TP 2 : un corrigé

Yiyang Yu

mars 2020

1. Mesurer la vitesse de la lumière

```r
c <- 3
sigma = sqrt(0.5)
```

$X_i = c + \xi_i, \quad \xi_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d

(1) Simuler 100 observations et les placer dans un vecteur **obs**.

```r
n <- 100
obs <- c + rnorm(n, mean=0, sd=sigma)
print(obs)
```

```
##   [1] 2.978935 2.882455 3.727018 2.229418 1.764426 2.087498 2.660183 2.970332
##   [9] 2.918710 3.513141 3.223935 2.702527 4.800909 3.230920 3.724212 3.198686
##  [17] 3.501737 2.710633 2.735516 3.274375 2.914122 4.006246 3.138854 2.473109
##  [25] 1.767533 1.998984 3.242994 1.817270 3.076729 2.717793 2.853122 2.781851
##  [33] 2.130281 3.636545 2.468330 4.642256 2.781451 3.445516 3.882892 2.233166
##  [41] 3.370820 3.860817 3.358915 2.747811 2.836867 2.997201 2.552775 2.856541
##  [49] 2.314063 2.921758 2.900724 2.437535 2.824297 2.797940 3.314178 1.947379
##  [57] 4.171924 2.603102 3.487086 1.838088 2.247504 2.813136 2.797047 2.432011
##  [65] 3.348541 2.970422 3.649608 2.957413 1.124329 1.612577 4.721920 4.539735
##  [73] 2.644826 3.655252 4.165214 3.283748 2.751741 2.245258 3.003320 2.828640
##  [81] 3.362272 3.966040 3.758284 3.849092 3.522047 3.077208 3.068751 2.128441
##  [89] 2.759951 4.264201 3.765539 2.584578 3.021000 1.952114 4.165976 2.280984
##  [97] 3.742355 3.031290 2.123939 3.157681
```

(2) On estime c par la moyenne arithmétique des observations. On simule 100 observations et on calcule l'estimateur $\bar{X}_{100}$ associé. On répète cette opération 50 fois (calculer la moyenne empirique de 100 obervations), on conserve à chaque fois l'estimateur $\bar{X}_{100}$.

En effet, $\mathbb{E}[X_i] = c$, $\mathbb{E}[\bar{X}_n] \to c$ pour $n \to +\infty$ en proba (LGN); on utilise $\bar{X}_n$ pour estimer c avec $n = 100$.

```r
list_moyenne_empirique = c(1:50)
for (i in 1:50){
  obs <- c + rnorm(n, 0, sigma)
  moyenne_empirique <- mean(obs)
  list_moyenne_empirique[i] = moyenne_empirique
}
print(list_moyenne_empirique)
```

```
##  [1] 2.990144 2.942685 3.079497 2.930824 3.153587 2.997914 2.969673 3.121666
##  [9] 2.933795 3.096631 2.955550 3.045601 2.901923 3.050724 3.169369 3.002896
## [17] 3.061016 2.894463 2.958060 3.040694 2.928042 2.988392 3.018561 2.940746
## [25] 2.986157 2.934005 2.964514 3.071720 2.890320 3.065098 3.003429 3.037144
```

```
## [33] 3.036247 2.970120 3.122306 2.981712 3.030085 3.006590 3.068876 3.105456
## [41] 3.000479 2.929600 3.048010 2.960859 2.999672 3.030336 2.979259 2.946081
## [49] 3.010733 2.956634
```

Quelle est la variance empirique de ces 50 estimateurs ?

C'est variance des 50 exemples de $\bar{X}_{100}$

```
variance_empirique_list = var(list_moyenne_empirique)
print(variance_empirique_list)
```

```
## [1] 0.004407816
```

Répéter la procédure avec 1000 observations, que dire de la nouvelle variance empirique ? C'est variance des 50 exemples de $\bar{X}_{1000}$

```
m <- 1000
list_me = c(1:50)
for (i in 1:50){
  obs <- c + rnorm(m, 0, sigma)
  moyenne_empirique <- mean(obs)
  list_me[i] = moyenne_empirique
}
variance_empirique_list_me = var(list_me)
print(variance_empirique_list_me)
```

```
## [1] 0.0004655713
```

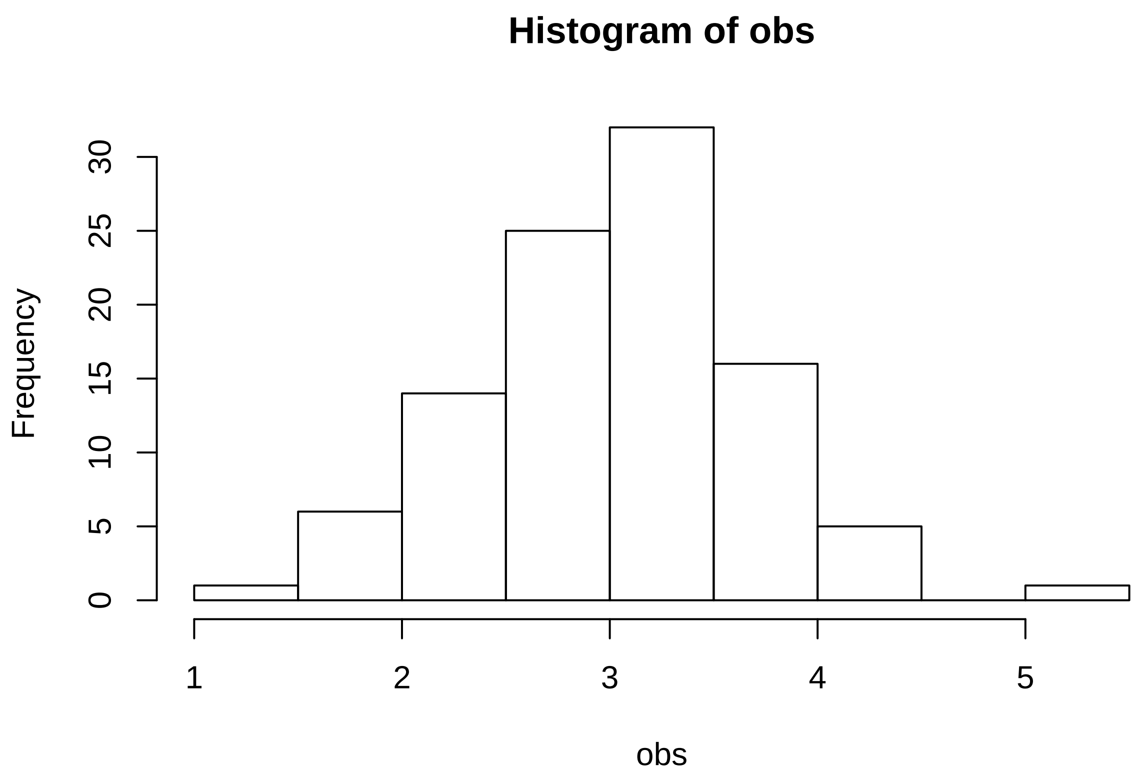
Quel principe est illustré ici ?

TCL: $\frac{\sqrt{n}(\bar{X}_n - c)}{\sigma} \to \mathcal{N}(0,1)$

donc $\bar{X}_n \to \mathcal{N}(c, \frac{\sigma^2}{n})$

(3) A nouveau on simule 100 observations. Tracer un histogramme des observations.

```
n <- 100
sigma = sqrt(0.5)
obs <- c + rnorm(n, 0, sigma)
hist(obs)
```



(4) Calculer la variance empirique des observations, ainsi que l'écart-type.

```
moyenne_empirique_obs = mean(obs)
variance_empirique_obs = var(obs)
ecart_type_obs = sqrt(variance_empirique_obs)
paste('Moyenne empirique:', moyenne_empirique_obs)
```

```
## [1] "Moyenne empirique: 3.04204648915584"
```

```
paste('Varience empirique:', variance_empirique_obs)
```

```
## [1] "Varience empirique: 0.419739420532375"
```

```
paste('écart-type empirique:', ecart_type_obs)
```

```
## [1] "écart-type empirique: 0.647872997224282"
```

Quelle proportion des observations se situe à plus d'un éart-type de la moyenne ?

```
print((sum(obs < moyenne_empirique_obs - ecart_type_obs) +
         sum(obs > moyenne_empirique_obs + ecart_type_obs))/n)
```

```
## [1] 0.28
```

A deux écart-types ?

```
print((sum(obs < moyenne_empirique_obs - 2*ecart_type_obs) +
         sum(obs > moyenne_empirique_obs + 2*ecart_type_obs))/n)
```

```
## [1] 0.06
```

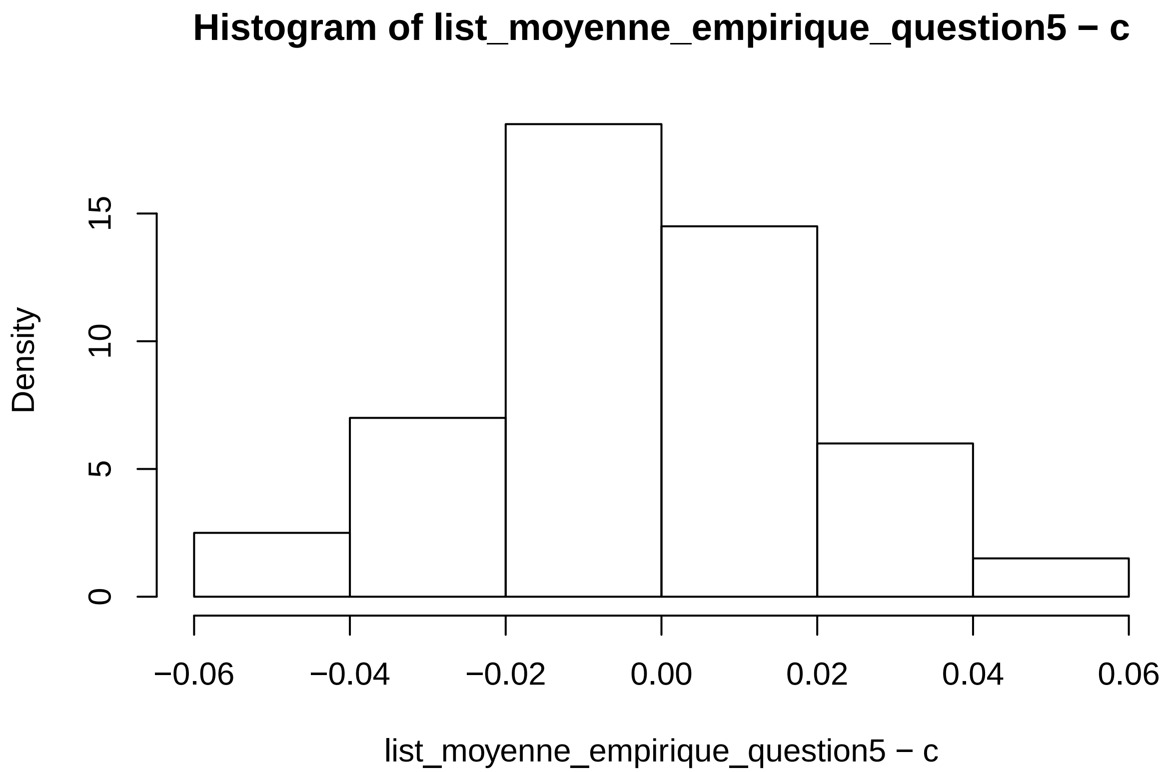
Rappelons la règle 68-95-99,7 (ou règle des trois sigmas)…

(5) Comme dans la question 2 générer un vecteur de 100 estimateur de c basés sur 1000 observations. On s'intéresse à loi de l'écart entre ces estimateurs et c. Comment faut-il renormaliser la différence

$\bar{X}_{1000} - c$? Tracer un histogramme de la distribution des estimateurs renormalisés, qu'observe-t'on ?

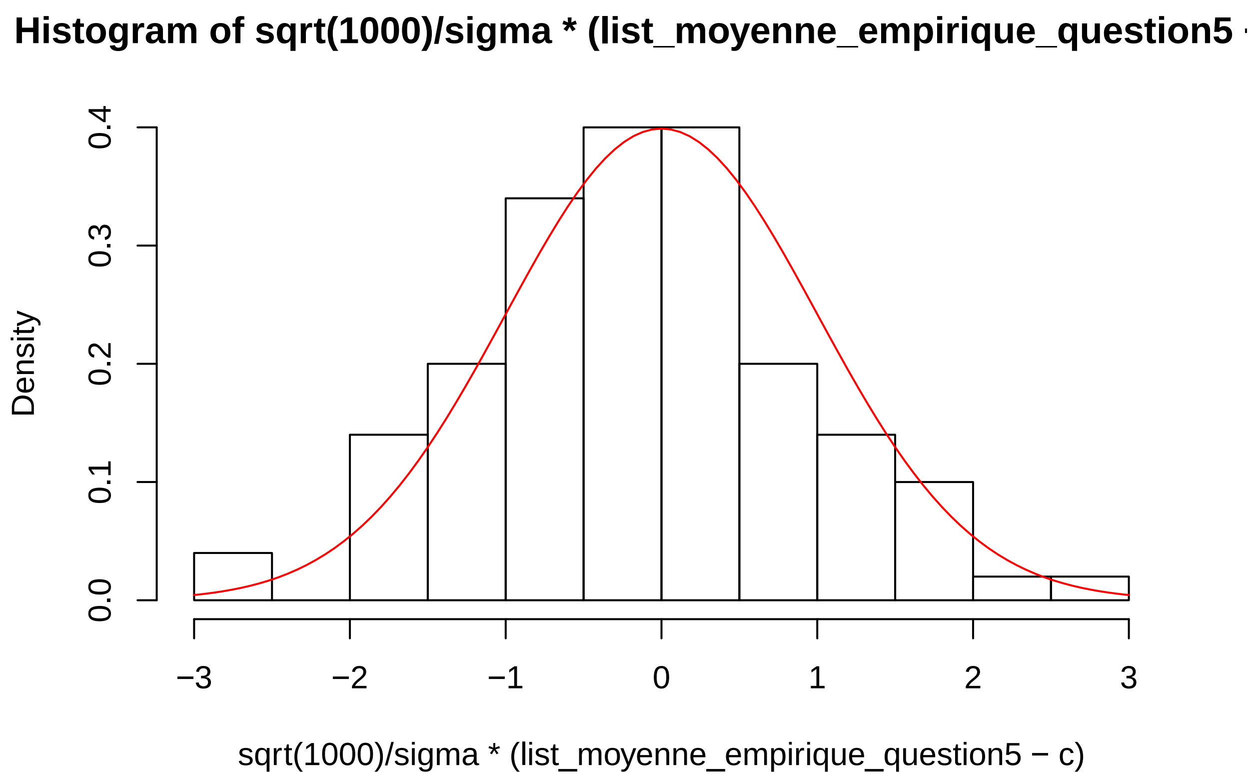
```
list_moyenne_empirique_question5 = c(1:100)
for (i in 1:100){
  obs <- c + rnorm(m, 0, sigma)
  moyenne_empirique <- mean(obs)
  list_moyenne_empirique_question5[i] = moyenne_empirique
}
```

```
hist(list_moyenne_empirique_question5 - c, freq=F)
```



On renormalise avec $\frac{\sqrt{n}}{\sigma}$, comme le suggère TCL. (En comparaison, on dessine la densité d'une loi normale centrée réduite en rouge.)

```
hist(sqrt(1000)/sigma*(list_moyenne_empirique_question5 - c), freq=F)
curve(dnorm(x,mean=0, sd=1), from=-3, to=3, add=T, col="red")
```

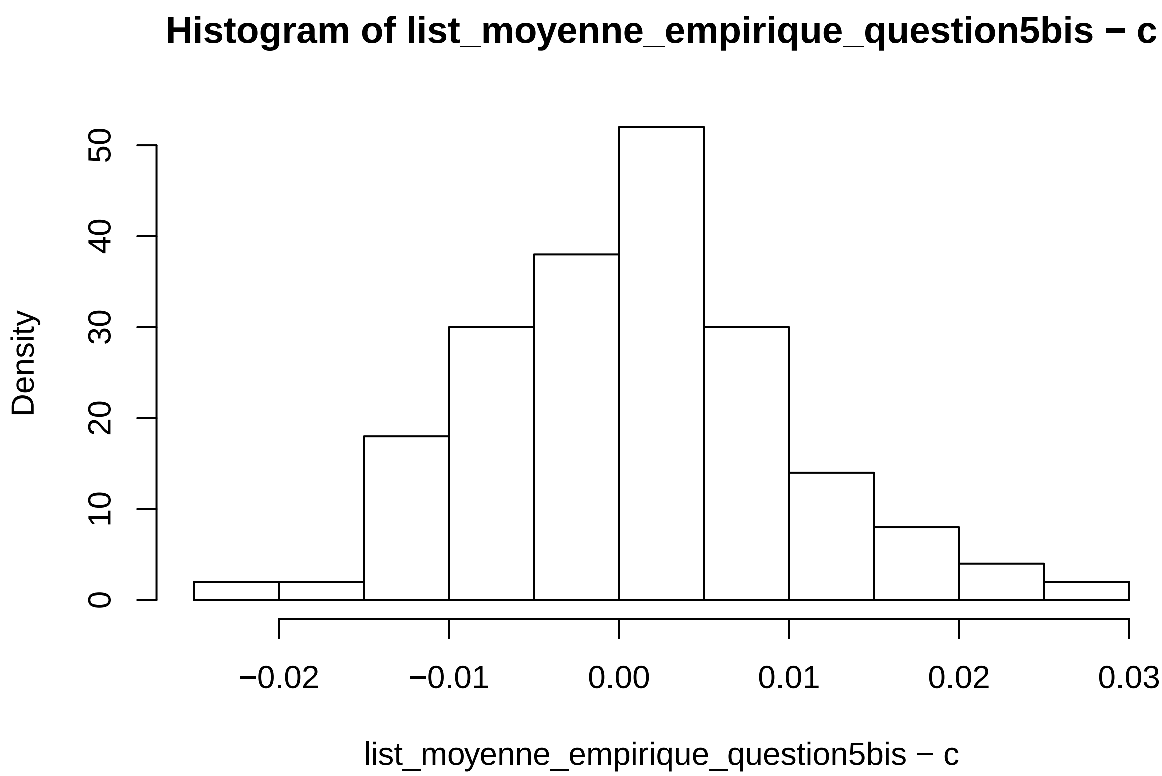


(5 bis) Reprendre cette question avec $\xi_i \sim \mathcal{U}([-0.5, 0.5])$.

```
list_moyenne_empirique_question5bis = c(1:100)
for (i in 1:100){
  obs <- c + runif(m, min=-0.5, max=0.5)
  moyenne_empirique <- mean(obs)
  list_moyenne_empirique_question5bis[i] = moyenne_empirique
}
print(var(obs))
```

```
## [1] 0.08122302
```

```
hist(list_moyenne_empirique_question5bis - c, freq=F)
```



Comme tout à l'heure, on renormalise avec $\frac{\sqrt{n}}{\sigma}$. (En comparaison, on dessine la densité d'une loi normale centrée réduite en rouge.)

```
hist(sqrt(1000)/sqrt(0.25/3)*(list_moyenne_empirique_question5bis - c), freq=F)
curve(dnorm(x,mean=0, sd=1),from=-3,to=3,add=T,col="red")
```