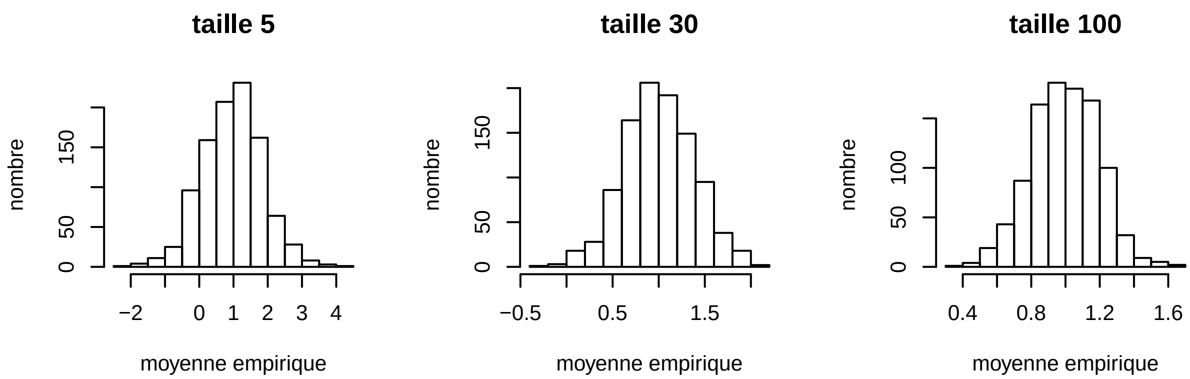
TPSTAT2

You ZUO

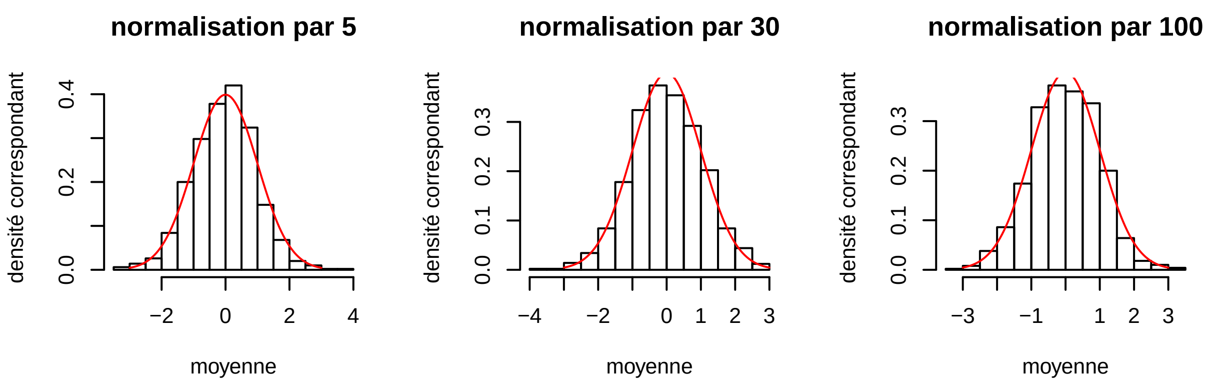
10/3/2019

1.Echantillon, Théorème Central Limite, Estimation Monte Carlo

<-1->1000 échantillons i.i.d gaussien



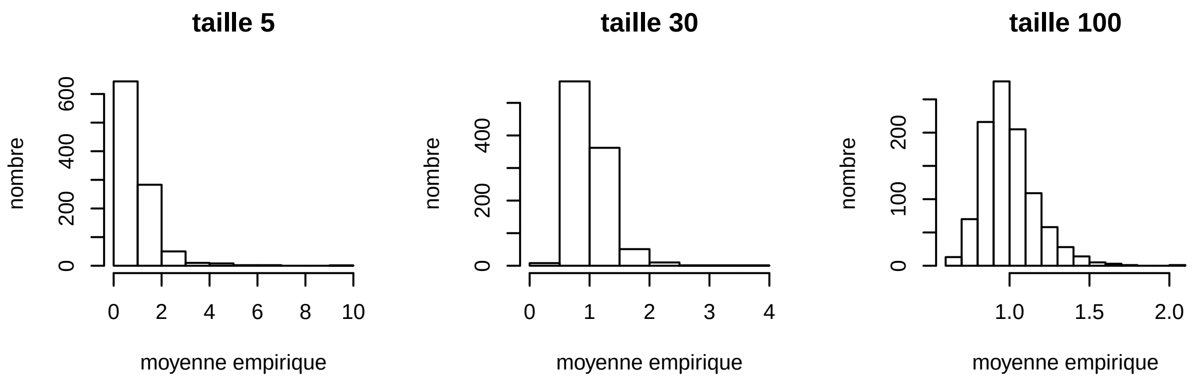
La moyenne empirique suit une loi théorique de $N(\mu, \frac{\sigma^2}{n})$.



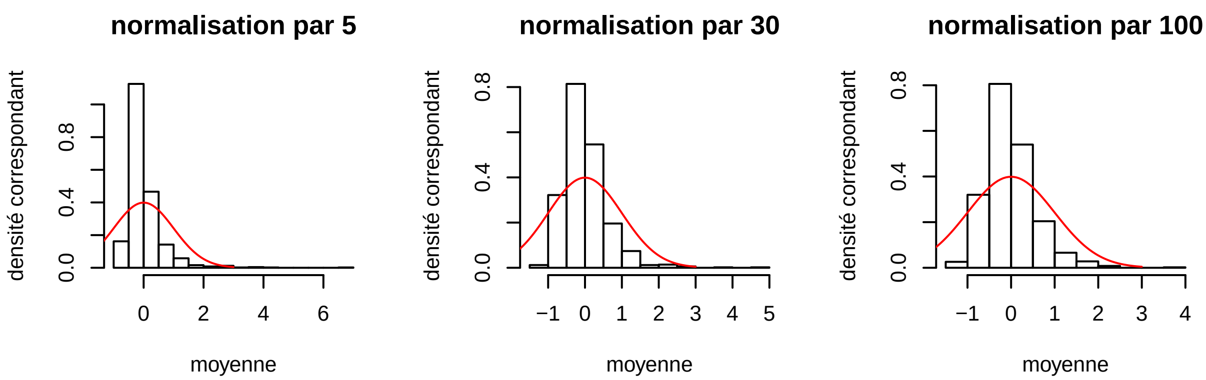
Selon les graphes au-dessus, on peut voir que plus la taille de l'échantillon est grande, plus l'intervalle des $U_{n,i}$ devient plus petit, et La distribution de ces valeurs se rapproche de plus en plus de 0.

<-2->1000 échantillons i.i.d Pareto $P(a, \alpha)$

Soient $m = 1,\ s = 3$:

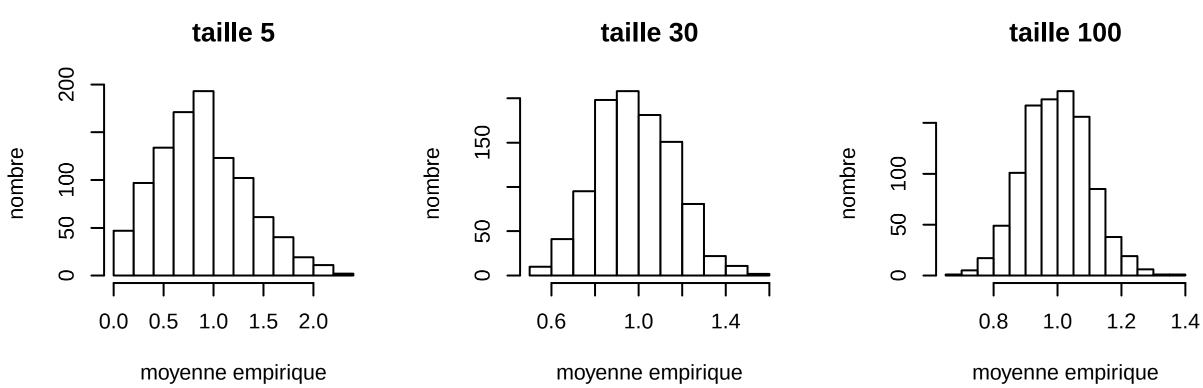


La moyenne empirique suit une loi théorique de $N(\frac{\alpha a}{\alpha-1}, (\frac{\alpha a}{\alpha-1})^2 \frac{\alpha}{n(\alpha-2)})$.

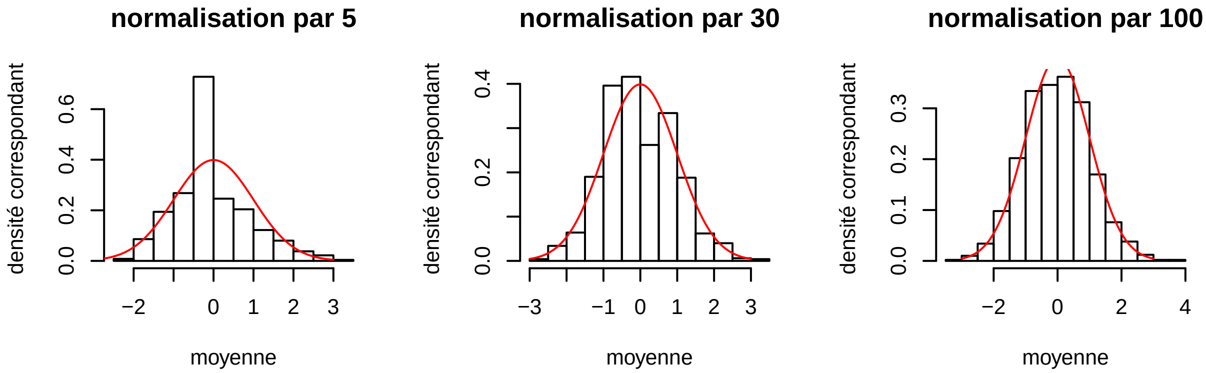


Selon les graphes au-dessus, on peut voir que plus la taille de l'échantillon est grande, plus l'intervalle des $U_{n,i}$ devient plus petit, et La distribution de ces valeurs se rapproche de plus en plus de 0.

<-3->1000 échantillons i.i.d de loi de Poisson



La moyenne empirique suit une loi théorique de $N(\lambda, \frac{\lambda}{n})$.



Selon les graphes au-dessus, on peut voir que plus la taille de l'échantillon est grande, plus l'intervalle des $U_{n,i}$ devient plus petit, et La distribution de ces valeurs se rapproche de plus en plus de 0.

<-4->La conclusion

Lorsque on veut estimer un paramètre d'une distribution, on a généralement besoin de construire une statistique T, puis on estimera le paramètre en calculant la distribution d'échantillonnage de la statistique, c'est-à-dire $E[T(X_1, ..., X_n)]$ et $V[T(X_1, ..., X_n)]$. Quant à comment n influence la qualité de cette approximation, on peut déduire selon des expérimentations précédentes que:

Plus la taille de l'échantillon n est grande, plus l'intervalle des resultats de calculs de $T(X_1, ..., X_n)$ devient plus petit, et La distribution de ces valeurs se rapproche de plus en plus de $E[T(X_1, ..., X_n)]$, et en même temps la vraie valeur du paramètre que l'on cherche.

Ce résultat est également une bonne preuve du théorème de la limite centrale.

2.Moyenne et dispersion

<-1->Inégalité de Bienaymé-Tchebychev

Pour tout réel strictement positif α,

$$P(|X - E[X]| \geq \delta) \leq \frac{\sigma^2}{\delta^2}$$

<-2->Estimer par Monte Carlo les probabilités de déviation d'une variable aléatoire de sa moyenne.

2.a)

$$P(|X - \mu| \geq \delta) = P(1_{\{|X-\mu|\geq\delta\}} = 1)$$

soit $Z = 1_{\{|X-\mu|\geq\delta\}}$, alors la valeur de Z est soit 1 soit 0, ainsi on peut calculer sa valeur d'espérance:

$$E[Z] = 0 \times P(Z = 0) \ + 1 \times P(Z = 1) = P(Z = 1)$$

Donc, par les formules au-dessus, on a déduit que:

$$P(|X - \mu| \geq \delta) = E[1_{\{|X-\mu|\geq\delta\}}] = E[Z]$$

2.b)

Soient $\mu = 4,\ \sigma = 2,\ \delta = 1$

```
## Gaussien   Pareto  Poisson
##   0.6161   0.8762   0.8018
```

2.c)

On va calculer la fréquence que $P(|X - E[X]| \geq \delta) \leq \frac{\sigma^2}{\delta^2}$ pour plusieurs δ en Faisant varier σ de 1.5 à 4 par pas de 0.5 .

```
##          delta=1   delta=2   delta=3
## Gaussien       1 1.0000000 1.0000000
## Pareto         1 0.8333333 0.6666667
## Poisson        1 1.0000000 1.0000000
```

La conclusion: en modifiant les valeurs de δ et σ, on a trouvé que les échantillons obéissant à la distribution gaussienne et à laquelle de Poisson peuvent très bien satisfaire l'inégalité de Bienaymé Chebyshev, mais lorsque la valeur de δ devient plus grande, l'échantillon de la distribution de Pareto des fois ne peut pas satisfaire cette inégalité.

2.d)

Le cas Gaussien $X \sim N(\mu, \sigma^2)$, $P(X \geq t) \leq exp(-\frac{(t-\mu)^2}{2\sigma^2})$ et pour tout $t \geq 0$ donc

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

car ici $t = \mu + \delta$

Pour les cas Gaussien avec $\mu = 1$

On va varier d'abord le δ de 1 à 4 par pas de 0.5 en fixant que $\sigma = 2$.

```
## $Chernoff
## [1] 0.8824969 0.7548396 0.6065307 0.4578334 0.3246525 0.2162652 0.1353353
##
## $`Monte Carlo`
## [1] 0.1553 0.1613 0.1536 0.1499 0.1541 0.1586 0.1556
##
## $satisfait
## [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
```

Puis on va varier la σ de 1 à 4 par pas de 0.5 en fixant que $\delta = 2$

```
## $Chernoff
## [1] 0.1353353 0.4111123 0.6065307 0.7261490 0.8007374 0.8493658 0.8824969
##
## $`Monte Carlo`
## [1] 0.1617 0.1581 0.1572 0.1650 0.1521 0.1563 0.1590
##
## $satisfait
## [1] FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
```

Le cas Poisson $X \sim P(\lambda)$, et donc pour tout $t \geq 0$

$$P(X \geq t) \leq exp(-uh(\frac{t}{\mu}))$$

avec $h(x) = (1+x)log(1+x) - x, x \geq -1$

On va varier d'abord le δ de 1 à 4 par pas de 0.5 en fixant que $\lambda = 2$.

```
## $Chernoff
## [1] 0.24525296 0.26199662 0.25000000 0.21711556 0.17397004 0.12995030
## [7] 0.09122291
##
## $`Monte Carlo`
## [1] 0.1447 0.1425 0.1387 0.1432 0.1434 0.1381 0.1486
##
## $satisfait
## [1]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
```

Puis on va varier la $\sqrt{\lambda}$ de 0.5 à 2 par pas de 0.5 en fixant que $\delta = 2$

```
## $Chernoff
## [1] 0.07104448 0.20286897 0.30203131 0.29794366 0.21827980 0.12646904
## [7] 0.06014901
##
## $`Monte Carlo`
## [1] 0.2229 0.1088 0.2668 0.2094 0.1921 0.1940 0.2128
##
## $satisfait
## [1] FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
```

<-3->

3.a)

Remarque: $P(\bar{X}_n - \mu \geq \delta) \leq exp(-\frac{n(\mu+\delta)^2}{2\sigma^2})$

On va calculer les bornes de Chernoff dans le cas échantillon pour $\bar{X}_n$ en variant $n = 20, 100, 1000$.

```
##                   20          100           1000
## Gauss   3.606563e-03 6.101937e-13 7.156109e-123
## Poisson 2.681004e-14 1.385119e-68  0.000000e+00
```

3.b)

Les résultats ci-dessus montrent que, à mesure que n grandit, la valeur de borne de Chernoff se rapproche de 0. C'est-à-dire que la probabilité que la différence entre $\bar{X}_n$ et μ ($resp.\lambda$) soit supérieure à une certaine distance devient de plus en plus petite, presque nulle. Par conséquent, on peut penser que la valeur de μ ($resp.\lambda$) converge vers la moyenne empirique en fonction de la probabilité. $\bar{X}_n$ est donc un estimateur de μ ($resp.\lambda$).

<-4->

4.a)

```
##              20       100     1000   10000
## [1,] -0.2498296 0.4845105 4.026143 23.3485
```

Selon les résultats on a trouvé que la moyenne empirique ne converge pas par rapport à n.

4.b)

Selon la formule qui utilise la fonction propre pour trouver l'espérance:

$$E(X^n) = i^{-n}\phi_X^{(n)}(0) = i^{-n}[\frac{d^n}{dt^n}\phi_X(t)]_{t=0}$$

Ici, on étudiera uniquement la distribution standard de Cauchy, c'est-à-dire que $\theta = 0$. Alors $\phi_0(t) = exp(-|t|)$ il n'est pas dérivable en $t = 0$, indiquant que son espérance n'exisite pas.

4.c)

```
##                   20       100         1000
## médiane -0.4993789 0.0895584 -0.04159588
```