# Tp Statistiques 2

Lapu Matthias | Amaël Kreis

## Simulation et convergence

### I. Variation sous-jacente et échantillonnage répété

1. Si X  E(0.5), quelle est la probabilité qu'on observe une valeur supérieure à 3?

On a

$$f(0.5) = 0.5e^{-0.5x}$$

Donc pour

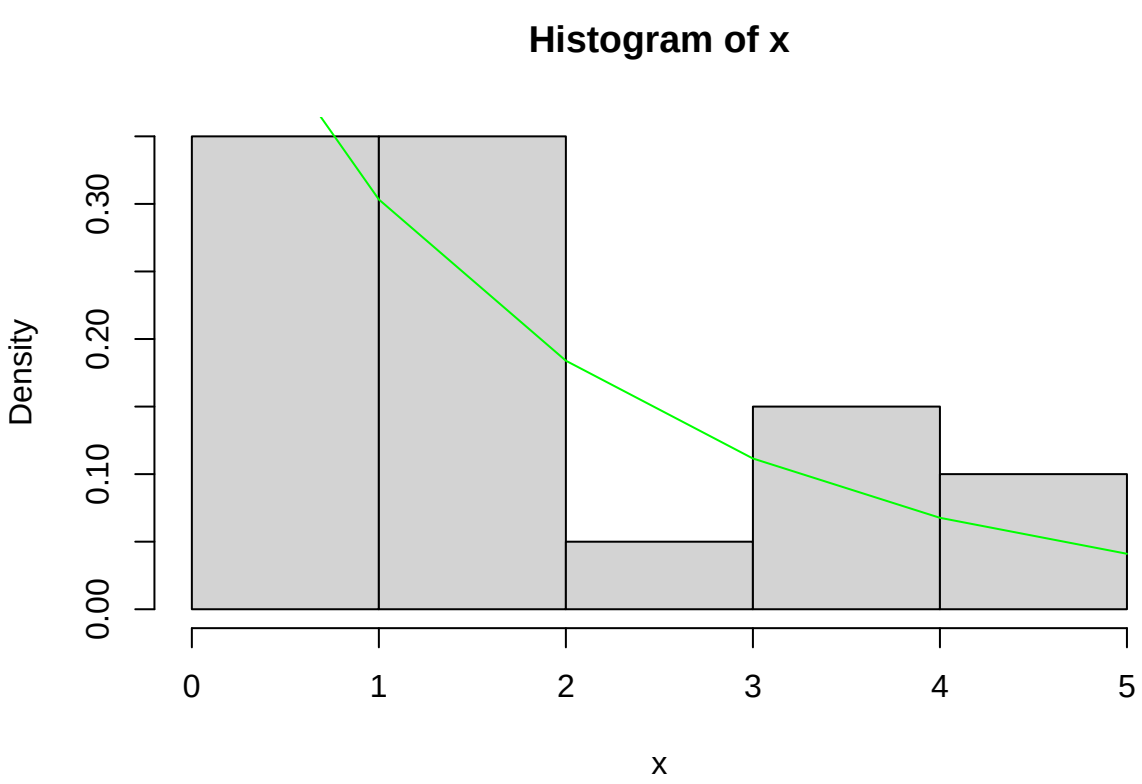
$$X \sim \xi(0.5)$$

On trouve

$$\int_3^{+\infty} \frac{1}{2} e^{\frac{-x}{2}}\, dx = \left[-e^{-\frac{x}{2}}\right]_3^{+\infty} = e^{-\frac{3}{2}} \approx 0.223$$

2.Simulez un échantillon de taille n = 20 d'un loi de E(0,5), créez un histogramme de votre échantillon et commentez la forme de votre histogramme. Superposer la vrai densité. Quelle est la probabilité empirique qu'on observe une valeur supérieure à 3 ?

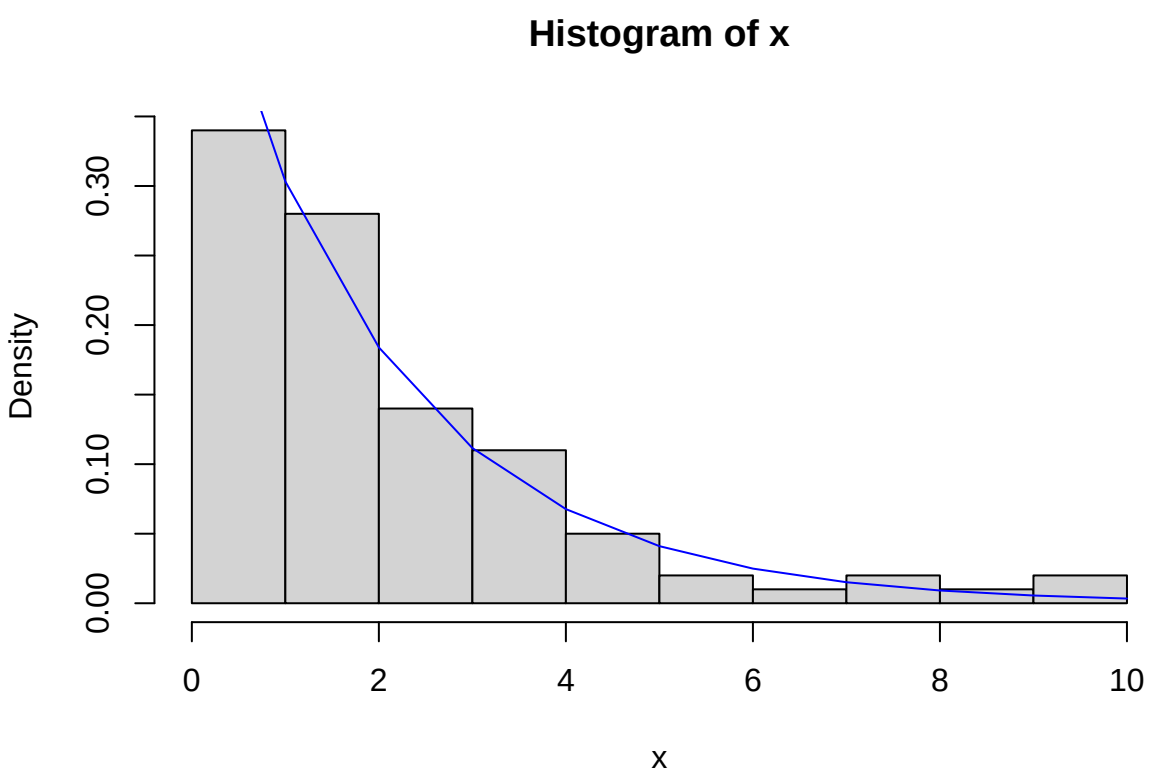
```
x<-rexp(20,0.5)
hist(x, freq=FALSE)
maxvalue <- ceiling(max(x))
lines(0:maxvalue,dexp(0:maxvalue, 0.5), col="green",)
```




Commentaire histogramme à insérer

3.Répétez cette opération 5 ou 6 fois et commentez les différences entre les histogrammes que vous obtenez à chaque fois. Utilisez la même limite sur les axes pour faciliter la comparaison. Notez également comment la probabilité empirique qu'on observe une valeur supérieure à 3 change.

4. Augmentez la taille de votre échantillon à 100 et répétez votre expérience. Que remarquez-vous?




## II. Variabilité aléatoire du maximum de l'échantillon

1. Simuler un échantillon de taille n = 10 d'une loi U(−1, 1) et enregistrez le maximum de l'échantillon.

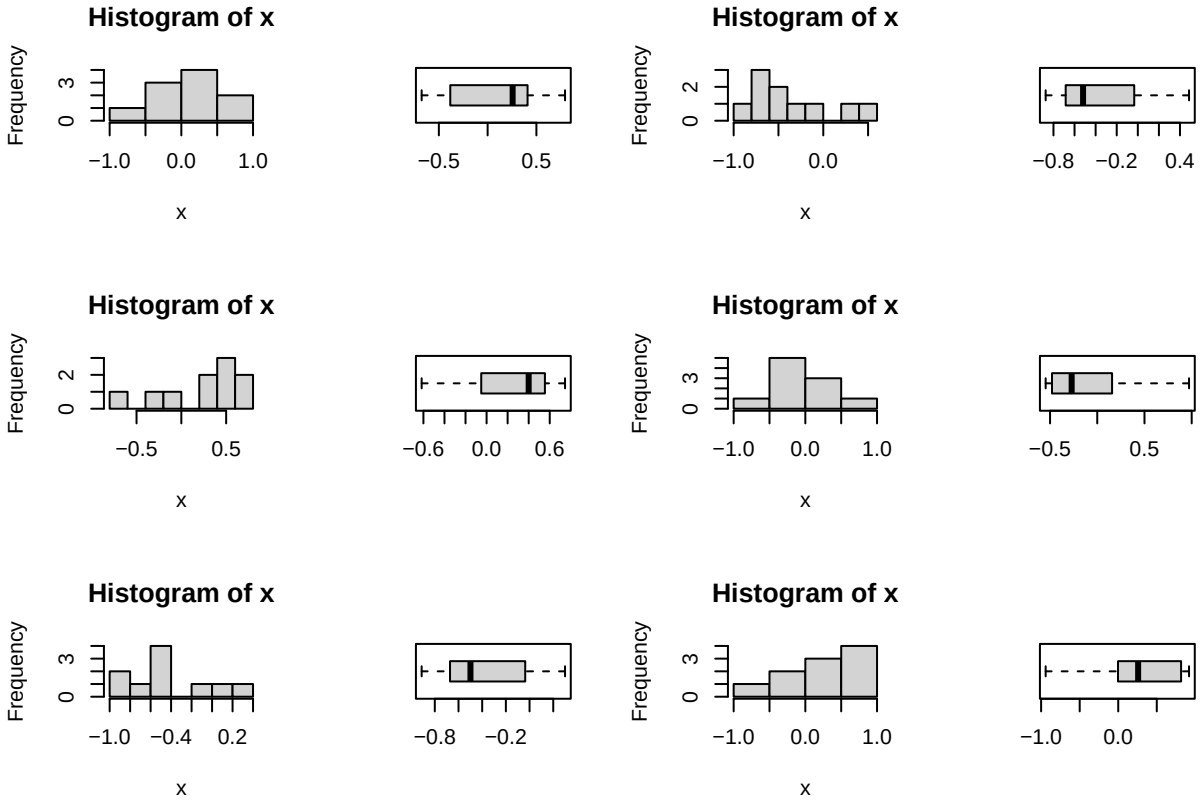
```
x <- runif(10, -1, 1)
max <- max(x)
```

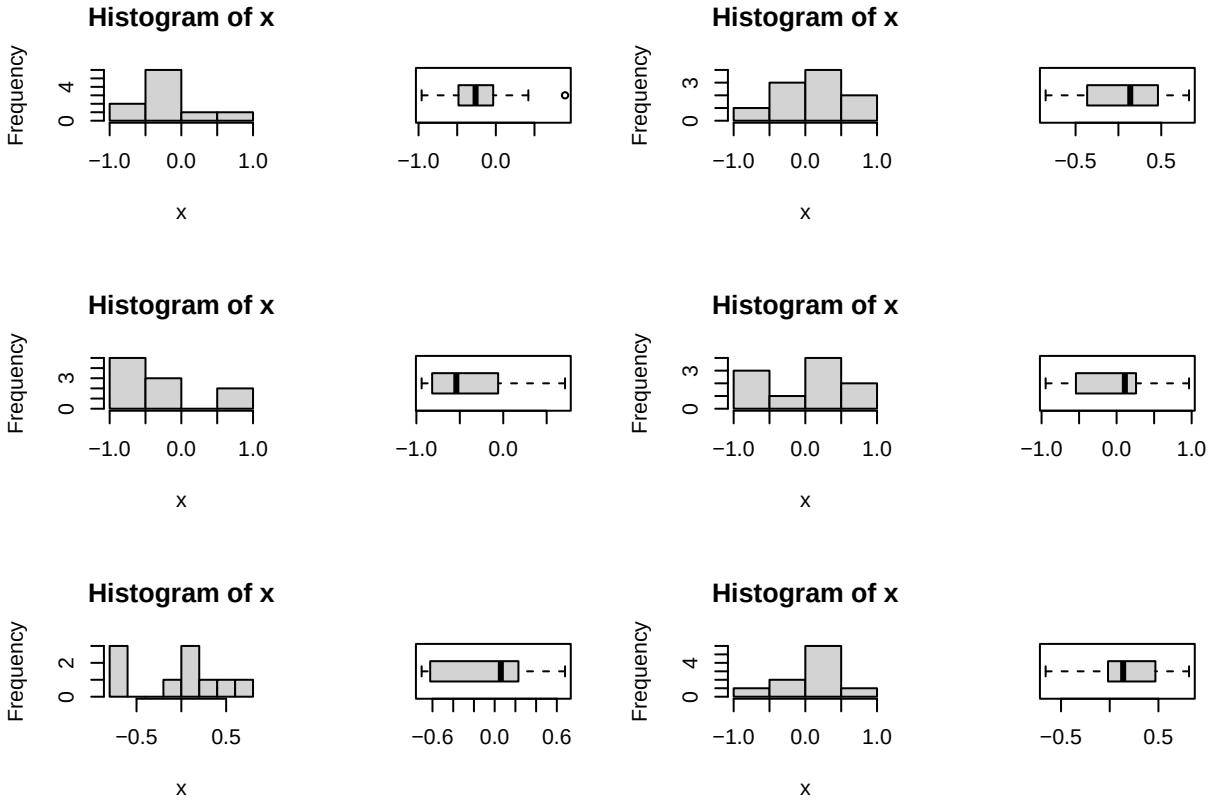
2. Répétez les deux étapes ci-dessus dix fois, en écrivant le maximum de l'échantillon à chaque fois. Commentez la variabilité des valeurs que vous obtenez pour les maxima de votre échantillon.

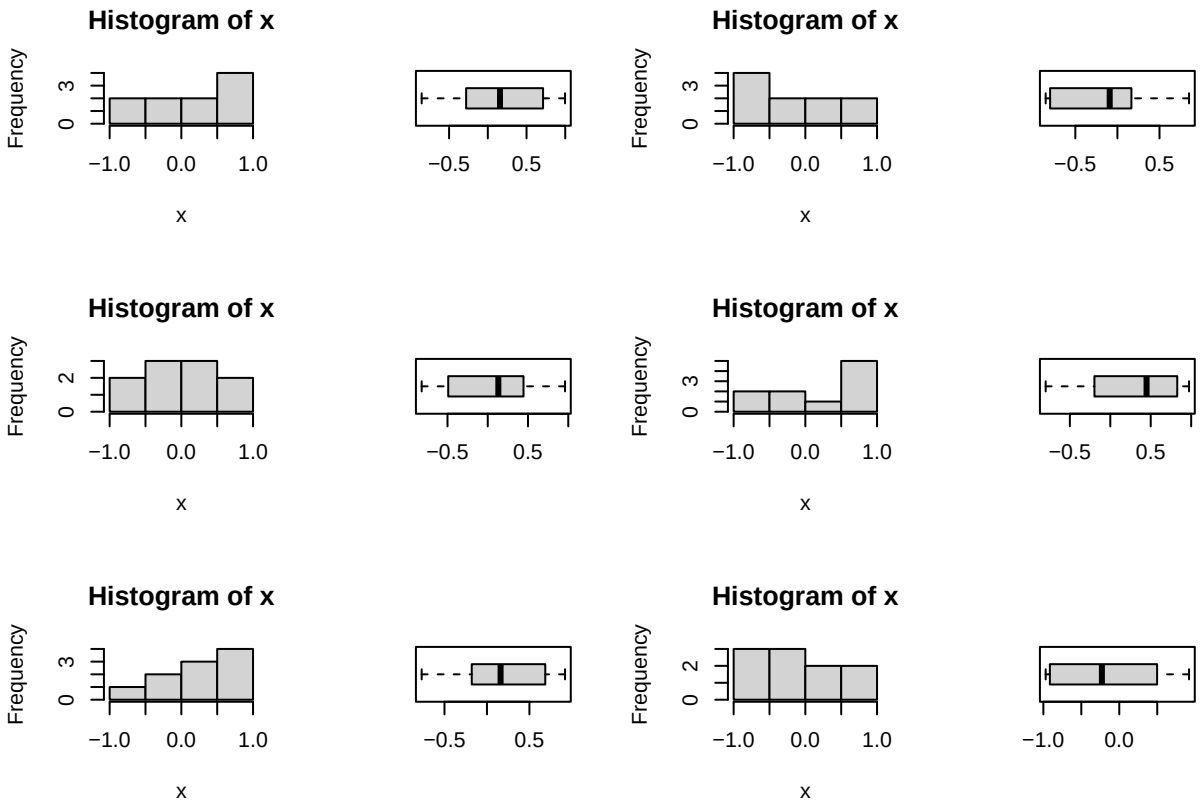
```
for (i in 1:10) {
  x <- runif(10, -1, 1)
}
```

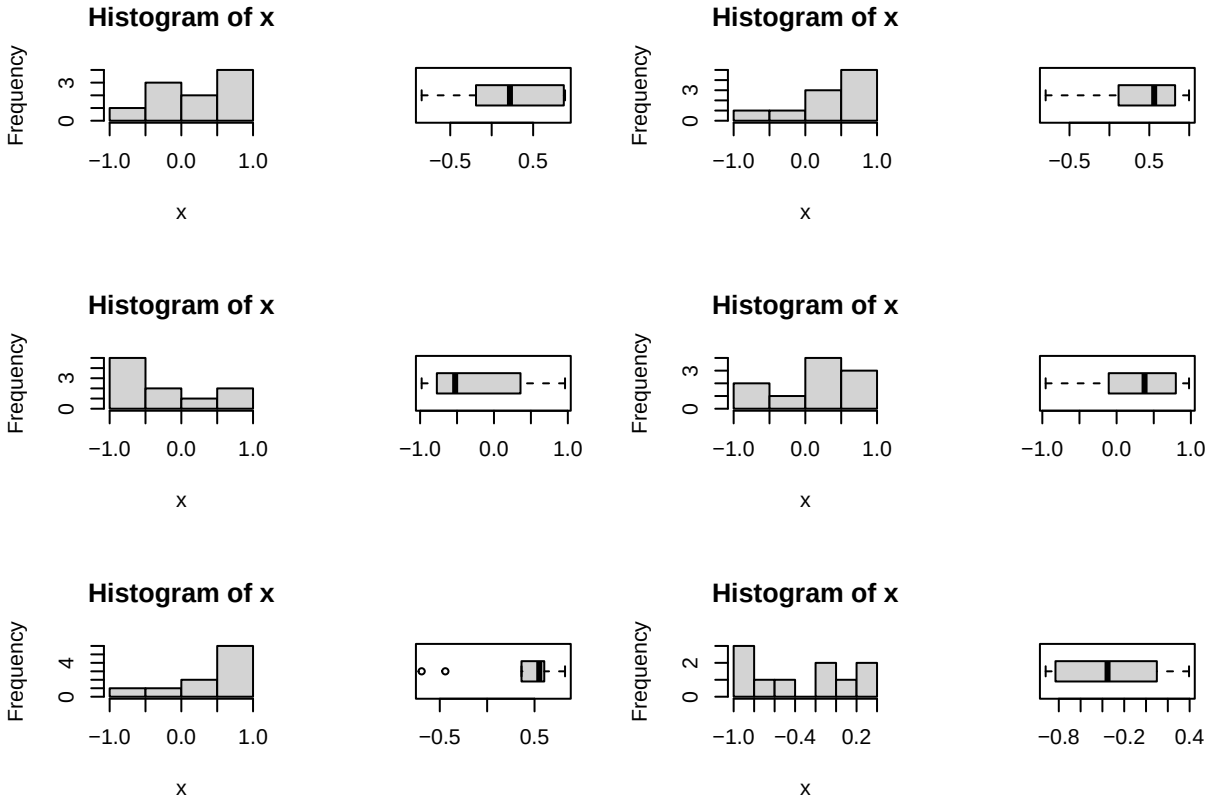
3. Répétez 100 fois et construisez un histogramme et une boîte à moustaches. Quelle est la loi dumaximum, M = max 1 i n X i où X i U(−1, 1) (TD1) ? Superposer la densité théoreique sur l'histogramme. Que remarquez-vous ?

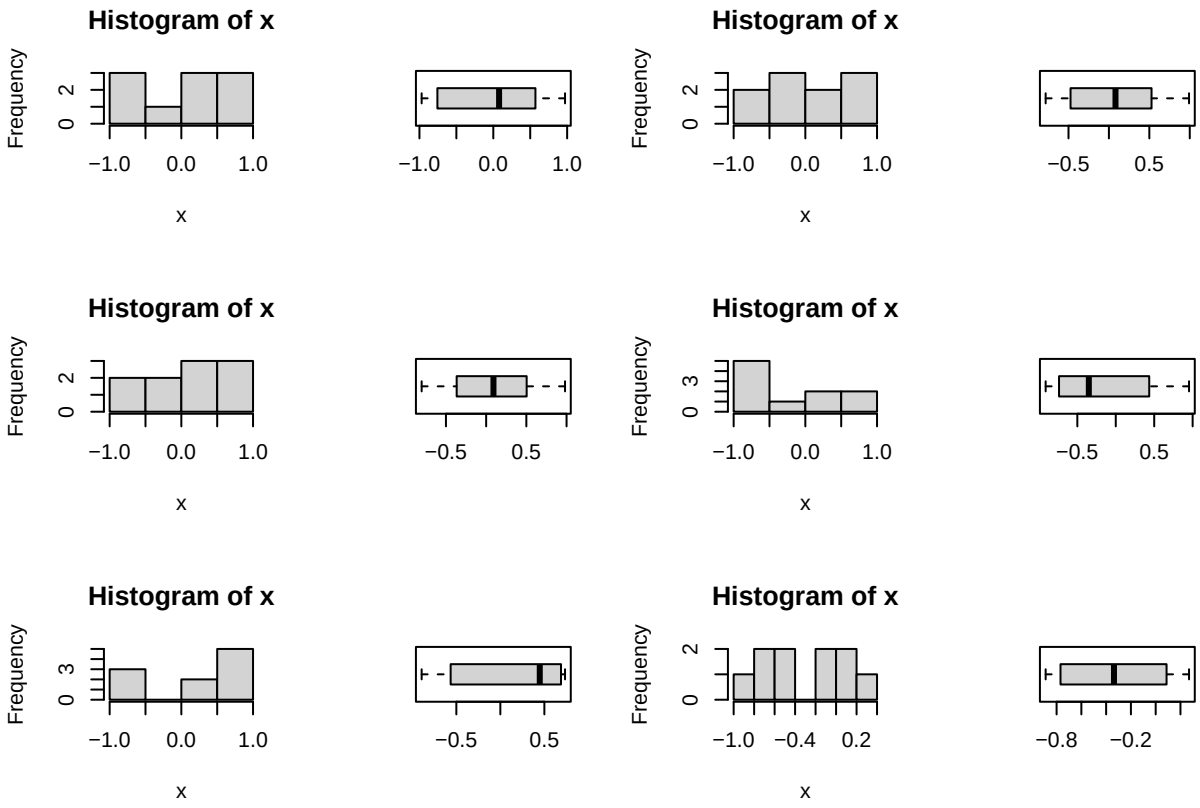
```
par(mfrow=c(3,4))
for (i in 1:100) {
  x <- runif(10, -1, 1)
  hist(x)
  boxplot(x, horizontal = TRUE)
}
```

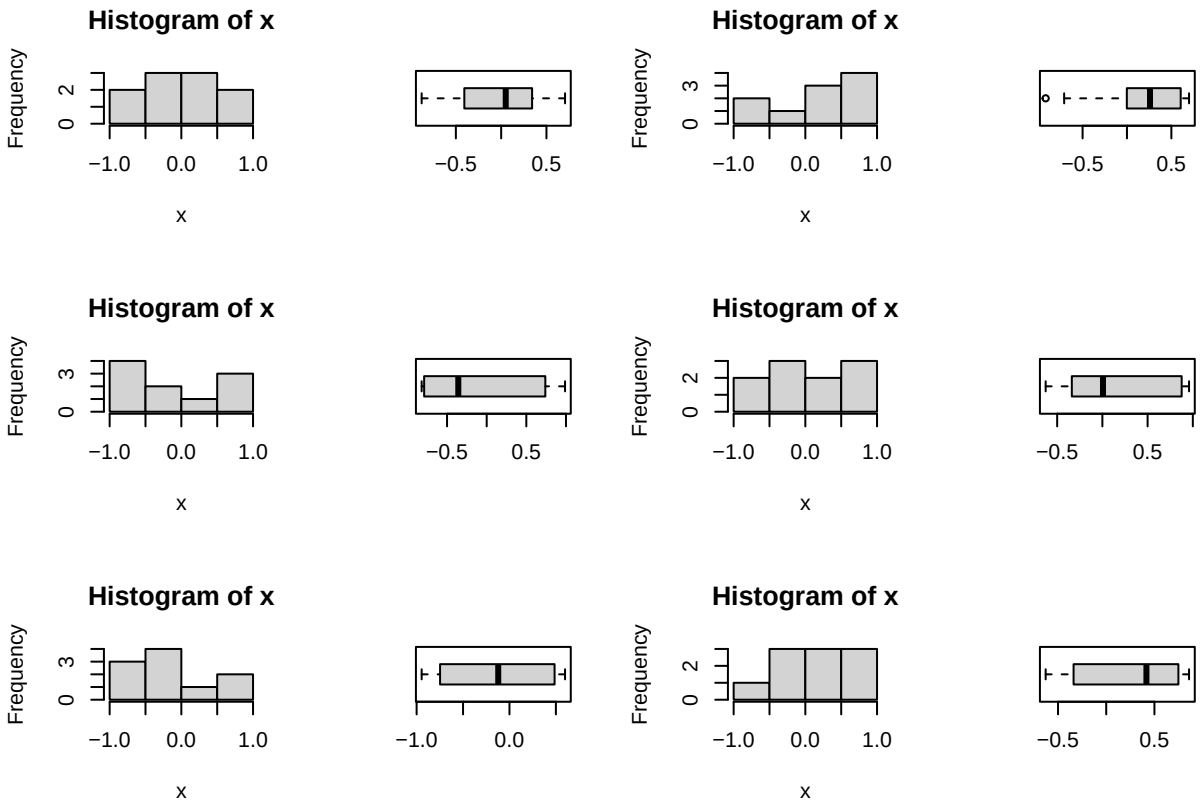

Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
x
−0.8
−0.2
0.4
Histogram of x
Frequency
0
2
−0.5
0.5
x
−0.6
0.0
0.6
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
−0.4
0.2
x
−0.8
−0.2
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0

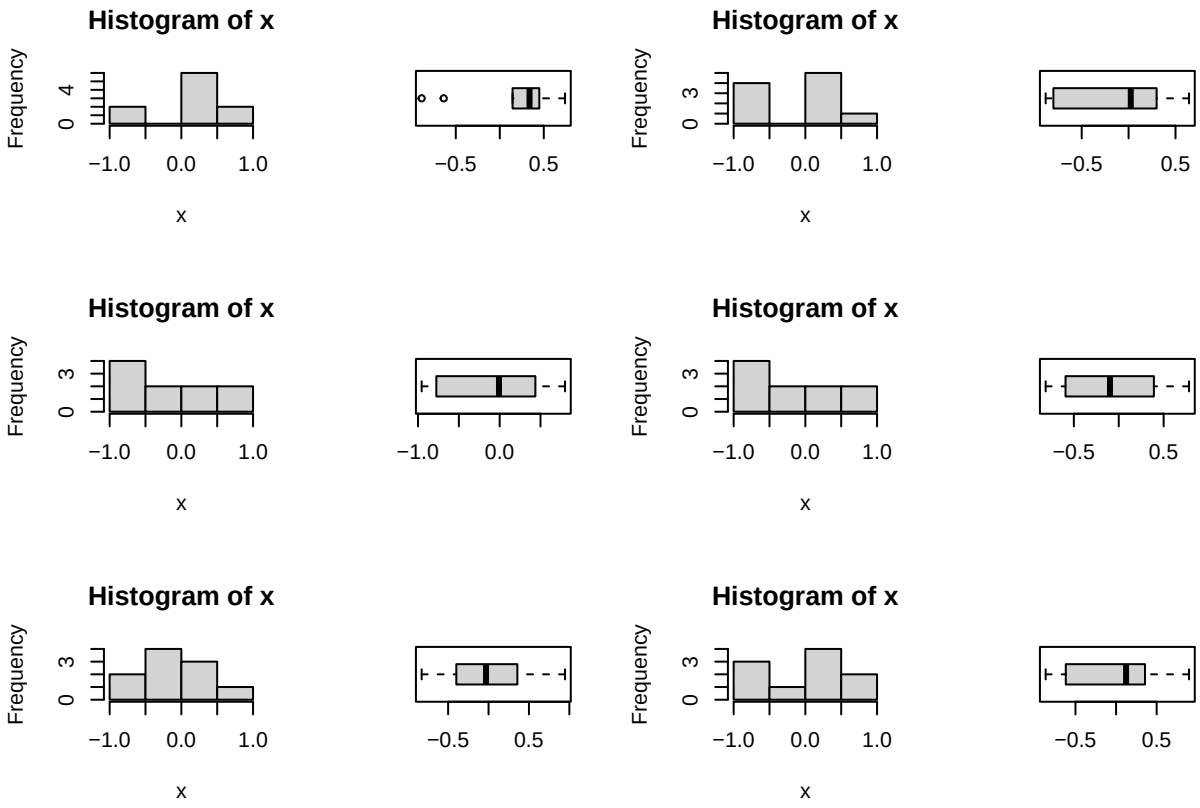

Histogram of x
Frequency
0 4
−1.0 0.0 1.0
x
−1.0 0.0
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−1.0 0.0
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−1.0 0.0 1.0
Histogram of x
Frequency
0 2
−0.5 0.5
x
−0.6 0.0 0.6
Histogram of x
Frequency
0 4
−1.0 0.0 1.0
x
−0.5 0.5

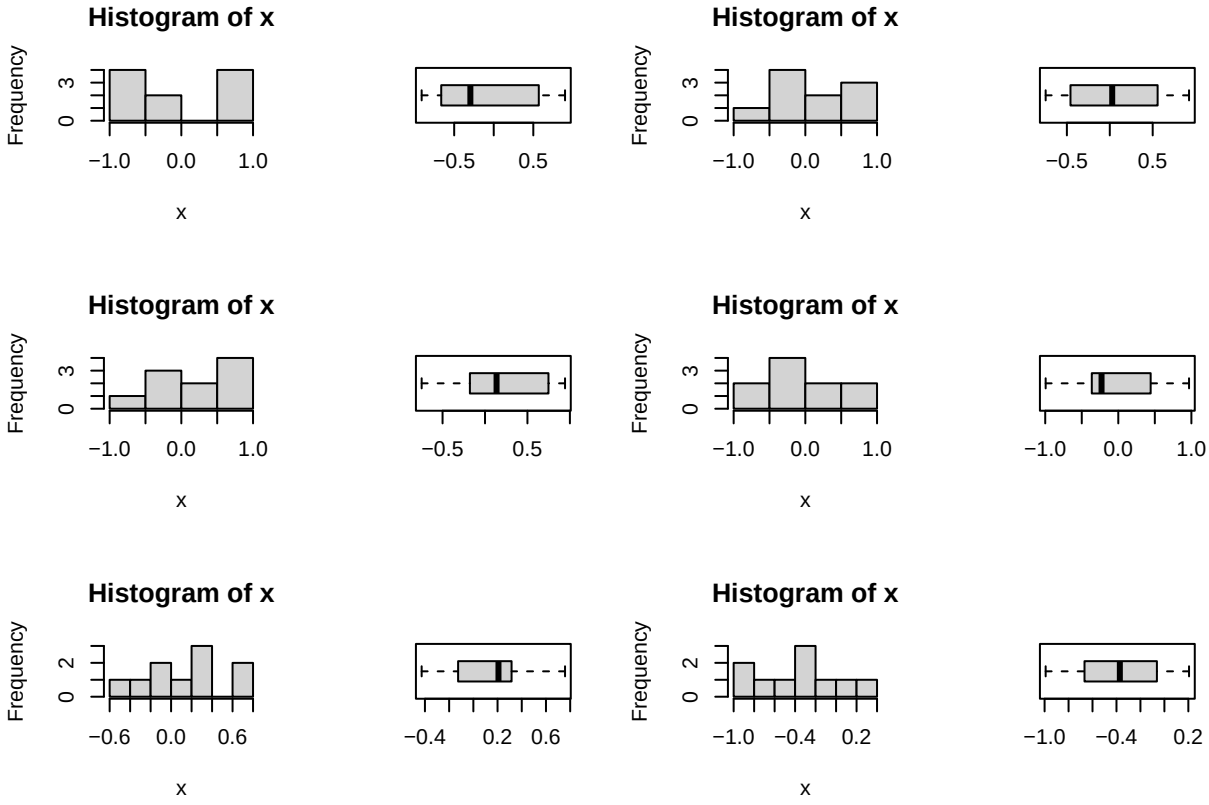

Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−1.0
0.0

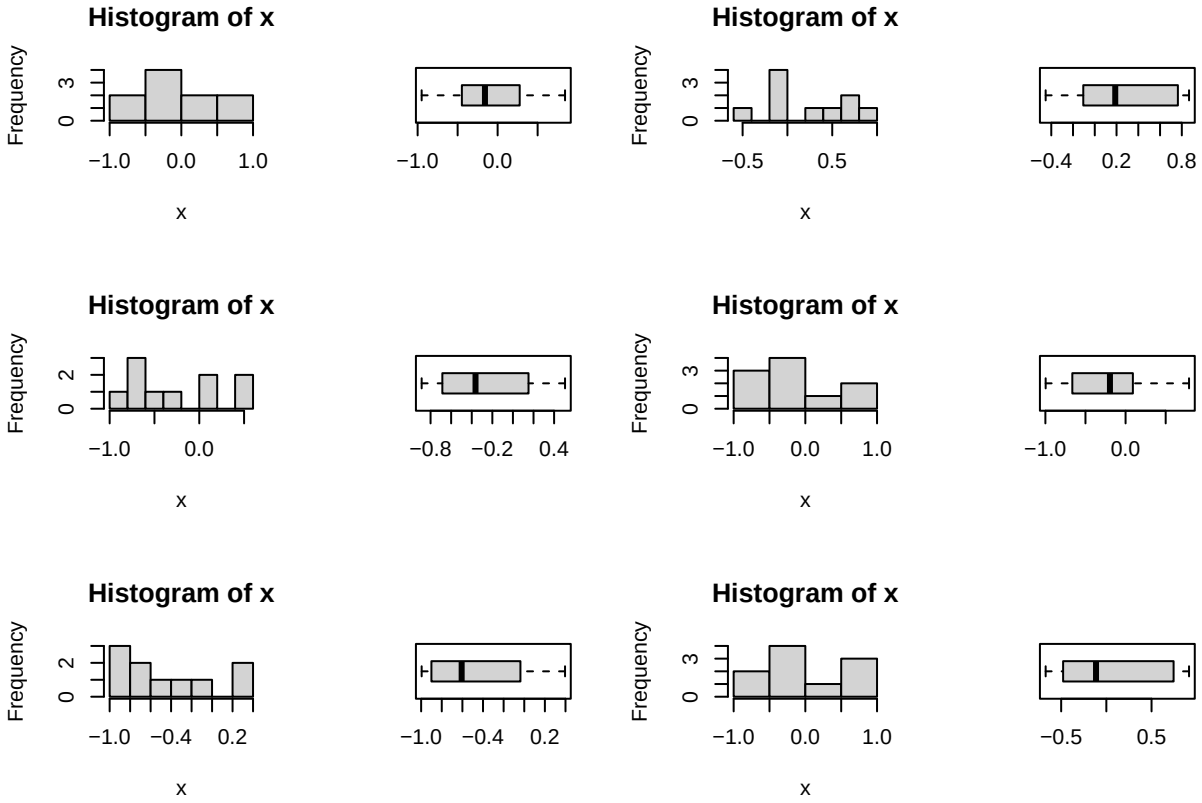

Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
4
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
−0.4
0.2
x
−0.8
−0.2
0.4

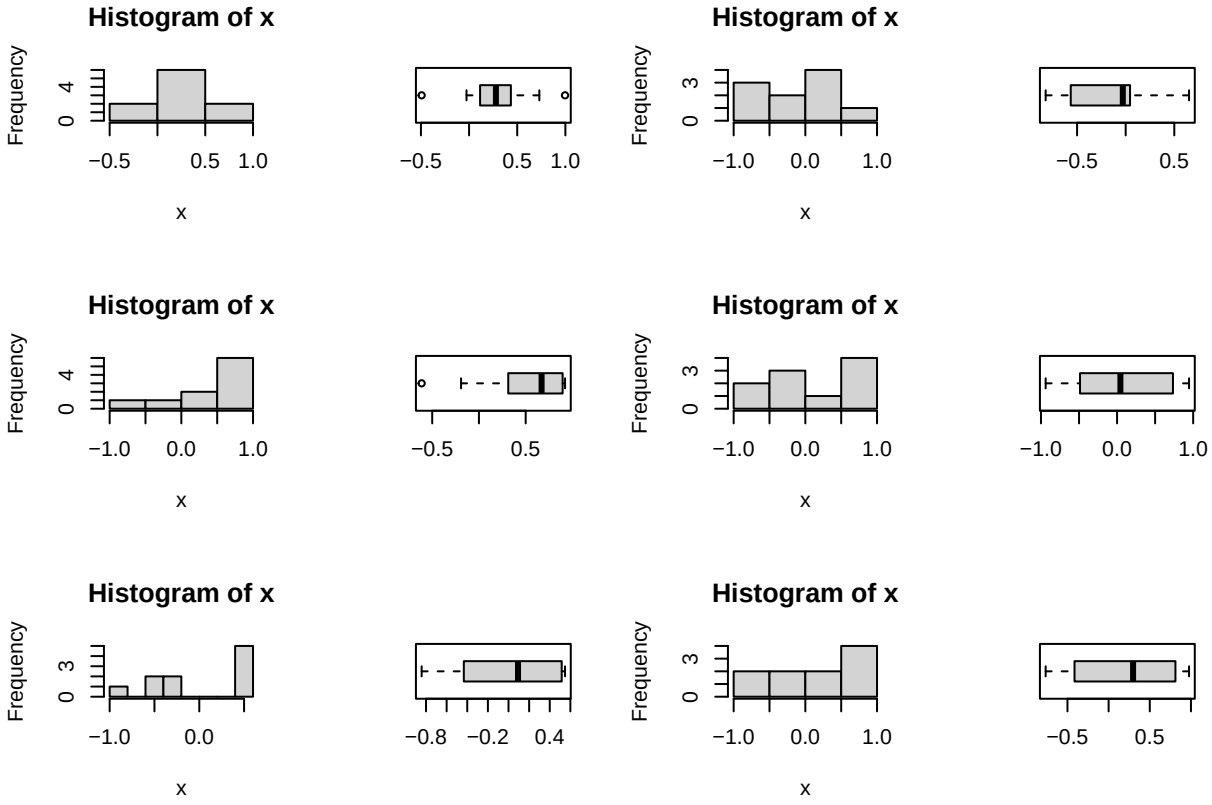

Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
−0.4
0.2
x
−0.8
−0.2

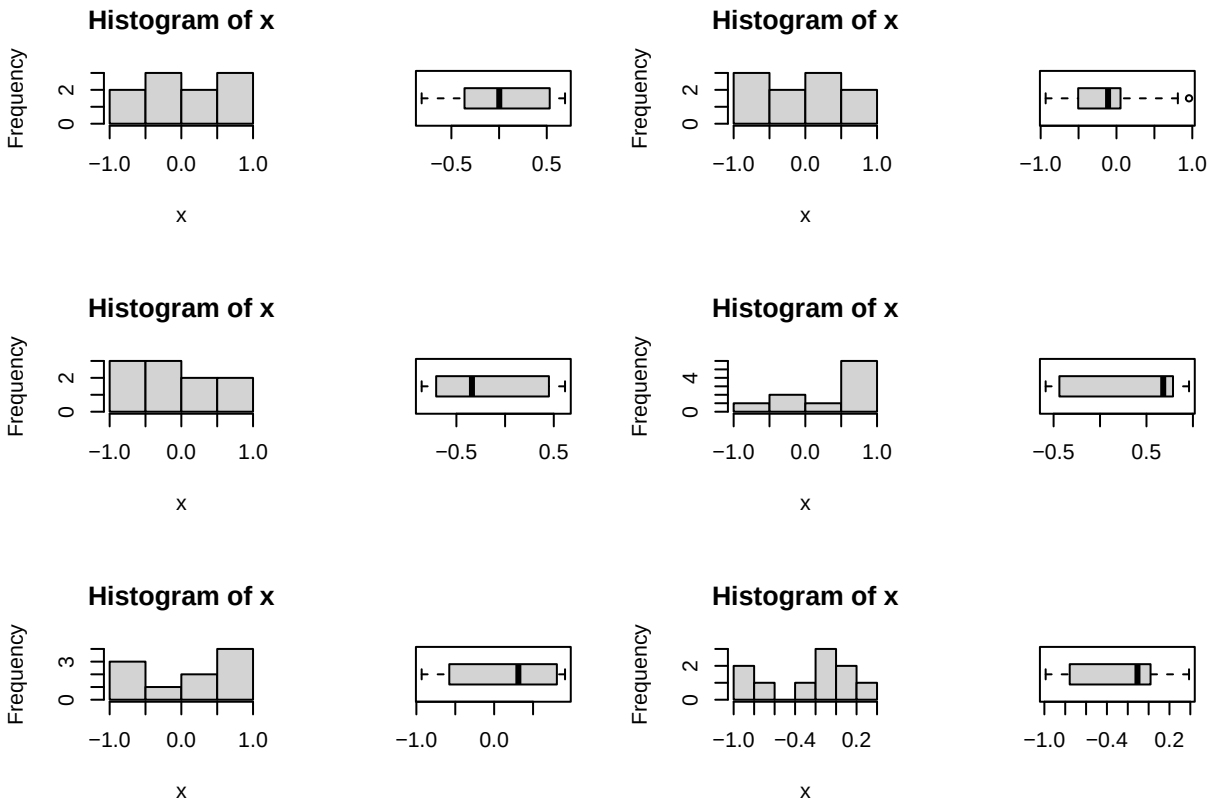

Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5

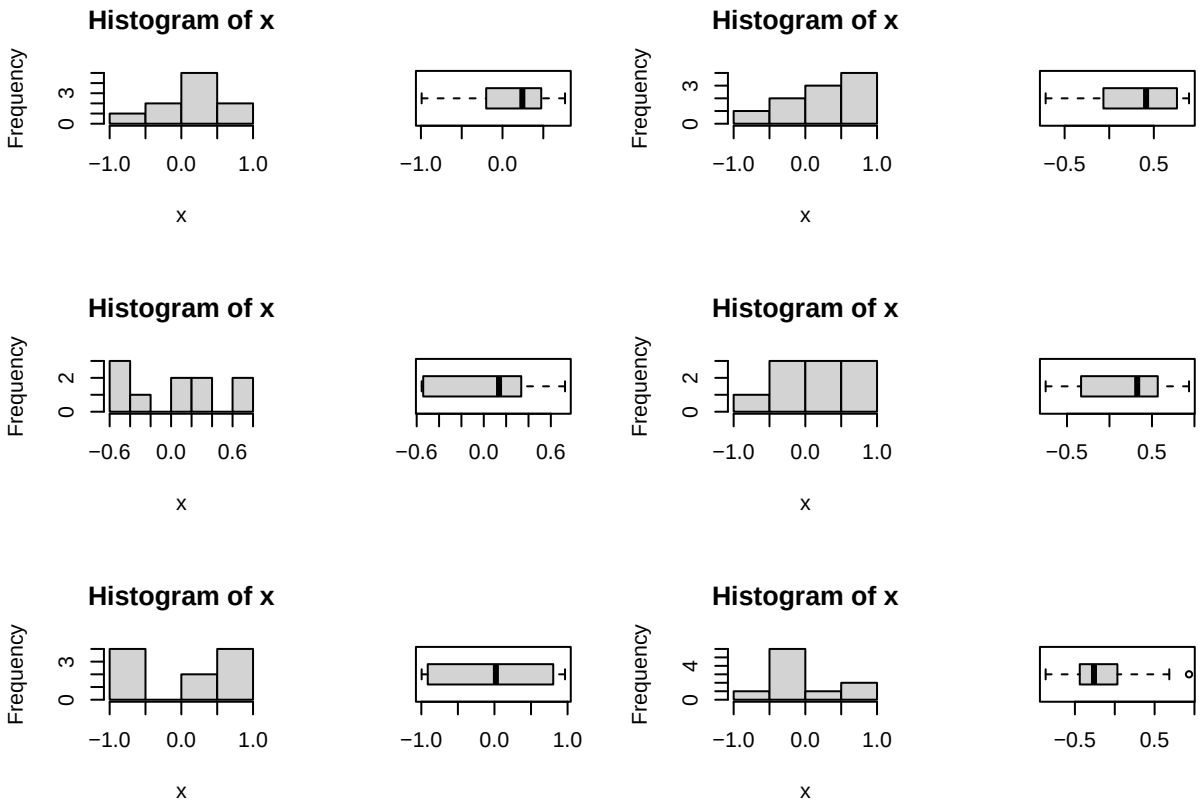

Histogram of x
Frequency
0 4
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−1.0 0.0
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−0.5 0.5

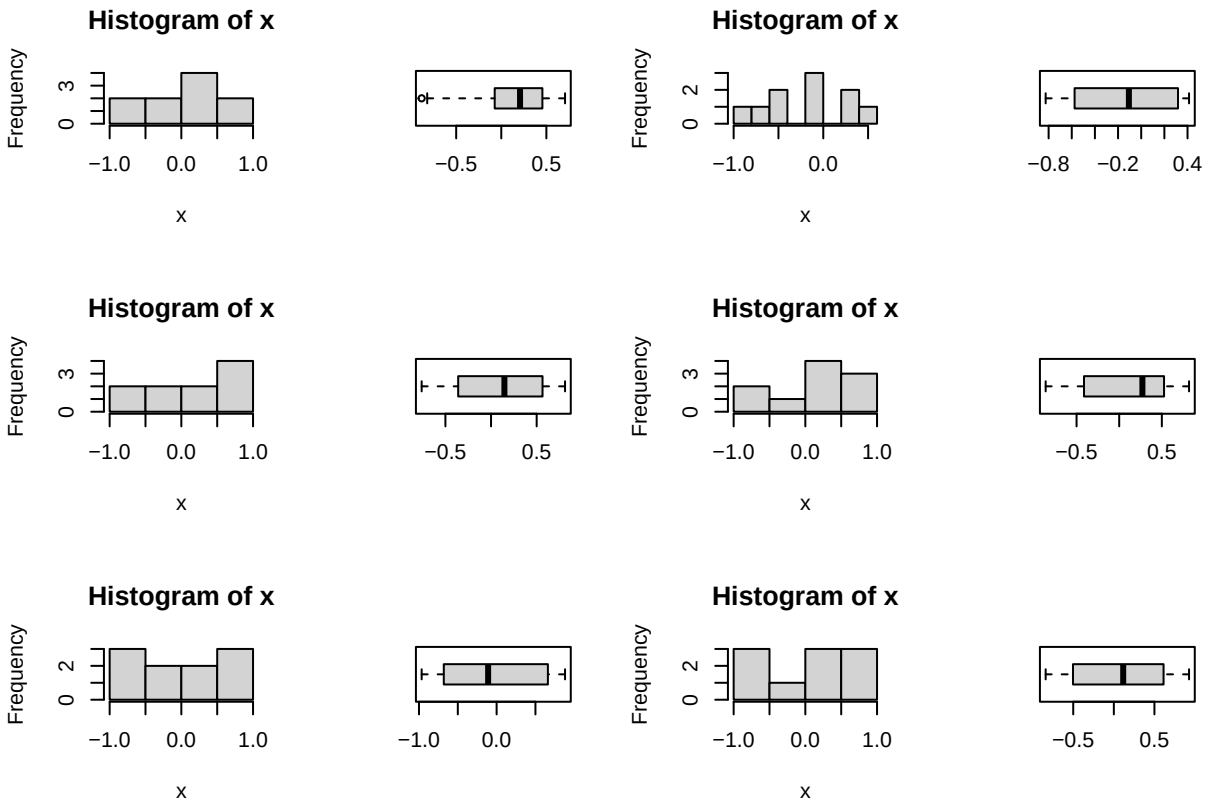

Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
2
−0.6
0.0
0.6
x
−0.4
0.2
0.6
Histogram of x
Frequency
0
2
−1.0
−0.4
0.2
x
−1.0
−0.4
0.2

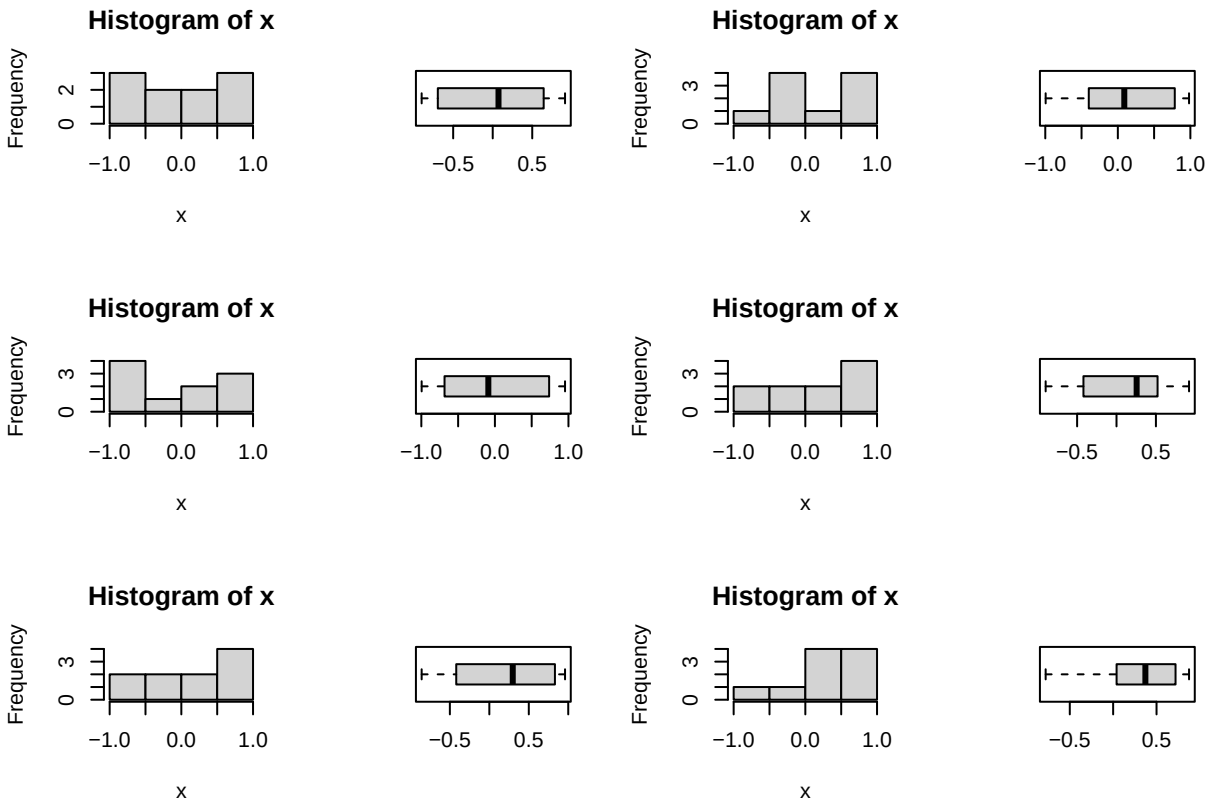

Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
Histogram of x
Frequency
0
3
−0.5
0.5
x
−0.4
0.2
0.8
Histogram of x
Frequency
0
2
−1.0
0.0
x
−0.8
−0.2
0.4
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
Histogram of x
Frequency
0
2
−1.0
−0.4
0.2
x
−1.0
−0.4
0.2
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5

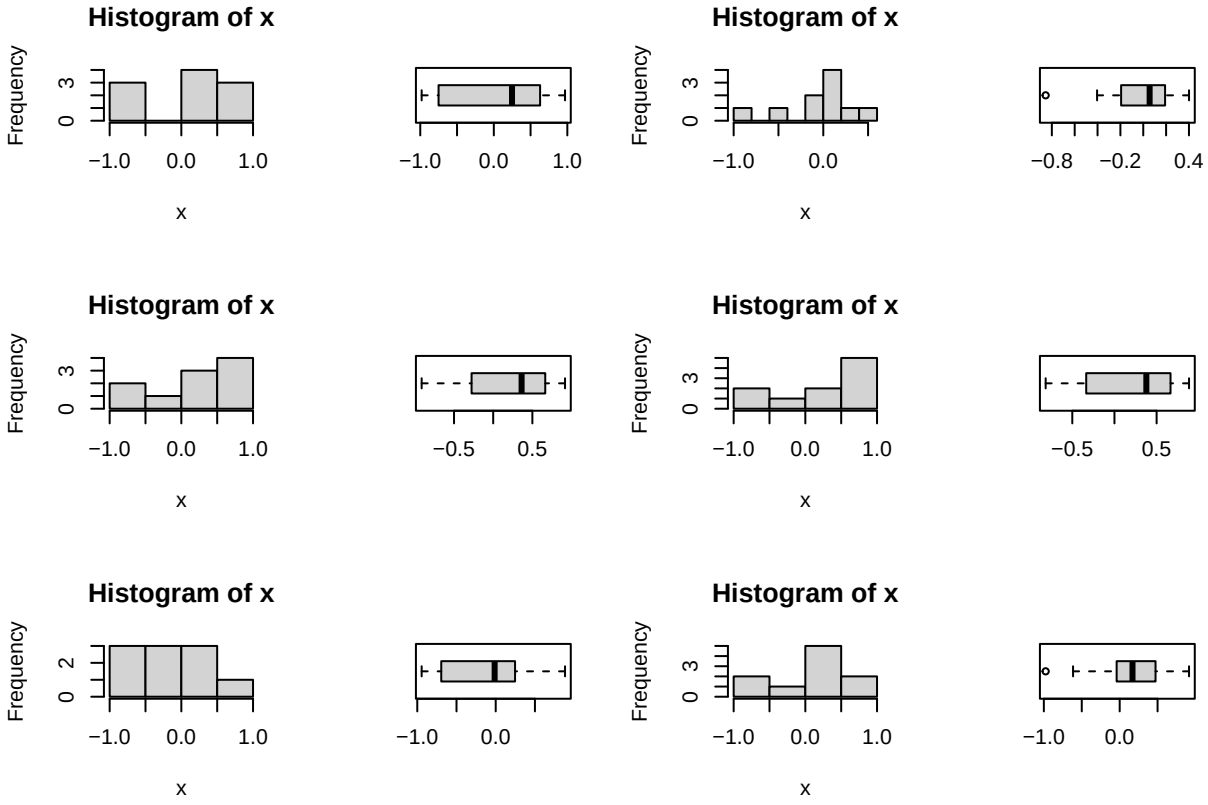

Histogram of x
Frequency
0 4
−0.5 0.5 1.0
x
−0.5 0.5 1.0
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 4
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−1.0 0.0 1.0
Histogram of x
Frequency
0 3
−1.0 0.0
x
−0.8 −0.2 0.4
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−0.5 0.5

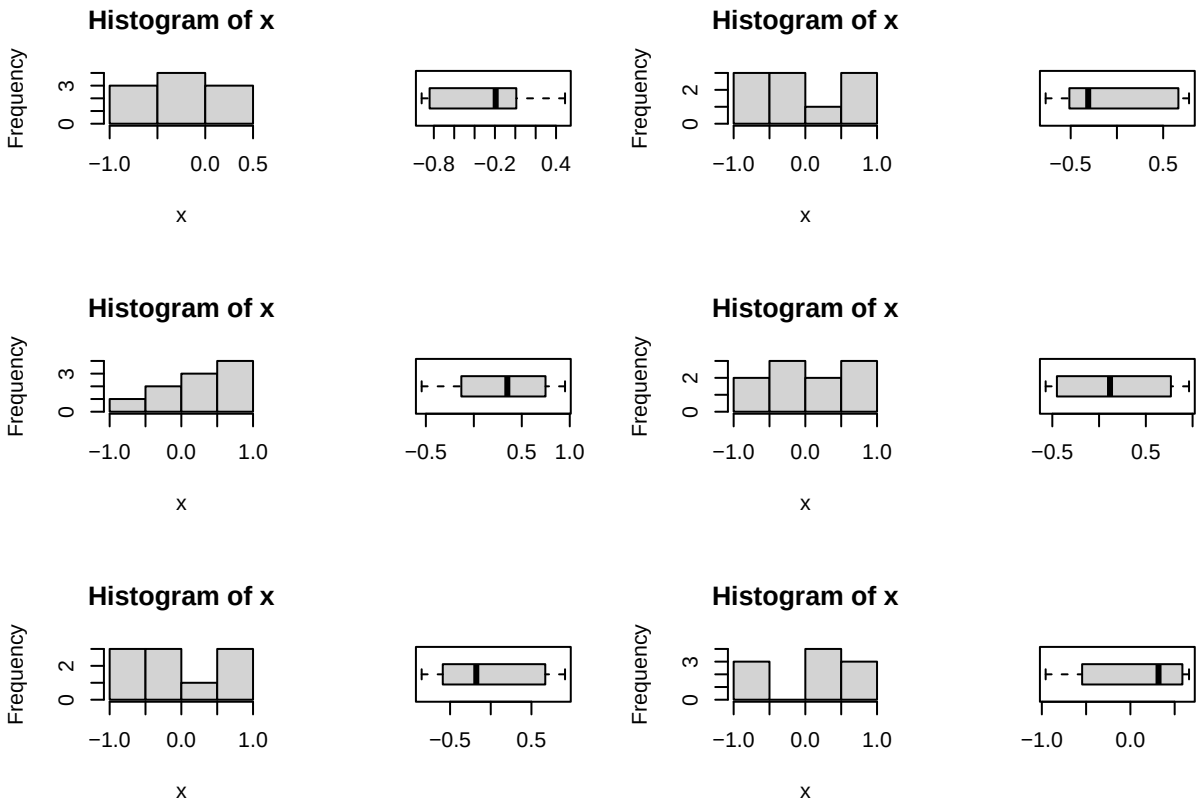

Histogram of x
Frequency
0 2
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 2
−1.0 0.0 1.0
x
−1.0 0.0 1.0
Histogram of x
Frequency
0 2
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 4
−1.0 0.0 1.0
x
−0.5 0.5
Histogram of x
Frequency
0 3
−1.0 0.0 1.0
x
−1.0 0.0
Histogram of x
Frequency
0 2
−1.0 −0.4 0.2
x
−1.0 −0.4 0.2

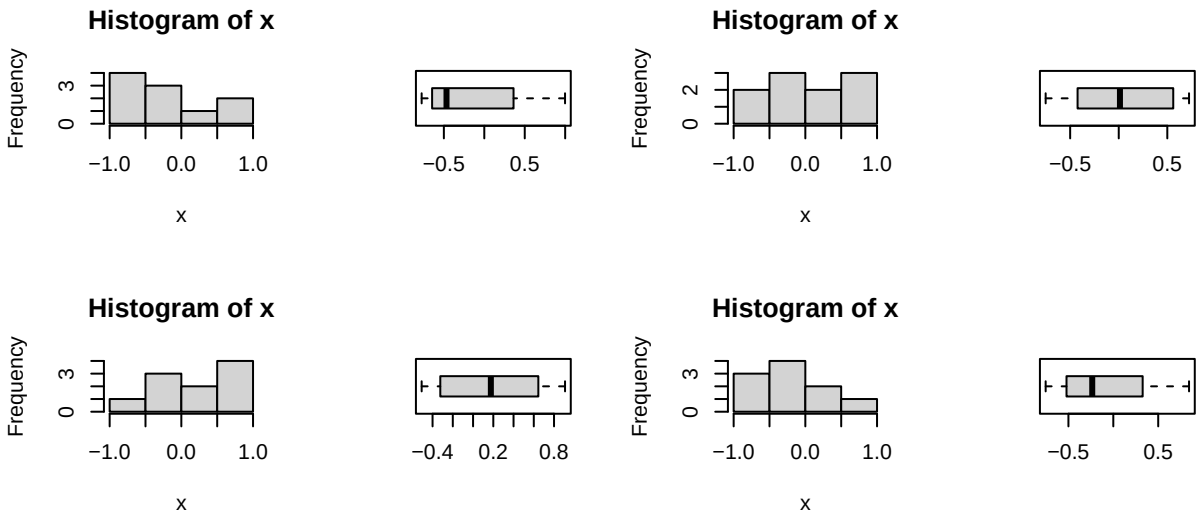

Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−0.6
0.0
0.6
x
−0.6
0.0
0.6
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
4
−1.0
0.0
1.0
x
−0.5
0.5


Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
x
−0.8
−0.2
0.4
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−1.0
0.0
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5


Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5


Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0
1.0
Histogram of x
Frequency
0
3
−1.0
0.0
x
−0.8
−0.2
0.4
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−1.0
0.0
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0


Histogram of x
Frequency
0
3
−1.0
0.0
0.5
x
−0.8
−0.2
0.4
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−0.5
0.5
1.0
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
2
−1.0
0.0
1.0
x
−0.5
0.5
Histogram of x
Frequency
0
3
−1.0
0.0
1.0
x
−1.0
0.0




4. Augmentez la taille de votre échantillon à 50 et répétez votre expérience. Que remarquez-vous? Sont-ils proches de la symétrie ?

```
x <- runif(50, -1, 1)
```

# Monte Carlo Methods

## Moyenne et phénomène de concentration.

1. Donner une borne de cette quantité en utilisant l'inégalité de Bienaymé Chebychev.

Inégalité de Bienaymé Chebychev

$$P(|\hat{\theta} - \theta| \geq \delta) \leq \frac{V(\hat{\theta})}{\delta^2}$$

Calculons la variance grâce à son caractère quadratique ainsi qu'à l'indépendance des X_i :

$$V(\hat{\theta}) = V(\frac{1}{n}\sum_{i=1}^{n}\psi(X)) = \frac{1}{n^2}V(\sum_{i=1}^{n}\psi(X)) = \frac{1}{n^2}\sum_{i=1}^{n}V(\psi(X)) = \frac{1}{n^2} * n\sigma^2 = \frac{\sigma^2}{n}$$

On retrouve donc une borne pour cette inégalité.

2. En supposant que a (Xi) b, donner une borne en utilisant l'inégalité de Hoeffding.

Posons :

$$S_n = \sum_{k=1}^{n} \psi(X_k)$$

D'après l'énoncé, nous savons que :

$$\frac{1}{n}\sum_{i=1}^{n} \psi(X_i) = \hat{\theta}$$

Ainsi :

$$\frac{S_n}{n} = \hat{\theta}$$

Donc :

$$S_n = n\hat{\theta}$$

D'après l'inégalité de Hoeffding nous savons que :

$$P(|S_n - E(S_n)| \geq t) \leq 2exp(\frac{-2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2})$$

Calculons l'esperance de S_n:

$$E(S_n) = E(\sum_{k=1}^{n} \psi(X_k)) = \sum_{k=1}^{n} E(\psi(X_k)) = \sum_{k=1}^{n} \theta = n\theta$$

Ainsi :

$$P(|n\hat{\theta} - n\theta| \geq t) \leq 2exp(\frac{-2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2})$$

$$P(|\hat{\theta} - \theta| \geq \frac{t}{n}) \leq 2exp(\frac{-2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2})$$

On pose : $\frac{t}{n} = \delta$

$$P(|\hat{\theta} - \theta| \geq \delta) \leq 2exp(\frac{-(2n\delta)^2}{\sum_{k=1}^{n}(b_k - a_k)^2})$$

3. De combien d'échantillons auriez-vous besoin pour que la probabilité que $= 2$ soit inférieur à 1%.

Servons-nous de l'inégalité de Bienaymé-Tchebychev que nous avons trouvé lors du 1.

$$P(|\hat{\theta} - \theta| \geq \delta) \leq \frac{\delta^2}{n\sigma^2}$$

Remplaçons par :

$$\delta = 2\sigma$$

$$\delta^2 = 4\sigma^2$$

Ainsi :

$$\frac{\sigma^2}{n\delta^2} = \frac{1}{4n}$$

Or l'énoncé demande à ce que la probabilité soit inférieur à 1% , c'est-à-dire que :

$$\frac{1}{4n} = 0.01 \Longrightarrow n = 25$$

Afin que la probabilité soit inférieur à 1%, il faut 25 échantillons.

## Application pour l'estimation de probabilité

1. Pour la question 1 :

Le paramètre d'intérêt est la moyenne de la fonction exponentielle, c'est à dire 2

Un estimateur serait :

$$E(\epsilon(\theta)) = \frac{1}{\theta}\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i\hat{\theta_n} = \frac{1}{\bar{X_n}}$$

La moyenne empirique tend presque surement vers la moyenne , ainsi l'estimateur est donc consistant. Il faut ensuite utiliser l'inégalité de Hoeffding.

L'énoncé demande à ce que E(Z) = garantie probabiliste de l'erreur. Il faut trouver Z .

On pose Z =

$$Z = \psi(|\hat{\theta} - \theta|)$$

En calculant l'esperance nous avons donc :

$$E(Z) = \int \psi(|\hat{\theta} - \theta|)f(x)dx$$

## Théorème Central Limite et Estimation Monte Carlo

1. Vérifier que l'espérance théorique d'une loi de Pareto est E [X] = a/ −1.

$$P(X \leq t) = (1 - \left(\frac{a}{t}\right)^{\alpha}),\ avec\ x \geq a$$

Donc :

$$E(X) = \int_0^{+\infty} 1 - P(X \leq t)dt = \int_0^{+\infty} P(X > t)dt = a + a^{\alpha}\int_a^{+\infty} \frac{1}{t^{\alpha}}dt = a + \frac{a}{\alpha - 1} = \frac{\alpha a}{\alpha - 1}$$

2. Simuler N = 1000 échantillons i.i.d de loi commune Pareto P(a, ) (avec votre choix de paramètres)

de taille n = 5, 30, 100 et calculer les moyennes et variances empiriques ¯Xn,i et Sn,i, i = 1, . . . ,N.

```
##
## Attachement du package : 'EnvStats'

## Les objets suivants sont masqués depuis 'package:stats':
##
##     predict, predict.lm

## L'objet suivant est masqué depuis 'package:base':
##
##     print.default

## [1] "Moyenne empirique n = 5"

## [1] 1.518912 1.510638 1.535794 1.505299 1.501118

## [1] "Moyenne empirique n = 30"
```

```
##  [1] 1.538641 1.483065 1.520743 1.496227 1.537632 1.504322 1.515873 1.566331
##  [9] 1.526323 1.488805 1.471626 1.494973 1.502167 1.497068 1.489579 1.480110
## [17] 1.470246 1.525653 1.466384 1.493883 1.508913 1.528535 1.584898 1.492900
## [25] 1.508871 1.523257 1.545045 1.532436 1.516653 1.500837
```

```
## [1] "Moyenne empirique n = 100"
```

```
##   [1] 1.538599 1.468208 1.482303 1.480237 1.542813 1.509425 1.521382 1.504604
##   [9] 1.535107 1.518828 1.560018 1.510673 1.487809 1.469342 1.482430 1.498384
##  [17] 1.471237 1.520373 1.523940 1.511042 1.474893 1.491754 1.478561 1.562914
##  [25] 1.458626 1.463396 1.492383 1.514316 1.443792 1.534109 1.518094 1.461371
##  [33] 1.501966 1.524796 1.469960 1.514799 1.460608 1.524833 1.526233 1.468527
##  [41] 1.485446 1.496101 1.475213 1.507071 1.524725 1.453330 1.544675 1.495875
##  [49] 1.513097 1.483541 1.499607 1.468099 1.460699 1.518438 1.486279 1.500034
##  [57] 1.500097 1.505795 1.475440 1.525479 1.473785 1.495190 1.488184 1.486683
##  [65] 1.461235 1.546584 1.452632 1.496237 1.479346 1.526385 1.486610 1.503796
##  [73] 1.504457 1.502024 1.539837 1.499566 1.526432 1.455722 1.490602 1.552318
##  [81] 1.496729 1.469700 1.506624 1.494100 1.522802 1.529896 1.518769 1.488578
##  [89] 1.504575 1.535055 1.502109 1.456779 1.494362 1.484670 1.467455 1.498036
##  [97] 1.457543 1.510542 1.487303 1.476691
```

```
## [1] "Variance empirique n = 5"
```

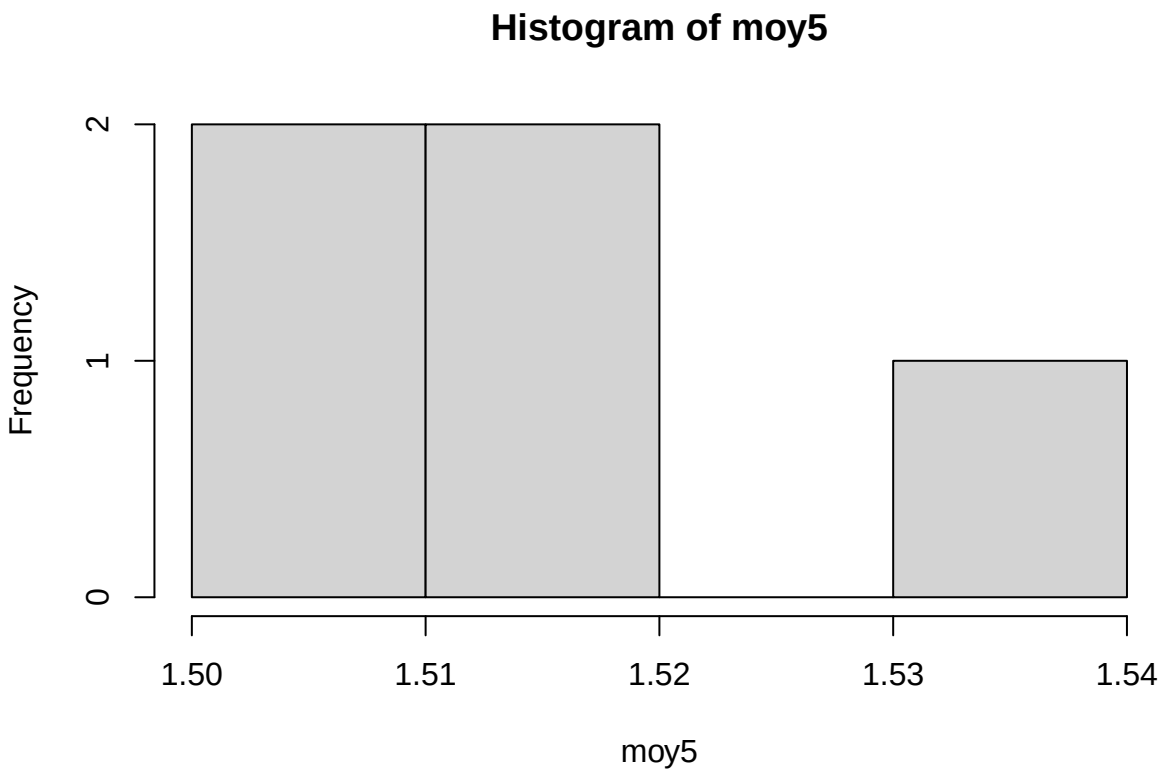
```
## [1] 0.002307094 0.002282029 0.002358663 0.002265925 0.002253354
```

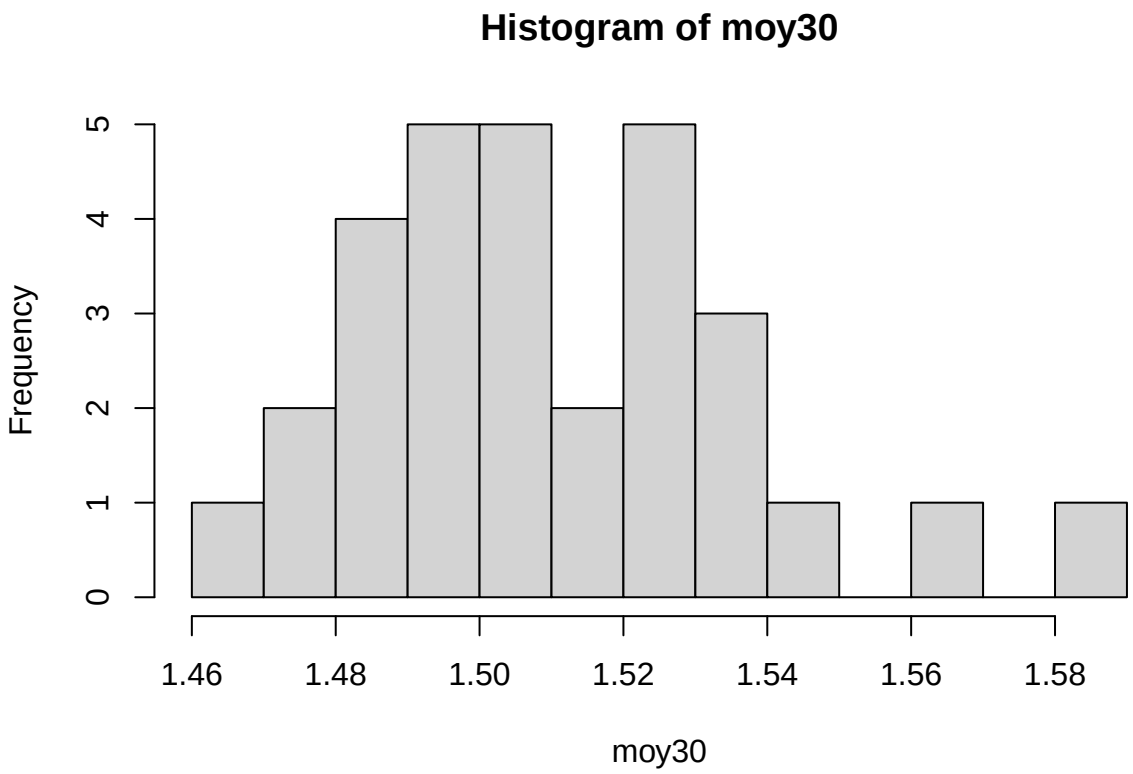
```
## [1] "Variance empirique n = 30"
```

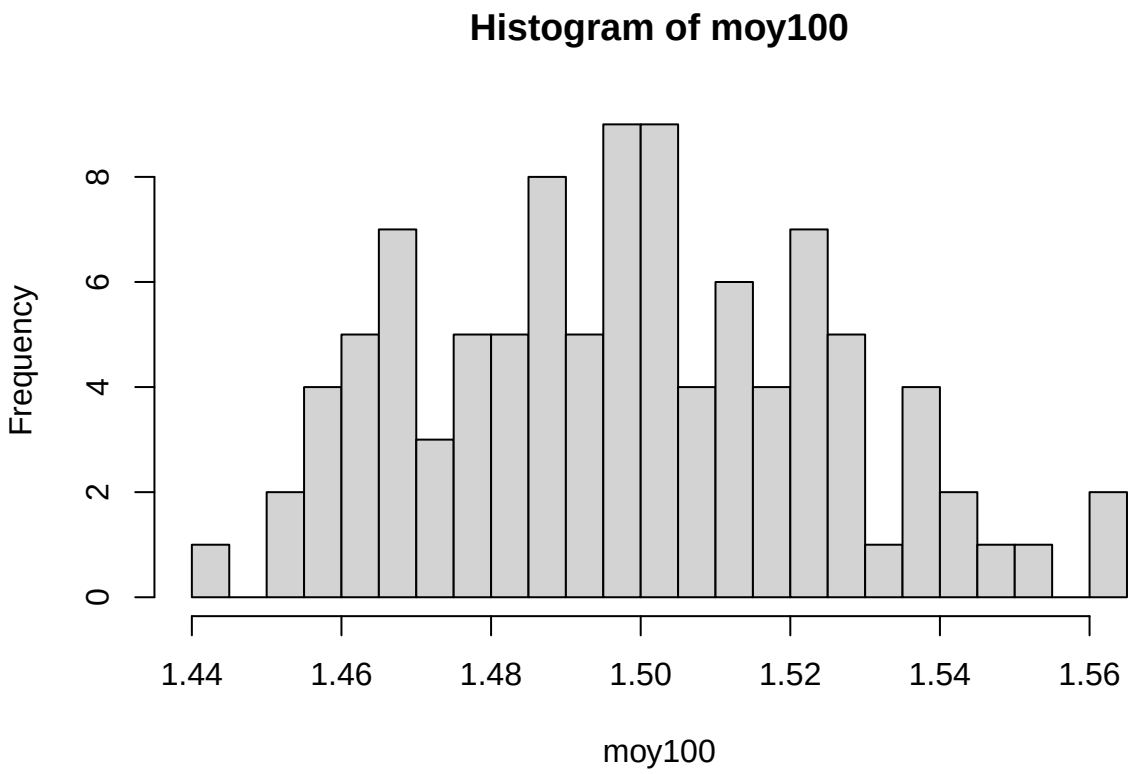
```
##  [1] 0.002367417 0.002199482 0.002312659 0.002238695 0.002364312 0.002262983
##  [7] 0.002297870 0.002453392 0.002329663 0.002216541 0.002165683 0.002234945
## [13] 0.002256504 0.002241212 0.002218847 0.002190726 0.002161622 0.002327618
## [19] 0.002150283 0.002231685 0.002276818 0.002336418 0.002511901 0.002228751
## [25] 0.002276691 0.002320311 0.002387163 0.002348361 0.002300237 0.002252511
```

```
## [1] "Variance empirique n = 100"
```

```
##   [1] 0.002367287 0.002155635 0.002197221 0.002191102 0.002380271 0.002278363
##   [7] 0.002314602 0.002263834 0.002356554 0.002306839 0.002433656 0.002282134
##  [13] 0.002213575 0.002158965 0.002197599 0.002245155 0.002164538 0.002311533
##  [19] 0.002322393 0.002283248 0.002175309 0.002225329 0.002186141 0.002442700
##  [25] 0.002127590 0.002141529 0.002227207 0.002293154 0.002084536 0.002353489
##  [31] 0.002304610 0.002135605 0.002255901 0.002325002 0.002160781 0.002294616
##  [37] 0.002133375 0.002325116 0.002329388 0.002156573 0.002206550 0.002238319
##  [43] 0.002176254 0.002271262 0.002324786 0.002112168 0.002386022 0.002237643
##  [49] 0.002289462 0.002200895 0.002248821 0.002155315 0.002133642 0.002305655
##  [55] 0.002209026 0.002250103 0.002250292 0.002267419 0.002176922 0.002327086
##  [61] 0.002172042 0.002235593 0.002214692 0.002210227 0.002135207 0.002391922
##  [67] 0.002110141 0.002238724 0.002188466 0.002329851 0.002210008 0.002261403
##  [73] 0.002263391 0.002256077 0.002371097 0.002248699 0.002329994 0.002119125
##  [79] 0.002221893 0.002409691 0.002240199 0.002160018 0.002269916 0.002232336
##  [85] 0.002318927 0.002340581 0.002306660 0.002215863 0.002263745 0.002356393
##  [91] 0.002256331 0.002122206 0.002233118 0.002204244 0.002153425 0.002244112
##  [97] 0.002124431 0.002281737 0.002212069 0.002180617
```

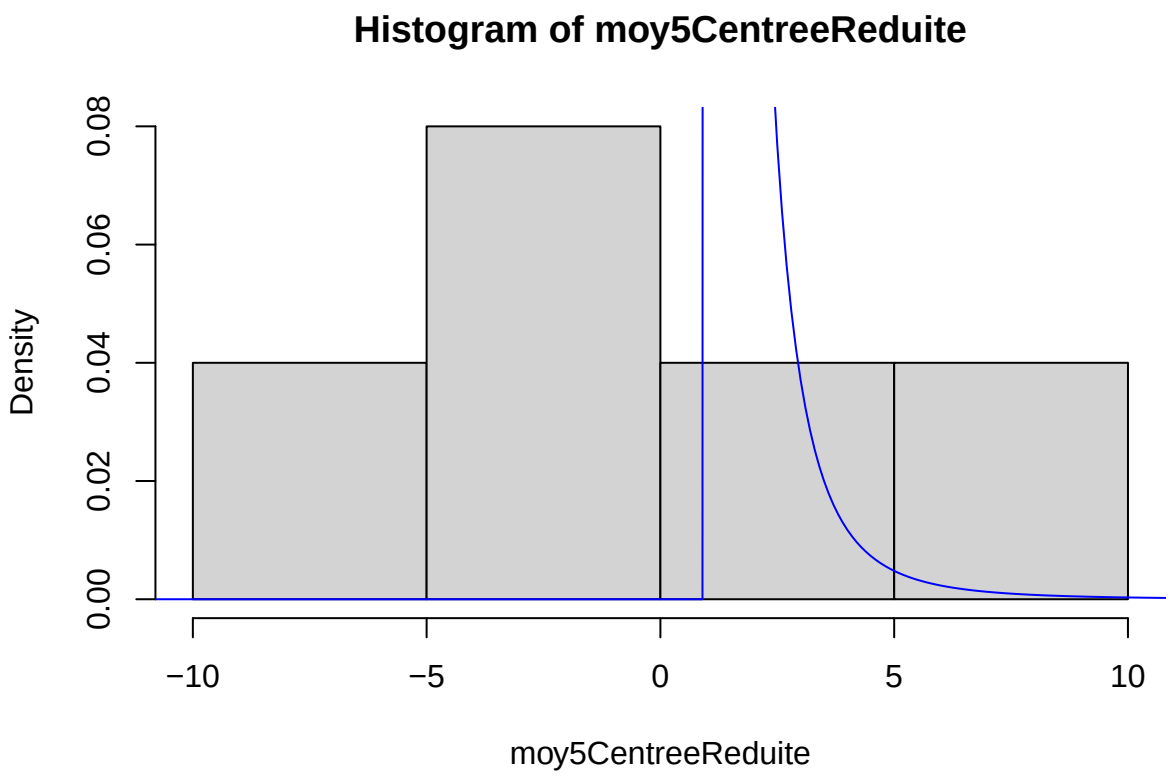

Histogram of moy5
Frequency
0
1
2
1.50
1.51
1.52
1.53
1.54
moy5


Histogram of moy30
Frequency
0
1
2
3
4
5
1.46
1.48
1.50
1.52
1.54
1.56
1.58
moy30

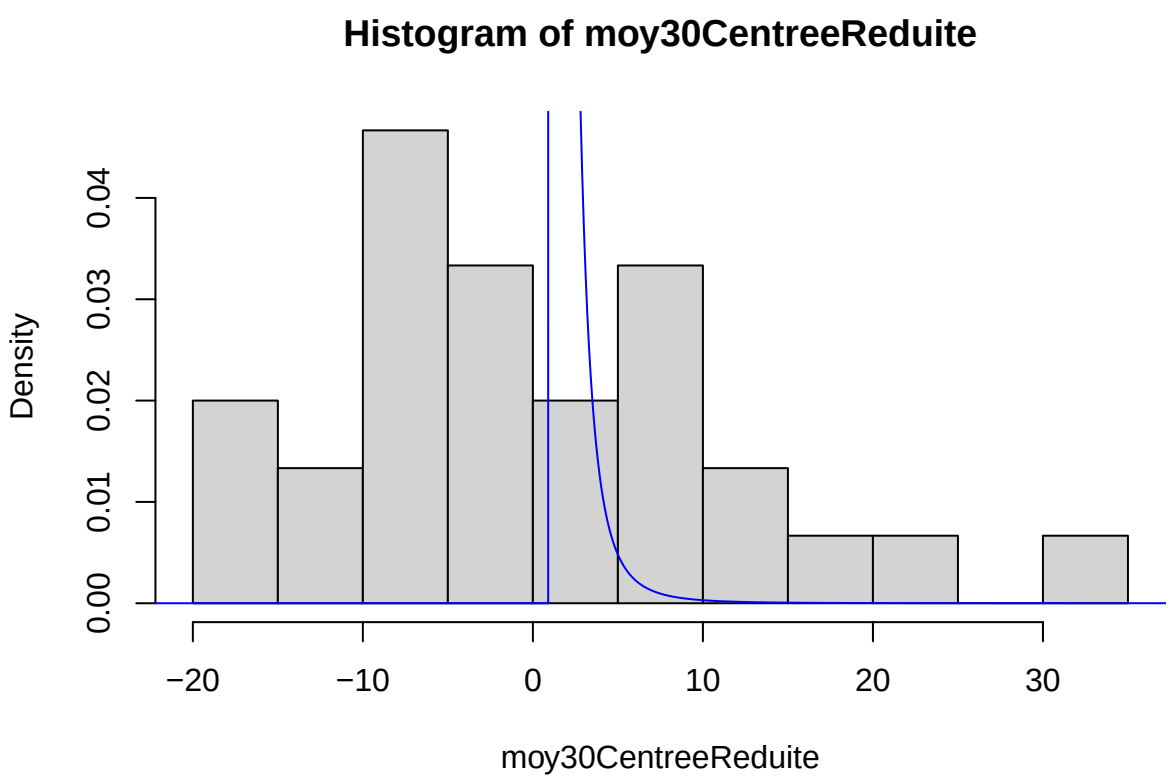



4. A l'aide d'une renormalisation adéquate (an, bn), montrer que Un,i = ¯Xn,i−an/ bn a une loi que vous pouvez approchez. Comparez histogramme de les moyennes empiriques normalisées, Un,i, et distribution théorique approchée. Quelle est l'influence de la taille de l'échantillon n sur la qualité de cette approximation?

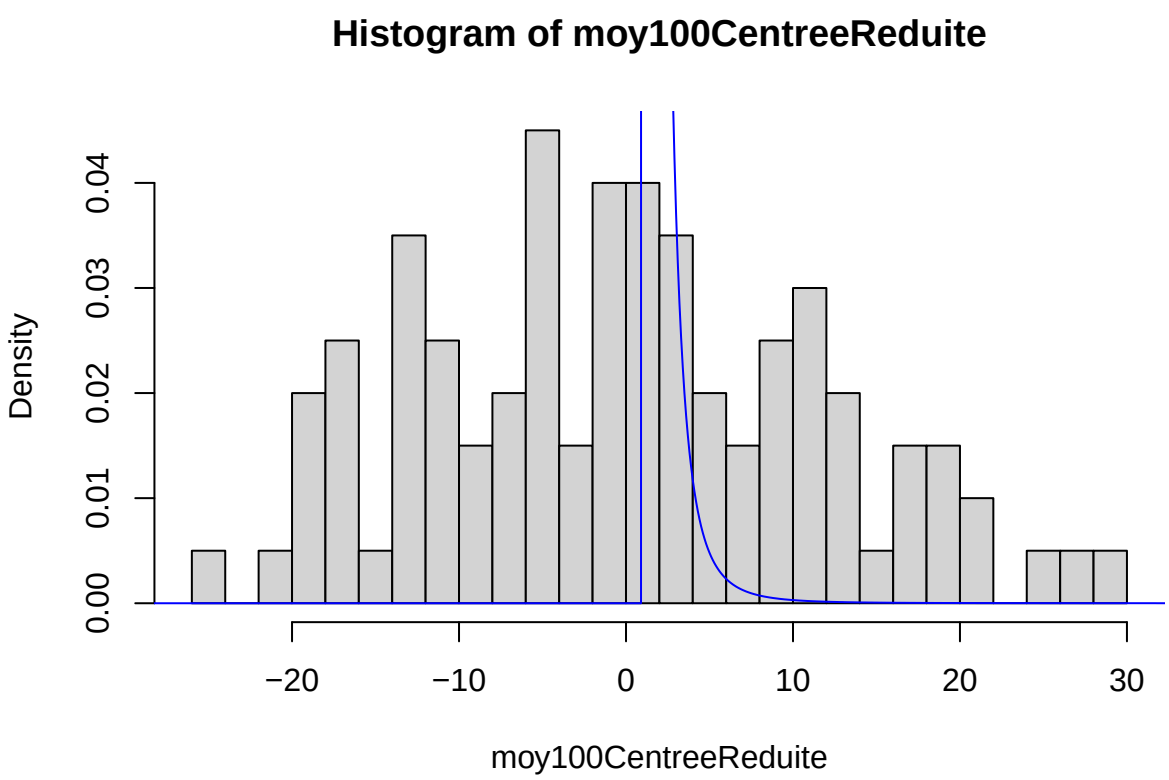
```
moy5CentreeReduite <- (moy5-mean(moy5))/mean(moy5^2/1000)
moy30CentreeReduite <- (moy30-mean(moy30))/mean(moy30^2/1000)
moy100CentreeReduite <- (moy100-mean(moy100))/mean(moy100^2/1000)

hist(moy5CentreeReduite,freq=FALSE)
lines(seq(-50,50,by=0.1),dpareto(seq(-50,50,by=0.1),1,3),col = "blue")
```




```
hist(moy30CentreeReduite,freq=FALSE,breaks = 10)
lines(seq(-50,50,by=0.1),dpareto(seq(-50,50,by=0.1),1,3),col = "blue")
```




```
hist(moy100CentreeReduite,freq=FALSE,breaks = 20)
lines(seq(-50,50,by=0.1),dpareto(seq(-50,50,by=0.1),1,3),col = "blue")
```




### Quand le théorème de central limite ne s'applique pas

1. Simuler un échantillon de taille n = 20 d'une loi de C(2) et calculer la moyenne empirique $\bar{X}n$.

Moyenne empirique:

```
## [1] -20.38199
```

2. Faites varier la taille de l'échantillon n = 20, 100, 1000 et 10000. Qu'en déduire ?

```
## [1] -20.38199
```

```
## [1] 0.7477084
```

```
## [1] -2.899143
```

```
## [1] 104.2197
```

On remarque que malgré le nombre élévé de l'échantillon la moyenne ne semble pas se stabiliser comme pour une loi normale.

3. Expliquer ce comportement

Nous savons d'après le cours de probabilités que la loi de cauchy n'admet pas d'espérance ni d'écart type. Cela explique donc le comportement de la moyenne malgré la taille de l'échantillon.

4. Quelle est la médiane d'une loi de cauchy ?

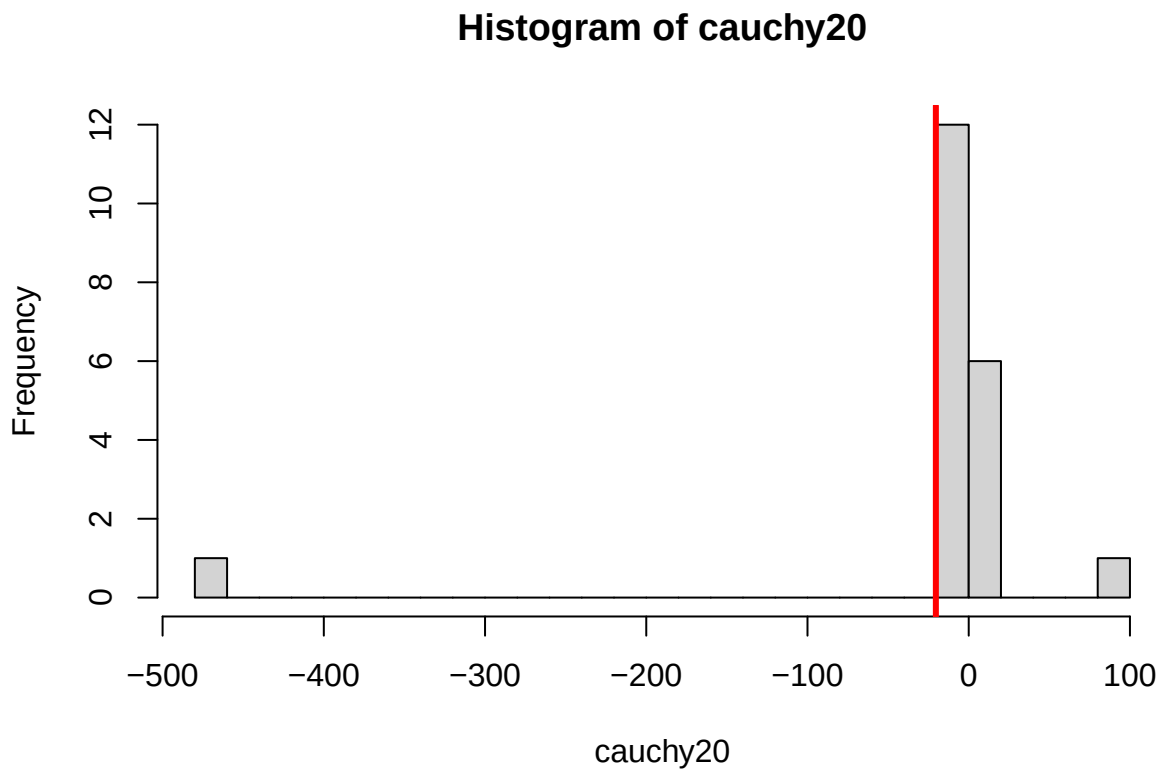
La courbe est symétrique ,la médiane d'une loi de cauchy est . D'après RStudio, quand la position n'est pas défini celle-ci est mise entre à 0. Par conséquent nous devons vérifier si la médiane semble proche de 0.

$$f(x,\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}\frac{1}{2} = \frac{1}{\pi}\int_{-a}^{a}\frac{dx}{1+(x-\theta)^2}F^{-1}(\frac{1}{2}) = \theta$$

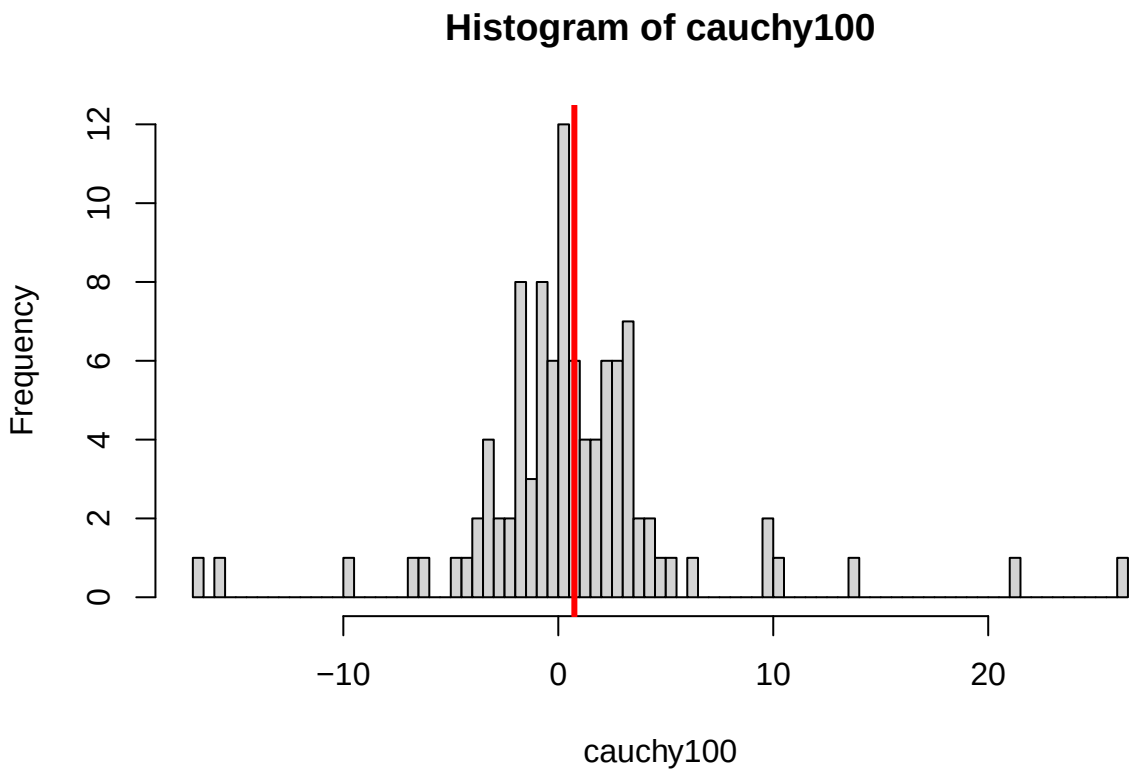
5. En déduire un estimateur de theta et evaluer la performance de cet estimateur sur les différents échantillons.

Nous pouvons essayer d'approximer theta , c'est-à-dire la médiane, cela revient donc à chercher une estimation du quantile en 0.5 . D'après le cours, les quantiles permettent de localiser les valeurs les plus fréquentes. Nous allons donc essayer d'estimer le quantile.

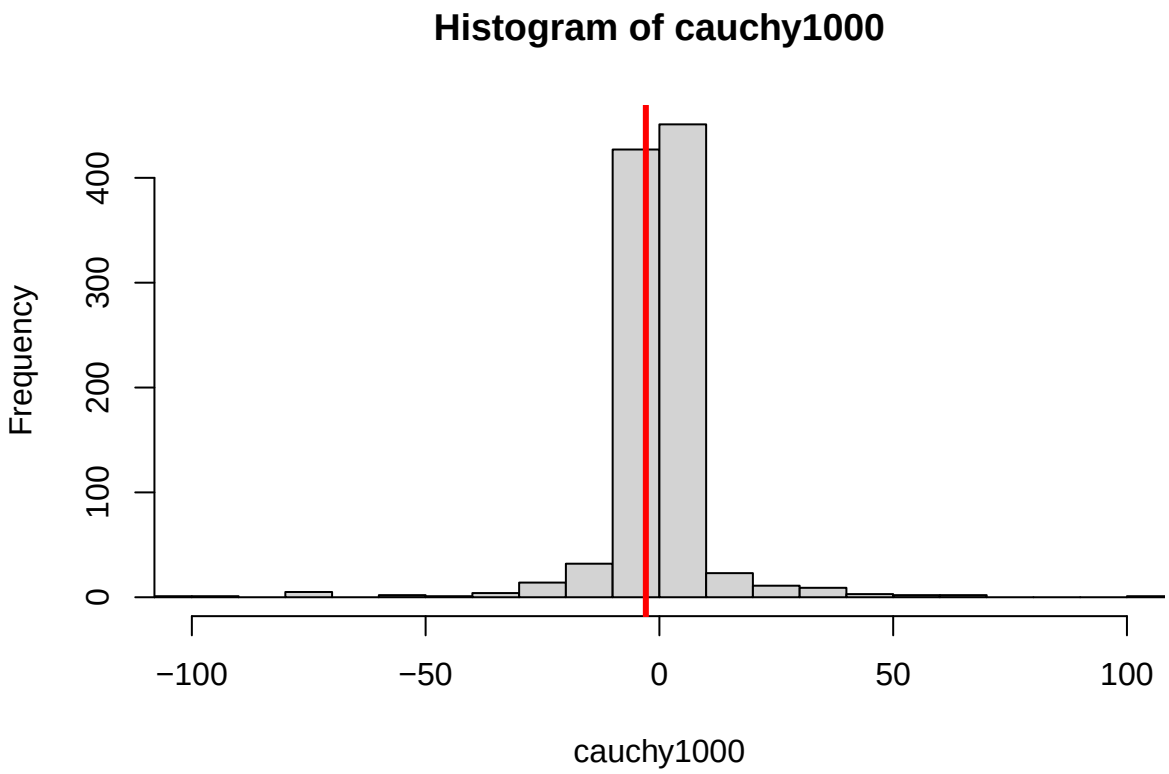
```
hist(cauchy20,breaks=20)
abline(v=mean(cauchy20),col="red",lwd=3)
```



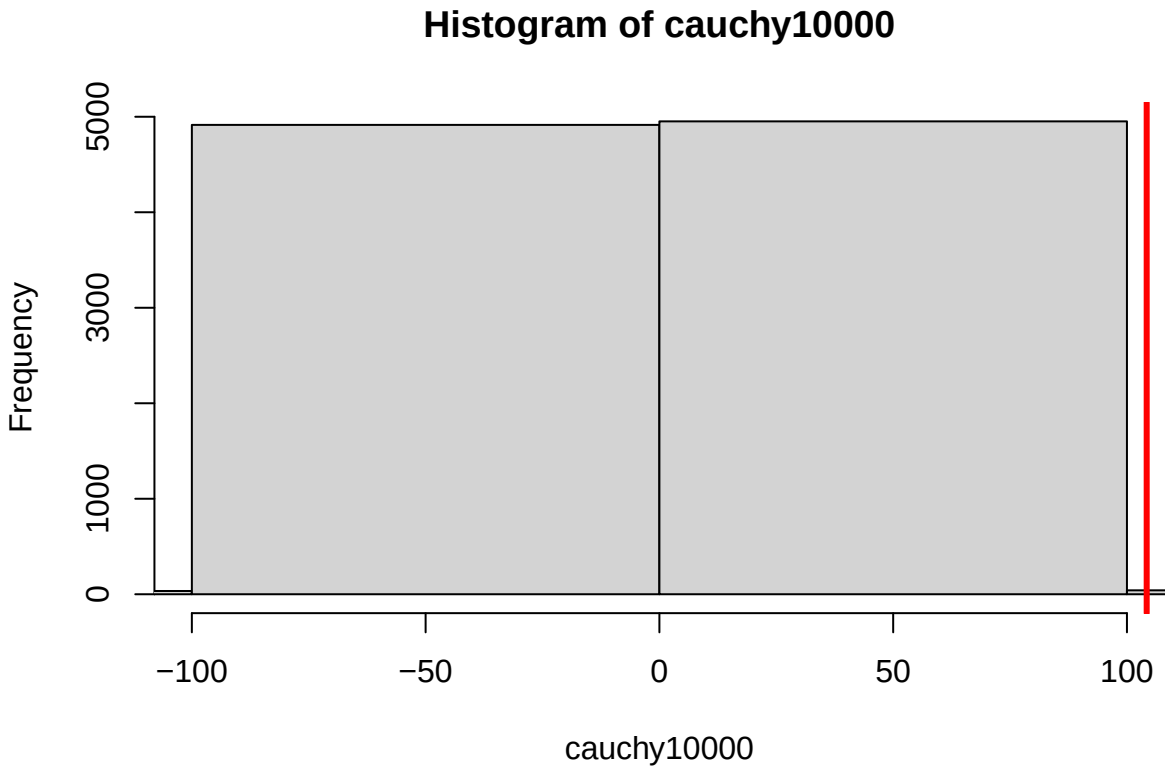

```
hist(cauchy100,breaks=100)
abline(v=mean(cauchy100),col="red",lwd=3)
```




```
hist(cauchy1000,xlim=c(-100,100),breaks=1000)
abline(v=mean(cauchy1000),col="red",lwd=3)
```




```
hist(cauchy10000,xlim=c(-100,100),breaks=10000)
abline(v=mean(cauchy10000),col="red",lwd=3)
```




D'après les graphiques nous pouvons remarquer que l'estimation de theta semble proche de la vrai valeur, n'ayant pas de moyen de calculer l'espérance de la loi cela semble être un bon estimateur, car celui-ci semble proche de 0, assez pour jugé les performances de cet estimateur comme suffisant.