rm = FALSE)
  }
 return(mean(mediances))
}
mediance(20)
```

```
## [1] -0.005702269
```

```r
mediance(100)
```

```
## [1] 0.001508898
```

```r
mediance(1000)
```

```
## [1] 0.0003411733
```