

TP3

11/03/2022

R Markdown

Démarche:

Définition des variables

On simule $N=1000$ échantillons de taille $n \in \{5, 30, 100\}$ pour les 3 premières questions.

On choisit de renormaliser uniformément chaque loi par sa moyenne empirique et variance empirique pour obtenir des lois centrées réduites.

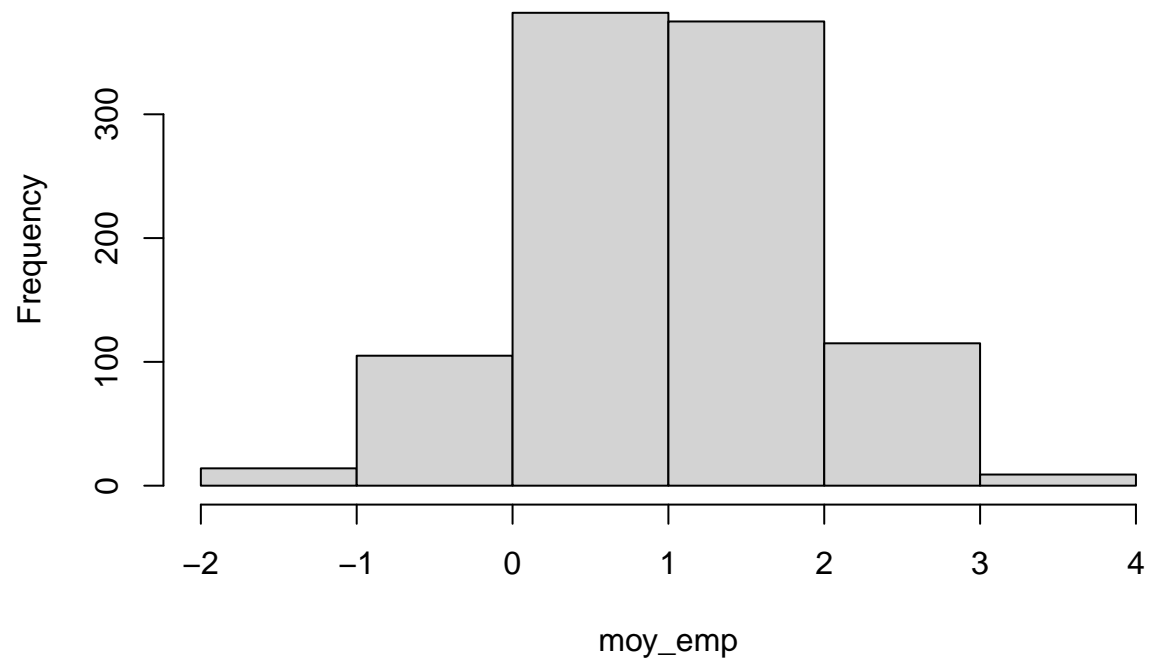
Moyenne empirique d'un échantillon

Variance empirique d'un échantillon

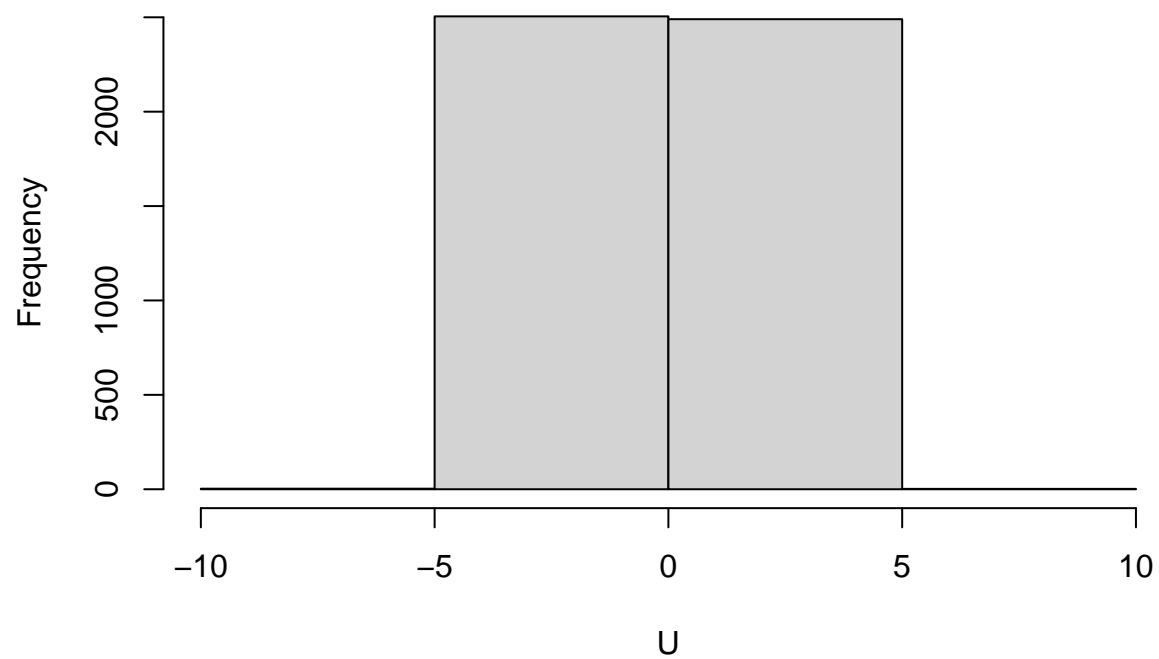
Question 1

Échantillons Gaussienne

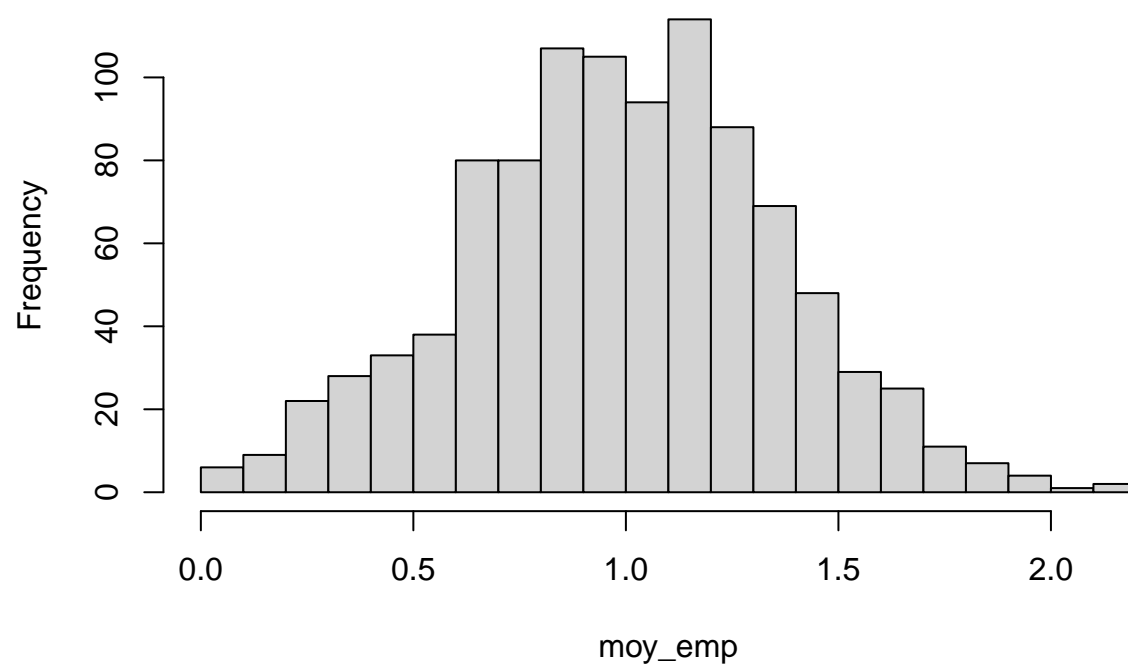
Moyenne empirique de gaussiennes pour $n = 5$



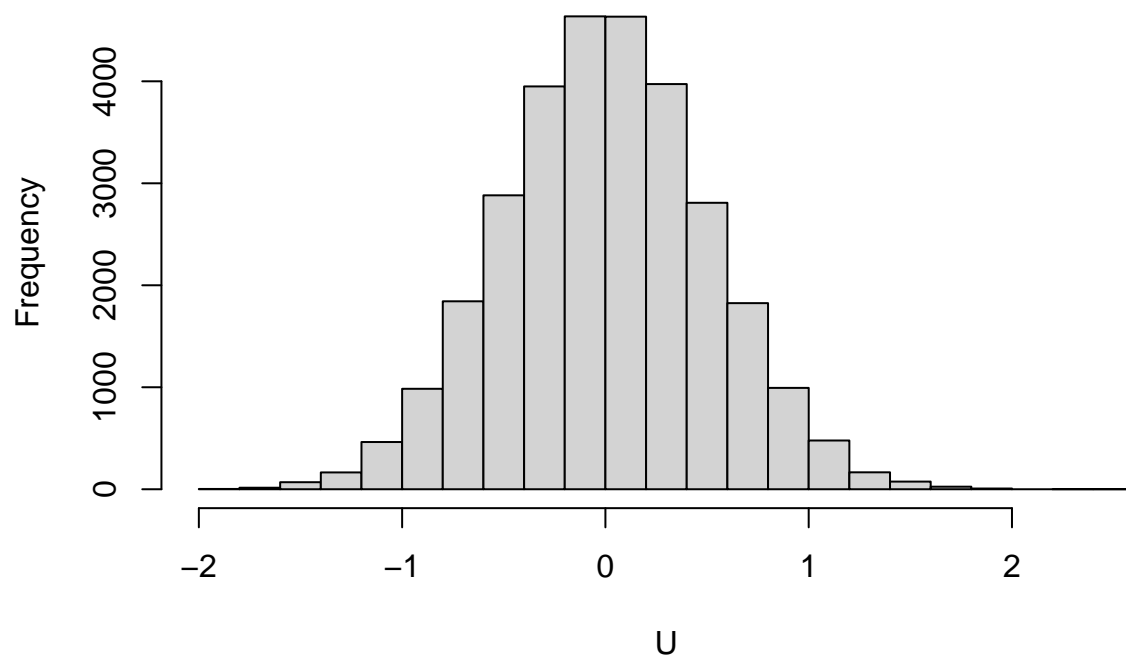
Loi empirique des $U_{n,i}$ pour $n = 5$



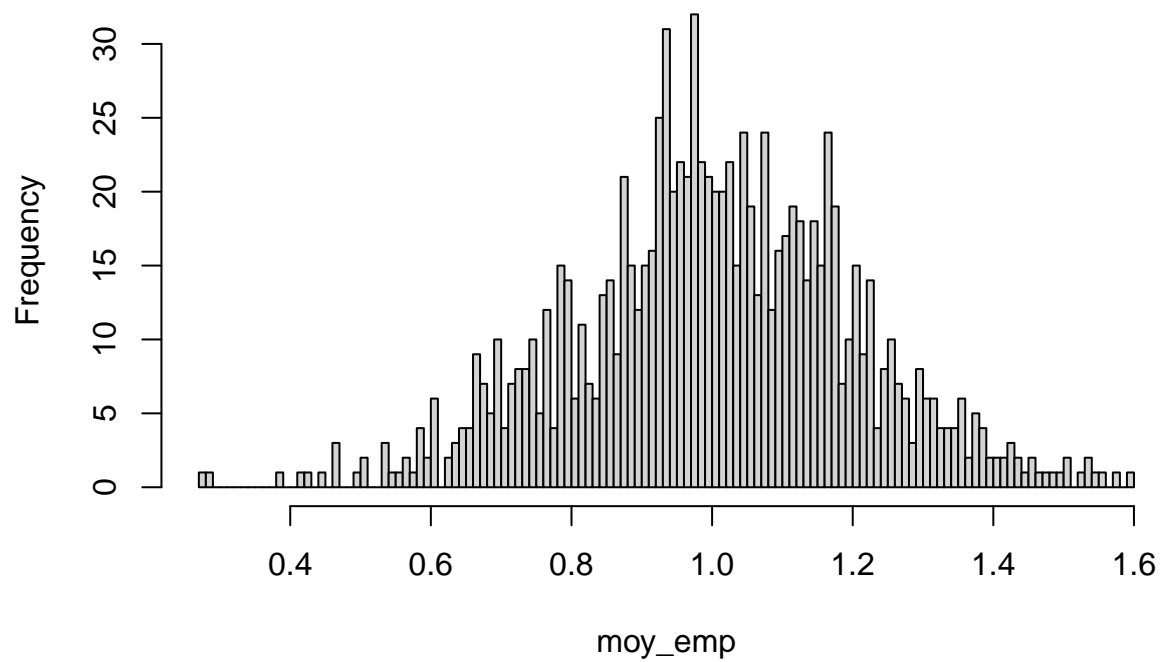
Moyenne empirique de gaussiennes pour n = 30



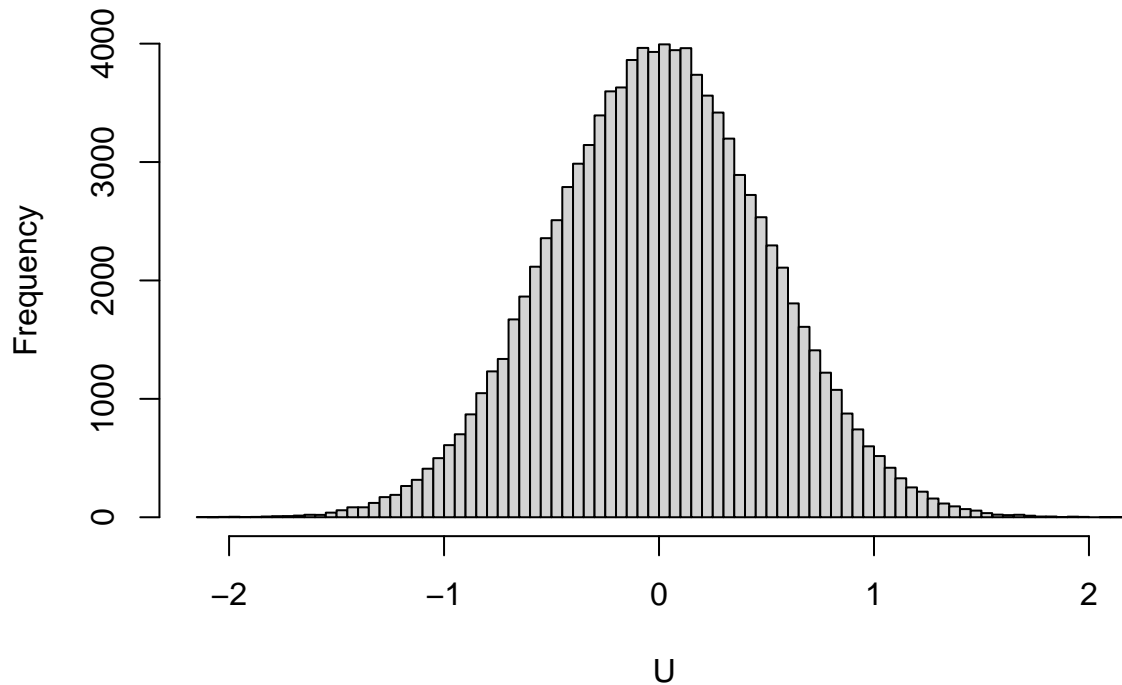
Loi empirique des $U_{n,i}$ pour $n = 30$



Moyenne empirique de gaussiennes pour $n = 100$



Loi empirique des $U_{n,i}$ pour $n = 100$



La moyenne empirique suit une loi gaussienne (avec quelques écarts).

On constate que les $U_{n,i}$ représentent la loi gaussienne à mesure que la taille augmente.

Question 2

On suppose

$$\alpha \geq 2$$

Démonstration de la formule de l'espérance d'une loi de Pareto

$$F_X(x) = \left(1 - \left(\frac{a}{x}\right)^\alpha\right), x \geq a$$

$$E(X) = \int_0^\infty 1 - F_X(t) dt = \int_0^\infty P(X > t) dt$$

donc

$$E(X) = a + \int_a^\infty \left(\frac{a}{t}\right)^\alpha dt = a + \frac{a}{\alpha - 1} = \frac{\alpha a}{\alpha - 1}$$

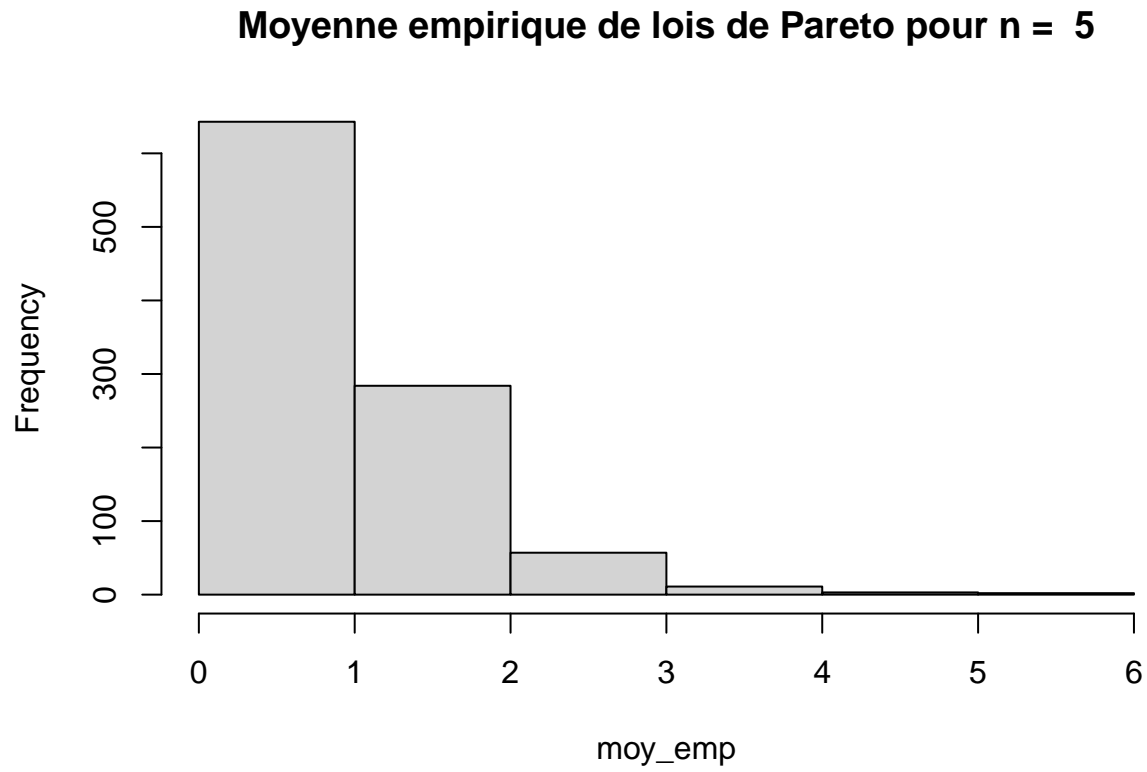
Ici: $\alpha=3$, $a=1$ $E(X) = 1.5$

[1] 1.012598

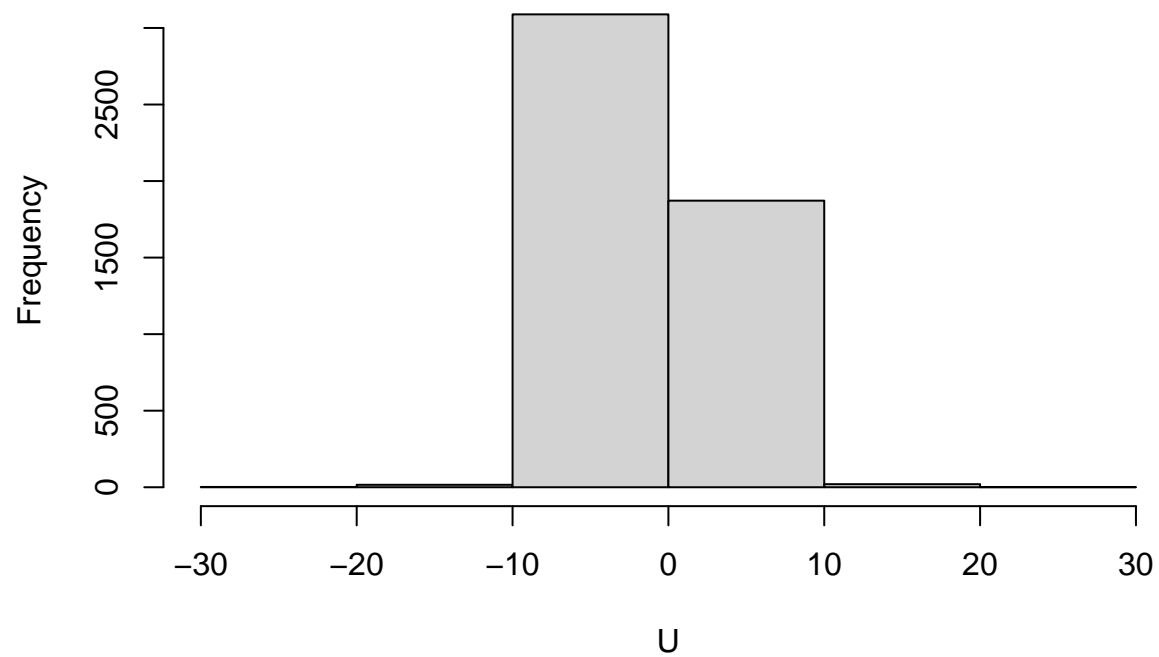
[1] 1.00195

```
## [1] 0.9942371
```

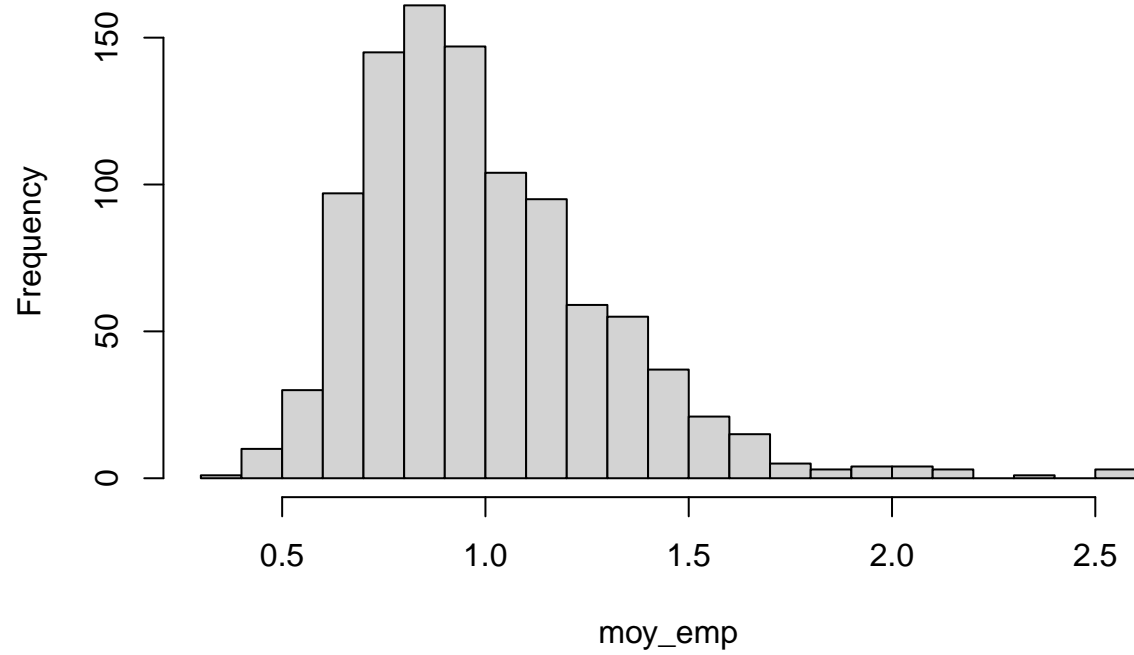
Échantillons Pareto



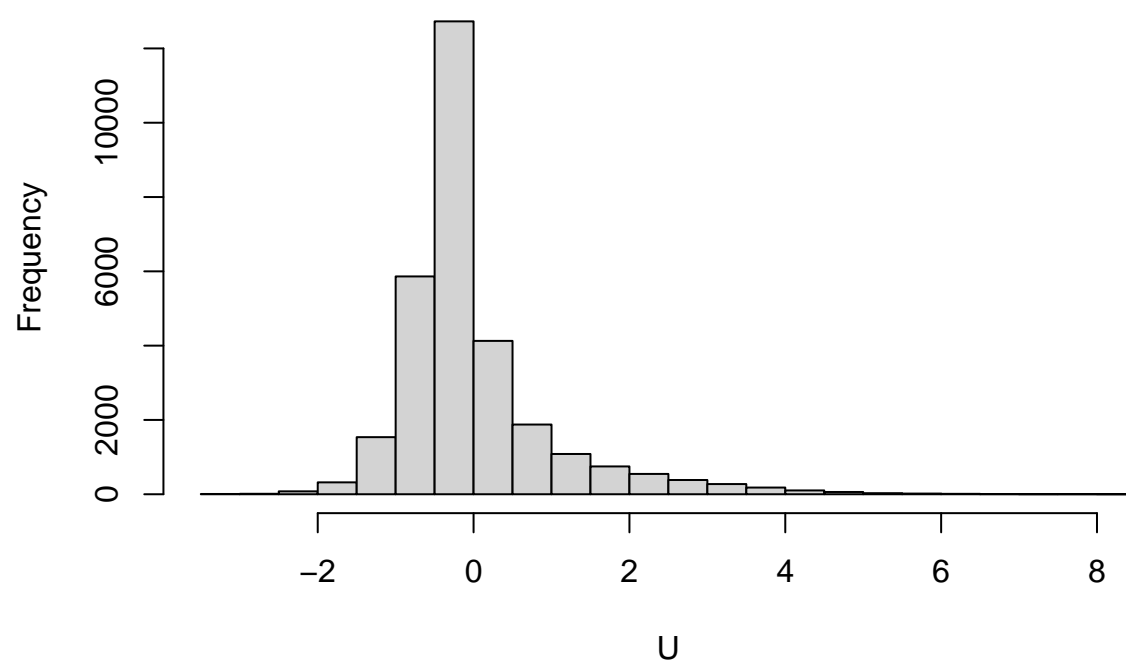
Loi empirique des $U_{n,i}$ pour $n = 5$



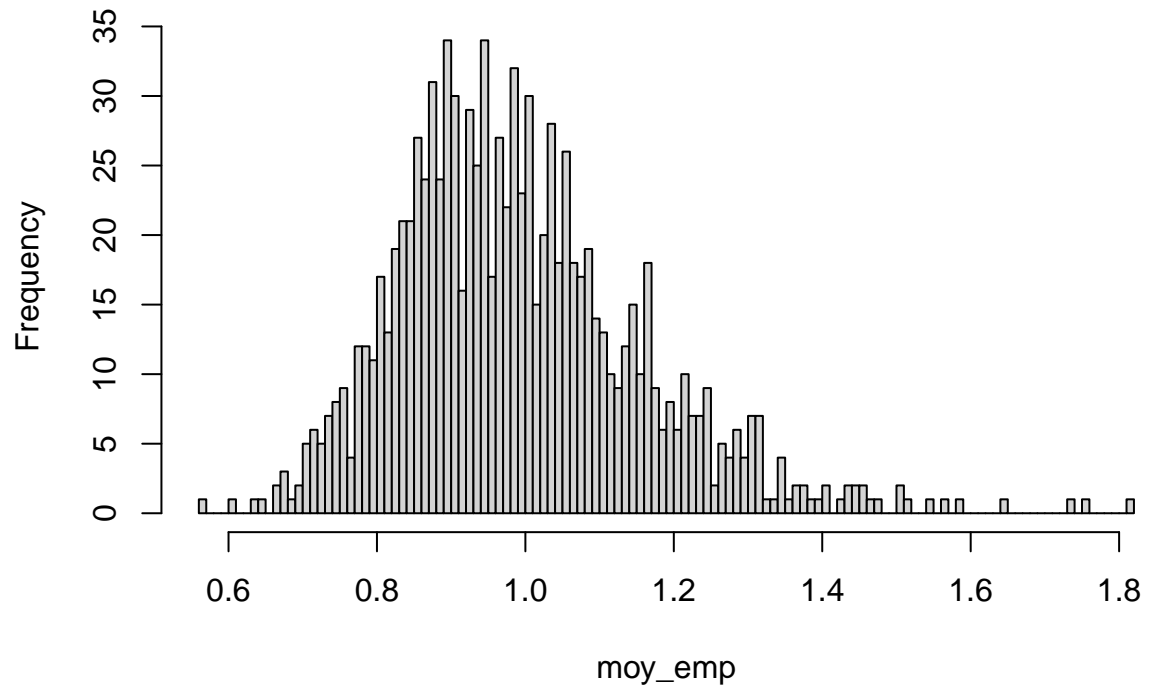
Moyenne empirique de lois de Pareto pour $n = 30$



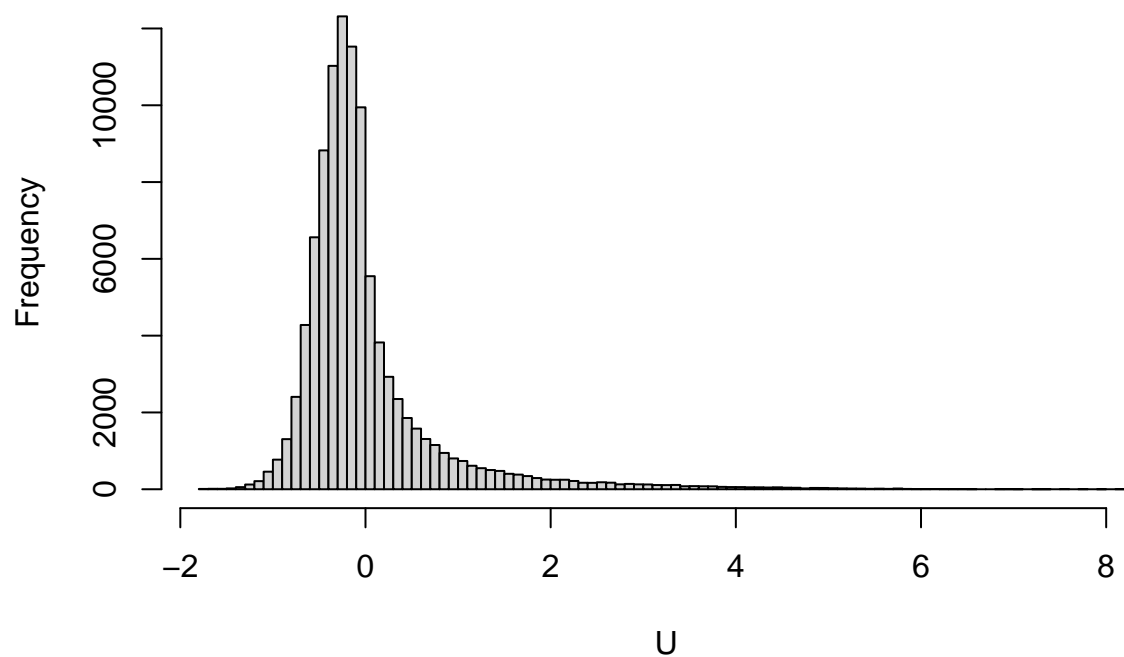
Loi empirique des $U_{n,i}$ pour $n = 30$



Moyenne empirique de lois de Pareto pour $n = 100$



Loi empirique des $U_{n,i}$ pour $n = 100$

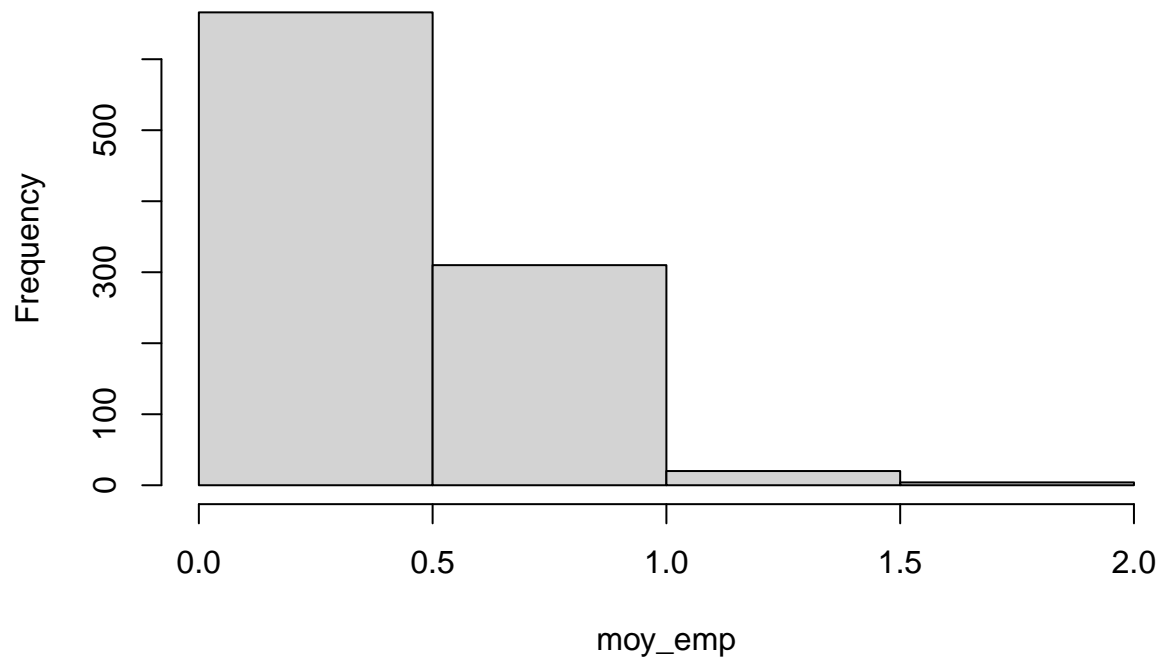


On constate que les $U_{n,i}$ représentent mieux la loi de pareto que les moyennes empiriques, à mesure que la taille augmente.

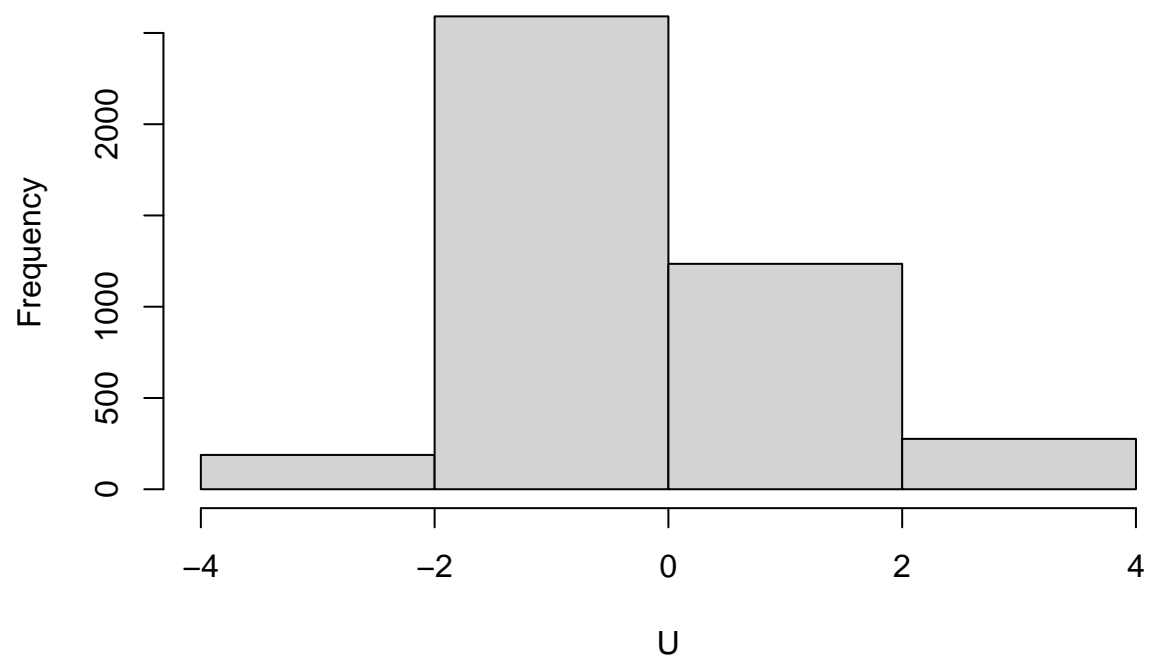
Échantillons Poisson

Question 3

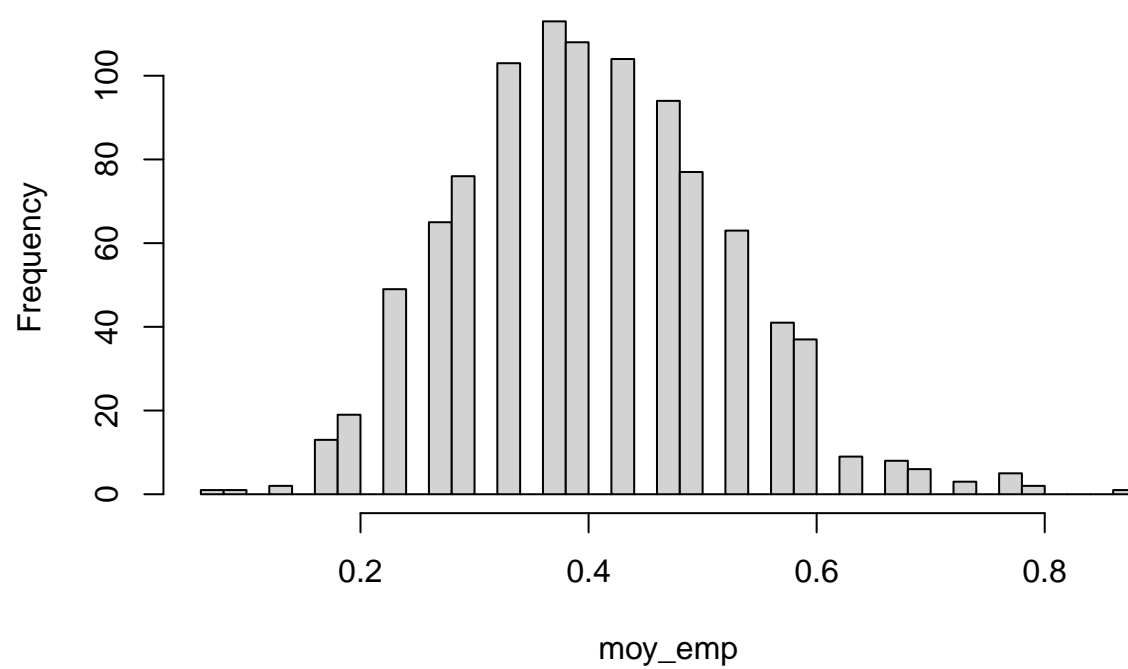
Moyenne empirique de lois de Poisson pour $n = 5$



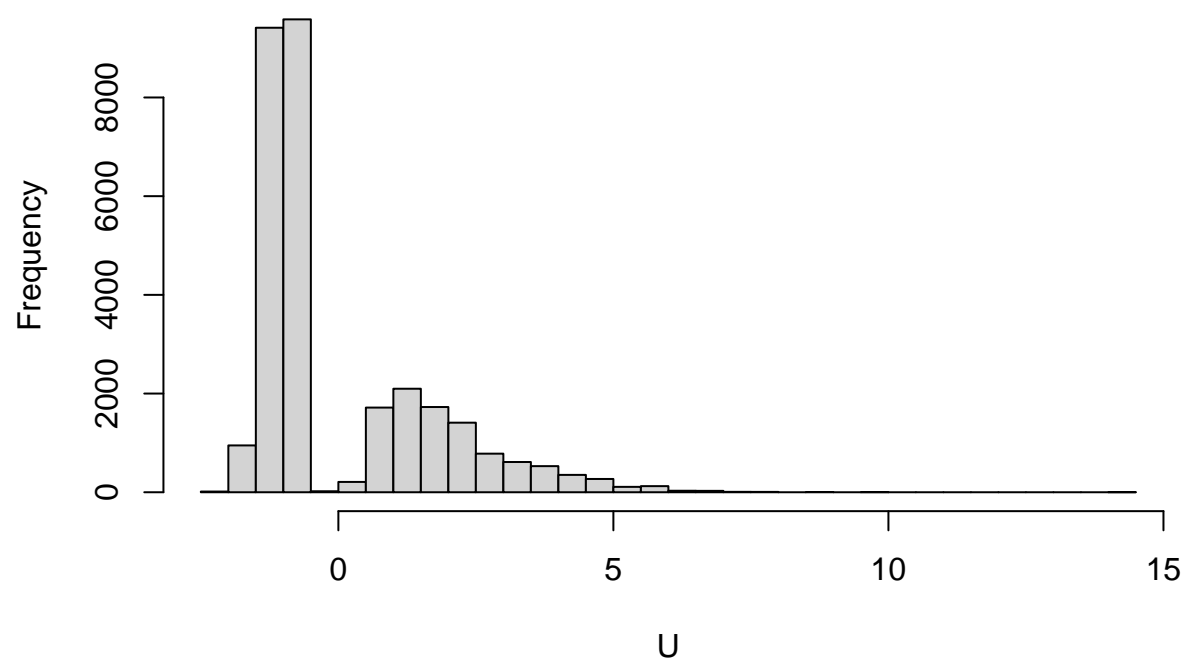
Loi empirique des $U_{n,i}$ pour $n = 5$



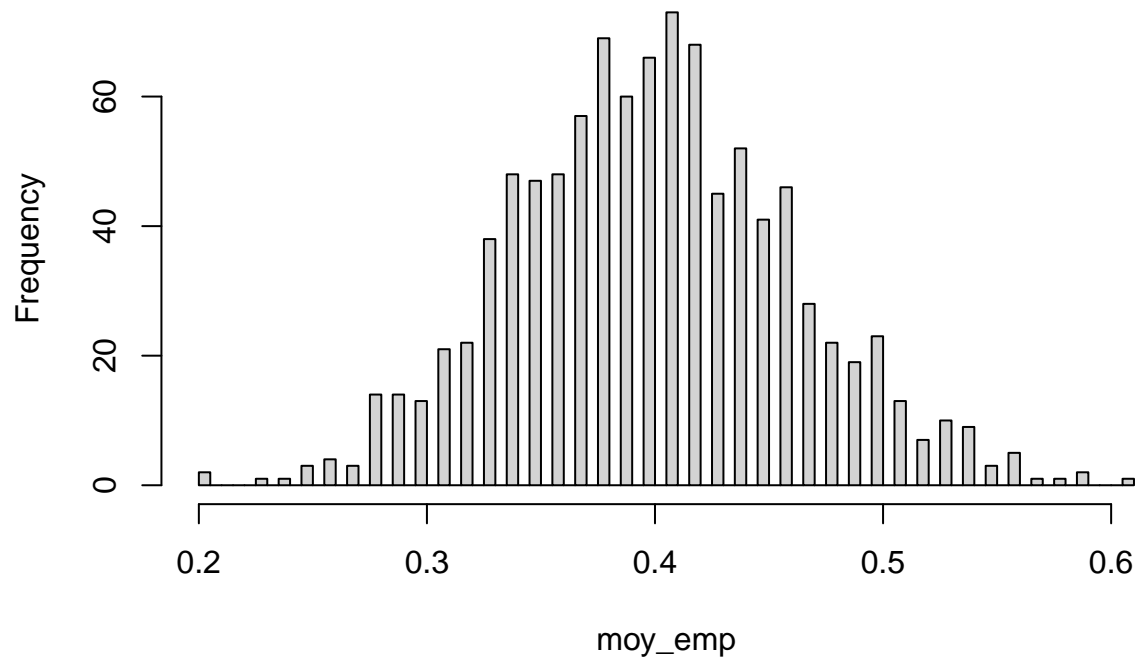
Moyenne empirique de lois de Poisson pour $n = 30$



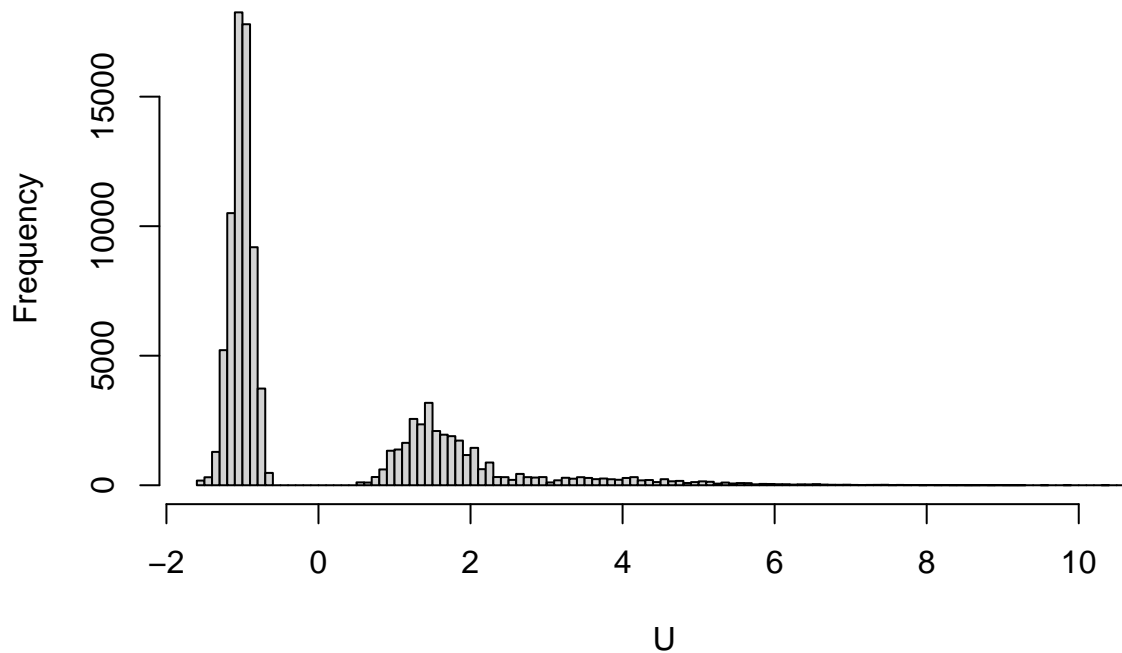
Loi empirique des $U_{n,i}$ pour $n = 30$



Moyenne empirique de lois de Poisson pour n = 100



Loi empirique des $U_{n,i}$ pour $n = 100$



Ici l'approximation par loi des moyennes empiriques occulte les événements rares (pics des $U_{n,i}$ entre -2 et 0) caractéristiques d'une loi de poisson pour obtenir en moyenne une gaussienne.

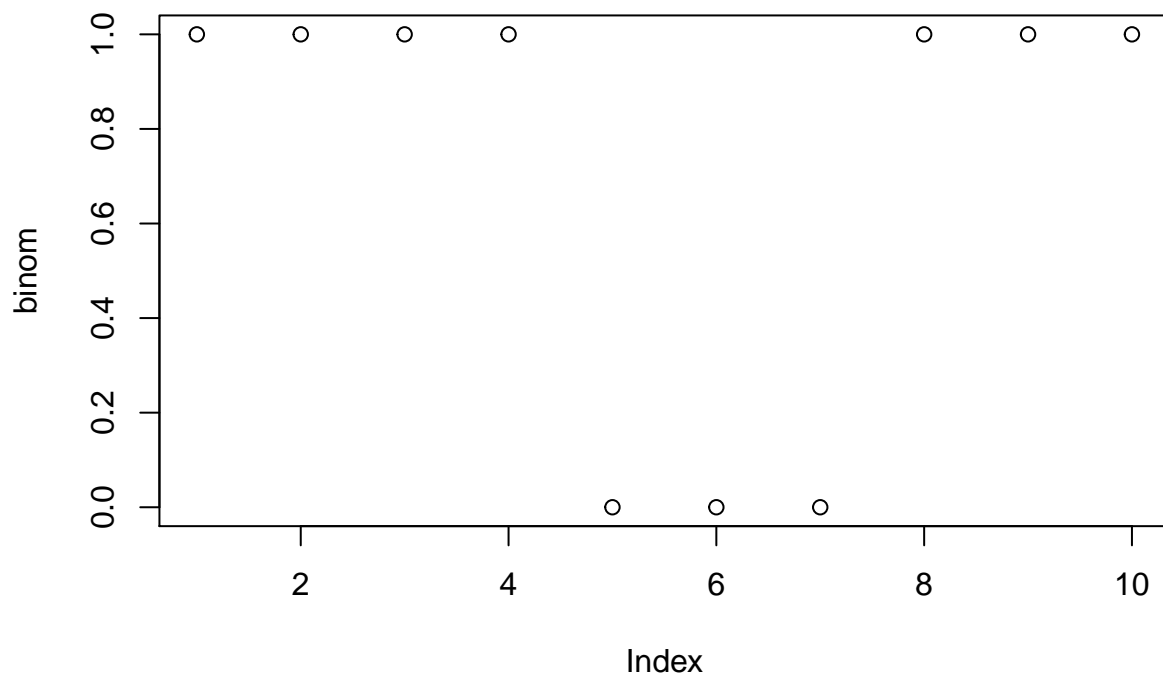
Question 4

On a vu d'après les 3 exemples précédent que la moyenne empirique n'est pas toujours un estimateur très fiable.

On doit donc calculer renormaliser uniformément chaque loi avec les $U_{n,i}$ pour obtenir un échantillon gaussien bien réparti et non biaisé.

Question 5

(Échantillon généré aléatoirement à chaque compilation)



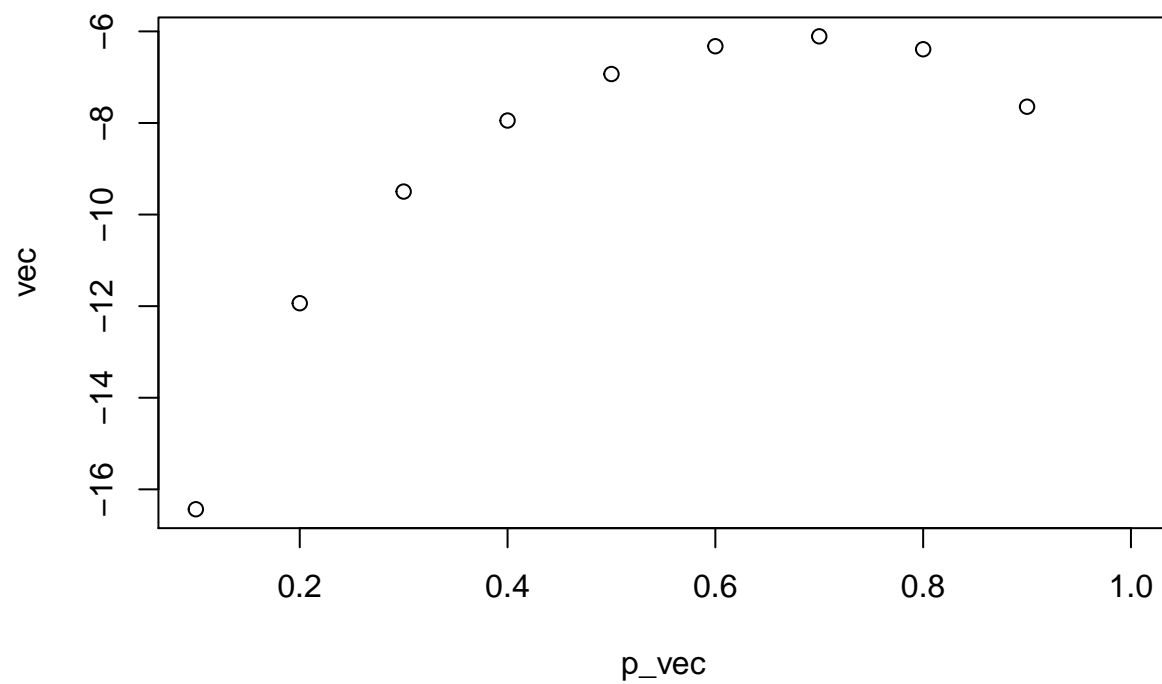
Pour estimer p il suffit d'estimer la moyenne de l'échantillon de bernoulli.

Question 6

$L_{\text{bern}} = p^s * (1 - p)^{n-s}$ où s =nombre de succès

```
L_bern <- function (p, binom) {
  n <- length(binom)
  sx <- sum(binom)
  return(log(p^sx*(1-p)^(n-sx)))
}
```

Question 7



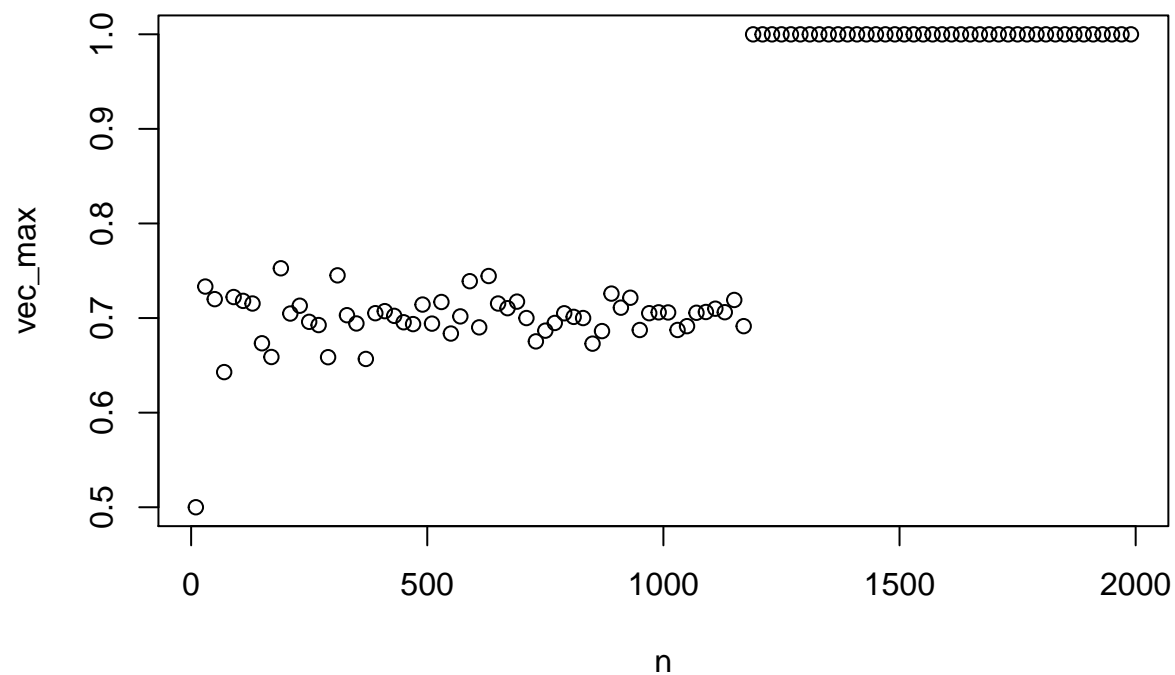
On remarque que la vraisemblance suit une courbe en cloche et que maximum de vraisemblance s'approche de $p = 0.7$ / est atteint entre 0.6 et 0.8 selon l'échantillon généré

Question 8

On s'approche du maximum de vraisemblance :

```
## [1] 0.7000058
```

Question 9



On constate que de $n=0$ à $n \approx 1300$ $|L_{max} - 0,7| - > 0$, puis partir de $n \approx 1300$, une discontinuité assigne L_{max} à 1.

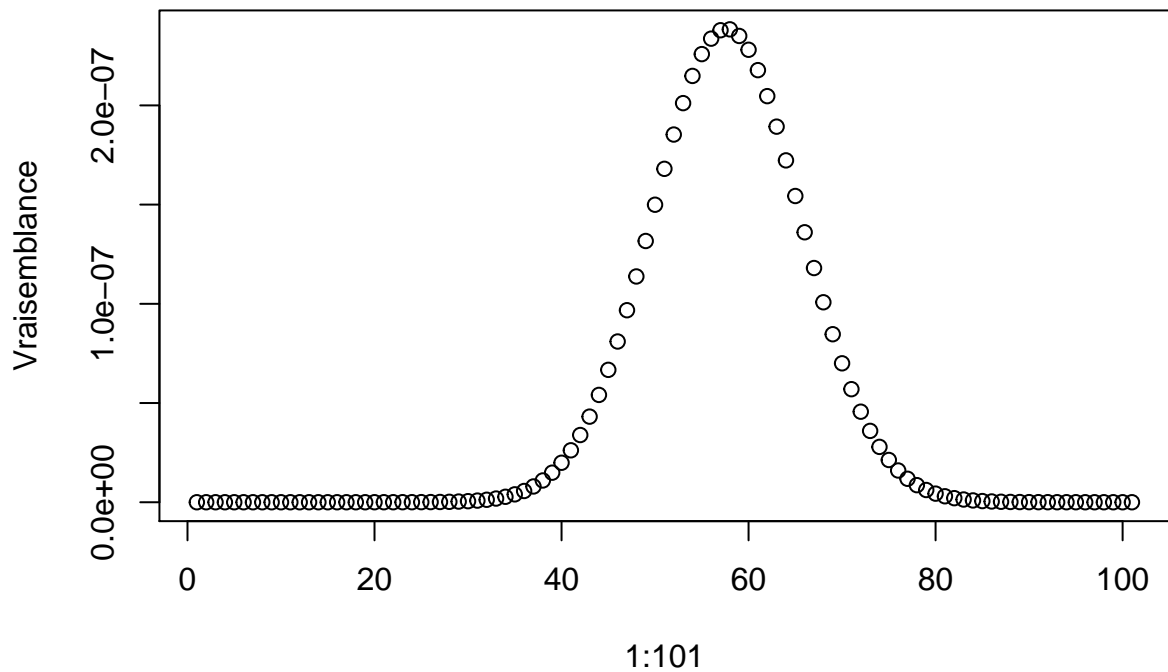
Pour combattre ces instabilités dues au produit des densités de dans le calcul de la vraisemblance on peut utiliser la log-vraisemblance pour obtenir des sommes de densités plus facile à différencier dès que la taille augmente

Question 10

$$L_{\text{norm}} = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2}$$

```
L_norm <- function (mu, std, X) {
  n = length(X)
  return(1/(std*sqrt(2*pi))**n * exp(-(1/(2*std**2))*sum((X-mu)**2)))
}
```

Question 11



Le maximum de vraisemblance est obtenu entre la 40ème et 60ème itération pour retrouver en moyenne à $i=50$ $\mu = 50 * 0.04 = 2$

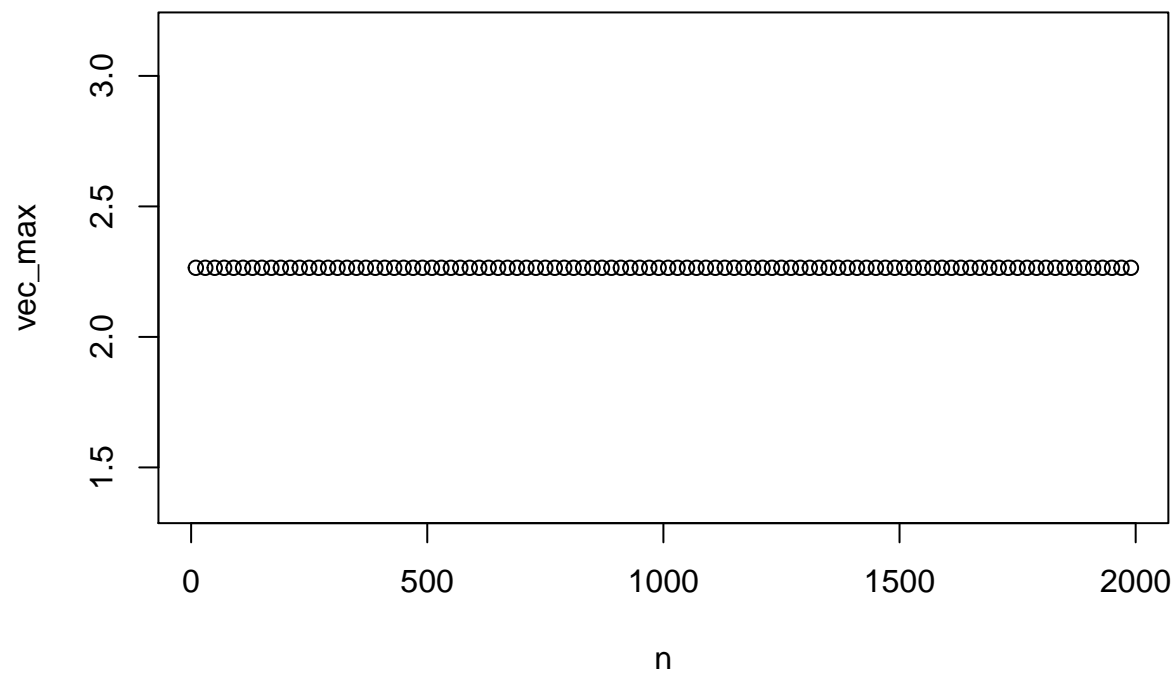
Question 12

On s'approche du maximum de vraisemblance :

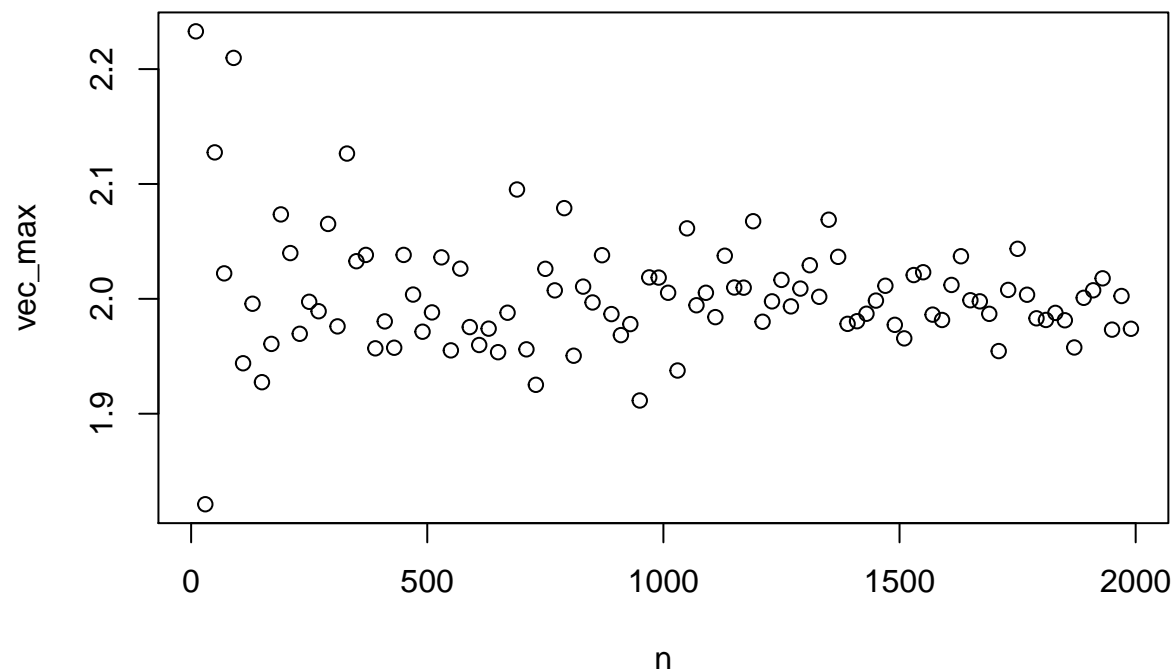
```
## [1] 2.264819
```

Question 13

On retrouve les instabilités de la vraisemblance , exacerbées par les exponentielles des gaussiennes, qui ne permettent donc pas de mesurer de convergence et produisent parfois une constante éloignée de μ :



Passons à la log vraisemblance en calculant des sommes:



Ici, il n'y a pas de problèmes d'instabilité numérique, et ce pour toute taille N d'échantillon, de plus la convergence vers $\lambda = 2$ est de plus en plus précise pour N grand.

Question 14

L'estimateur de maximum de vraisemblance pour l'écart-type est :

$$\sigma = \sum_{i=1}^n (x_i - \mu)^2$$

$$E(\sigma^2) = \frac{(n-1)\sigma^2}{n}$$

Ainsi l'écart type peut être biaisé et il serait intéressant de le faire varier

Maximum de vraisemblance de la concentration horaire maximale d'ozone (en $\mu\text{g}/\text{m}^3$) pour chaque site & saison :

