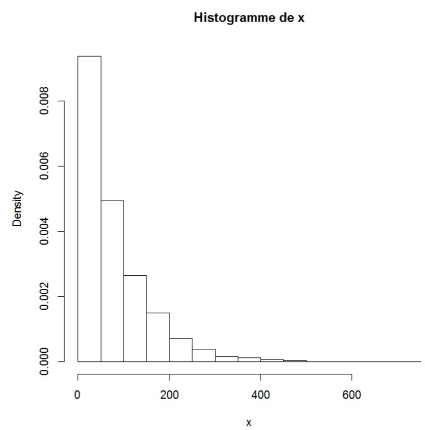
SOLUTION TP n° 7

Solution 1. On considère le jeu de données "rayons" disponible ici :

`https://chesneau.users.lmno.cnrs.fr/rayons.txt`

Dans celui-ci, on mesure les temps en secondes entre les arrivées de $n = 3935$ photons. L'histogramme associé à ces valeurs est :



Vu son allure, il est raisonnable de penser que le temps en secondes entre deux arrivées de photons est une *var* X suivant la loi gamma $\Gamma(\alpha, \beta)$, *i.e.* de densité :

$$f(x) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{si } x \geq 0, \\ 0 & \text{sinon,} \end{cases}$$

où $\Gamma :]0, \infty[\to [0, \infty[$ désigne la fonction gamma d'Euler définie par : $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

Ici, α et β sont des paramètres inconnus que l'on souhaite estimer ponctuellement à l'aide des données : $x_1, \ldots, x_n$ (observation / réalisation d'un n-échantillon $(X_1, \ldots, X_n)$ de X).

1. Montrer que
$$\mathbb{E}(X) = \frac{\alpha}{\beta}, \qquad \mathbb{V}(X) = \frac{\alpha}{\beta^2} :$$

   En faisant le changement de variable $y = \beta x$ et l'égalité : $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$, on obtient

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\beta}\right)^\alpha e^{-y} \frac{1}{\beta} dy = \frac{1}{\beta \Gamma(\alpha)} \int_0^\infty y^{(\alpha+1)-1} e^{-y} dy \\ &= \frac{1}{\beta \Gamma(\alpha)} \Gamma(\alpha + 1) = \frac{\alpha}{\beta}. \end{aligned}$$

On a $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Il reste à calculer $\mathbb{E}(X^2)$. En faisant de nouveau le changement de variable $y = \beta x$ et l'égalité : $\Gamma(\alpha+2) = (\alpha+1)\alpha\Gamma(\alpha)$, on obtient

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^{\infty} x^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} \left(\frac{y}{\beta}\right)^{\alpha+1} e^{-y} \frac{1}{\beta} dy = \frac{1}{\beta^2 \Gamma(\alpha)} \int_0^{\infty} y^{(\alpha+2)-1} e^{-y} dy \\
&= \frac{1}{\beta^2 \Gamma(\alpha)} \Gamma(\alpha+2) = \frac{\alpha(\alpha+1)}{\beta^2}.
\end{aligned}$$

D'où

$$\mathbb{V}(X) = \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2}.$$

2. Mettre les données dans un vecteur `x` :

```
x = scan("https://chesneau.users.lmno.cnrs.fr/rayons.txt")
```

3. En estimant ponctuellement $\mathbb{E}(X)$ par $\overline{x}$ et $\mathbb{V}(X)$ par s^2, donner des estimations ponctuelles de α et β par la méthode des moments : La question 1 entraîne $\alpha = \beta\mathbb{E}(X)$, donc on peut écrire :

$$\mathbb{V}(X) = \frac{\alpha}{\beta^2} = \frac{\beta\mathbb{E}(X)}{\beta^2} = \frac{\mathbb{E}(X)}{\beta},$$

soit

$$\beta = \frac{\alpha}{\beta^2} = \frac{\beta\mathbb{E}(X)}{\beta^2} = \frac{\mathbb{E}(X)}{\mathbb{V}(X)}.$$

Ainsi, une estimation de β par la méthode des moments est

$$b = \frac{\overline{x}}{s^2}.$$

D'autre part, comme $\alpha = \beta\mathbb{E}(X)$, une estimation ponctuelle de α est

$$a = b\overline{x} = \frac{\overline{x}}{s^2}\overline{x} = \frac{\overline{x}^2}{s^2}.$$

On considère les commandes :

```
b = mean(x) / sd(x)^2
a = mean(x)^2 / sd(x)^2
c(b, a)
```

Cela renvoie : `[1] 0.01266144 1.01209495`. Par la méthode des moments, une estimation ponctuelle de α est 1.01209495 et une estimation ponctuelle de β est $b = 0.01266144$.

4. Consulter les aides suivantes de R : `help(dgamma)` et `help(nlm)`.

5. On cherche maintenant à estimer ponctuellement α et β par la méthode du maximum de vraisemblance. Reproduire et comprendre les enjeux des commandes suivantes :

```
mlog = function(theta, x){
sum(-dgamma(x, shape = theta[1], rate = theta[2], log = TRUE))
}
nlm(mlog, c(1, 1), x = x)$estimate
```

Cela renvoie : `[1] 1.02634117 0.01283969`. En utilisant la méthode du maximum de vraisemblance, on a ainsi une estimation ponctuelle de α qui est 1.02634117 et une estimation ponctuelle de β qui est 0.01283969.

6. Quel est l'objet mathématique renvoyé par les commandes suivantes ?

```
mv = nlm(mlog, c(1, 1), x = x, hessian = TRUE)
mv$hessian
```

Cela renvoie :

```
            [,1]        [,2]
[1,]    6231.295   -305284.4
[2,] -305284.399 24121374.2
```

On obtient une estimation ponctuelle de la matrice Hessienne de la fonction de log-vraisemblance.

7. Quel sont les objets mathématiques renvoyés par les commandes suivantes ?

```
hat.alpha = mv$estimate[1]
hat.beta = mv$estimate[2]
z = qnorm((1 + 0.95) / 2)
inv.fish = solve(mv$hessian)
hat.alpha + c(-1, 1) * z * sqrt(inv.fish[1, 1])
hat.beta + c(-1, 1) * z * sqrt(inv.fish[2, 2])
```

Les commandes renvoient deux intervalles de confiance pour α et β au niveau 95% en utilisant la convergence asymptotique de l'estimateur du maxium de vraisemblance.

Solution 2. Soient $\theta \in \mathbb{N}^*$, X une *var* suivant la loi uniforme $\mathcal{U}(\{1, \ldots, \theta\})$:

$$\mathbb{P}(X = x) = \frac{1}{\theta}, \qquad x \in \{1, \ldots, \theta\},$$

$n \in \mathbb{N}^*$ et $(X_1, \ldots, X_n)$ un n-échantillon de X. Ici, θ est un paramètre inconnu que l'on souhaite estimer à l'aide de $(X_1, \ldots, X_n)$.

1. Déterminer un estimateur obtenu par la méthode des moments, et l'estimateur du maximum de vraisemblance : On a

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) = \sum_{x=1}^{\theta} x \times \frac{1}{\theta} = \frac{1}{\theta} \times \frac{\theta(\theta+1)}{2} = \frac{\theta+1}{2}.$$

Il vient $\theta = 2\mathbb{E}(X) - 1$. Ainsi, la méthode des moments nous donne l'estimateur

$$U_n = \frac{2}{n} \sum_{i=1}^{n} X_i - 1.$$

La fonction de vraisemblance est

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) = \frac{1}{\theta^n}, \qquad \sup(x_1, \ldots, x_n) \in \{1, \ldots, \theta\}.$$

Comme cette fonction est décroissante en θ, on a

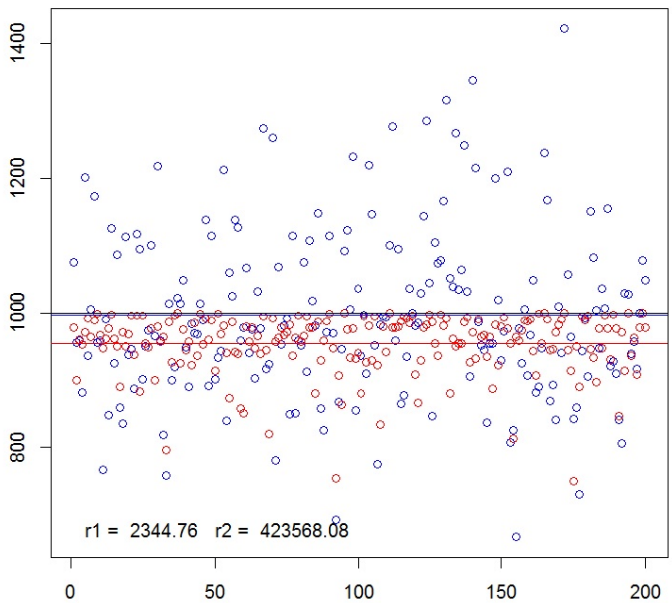
$$V_n = \operatorname*{Argmax}_{\theta \in \mathbb{N}^*} L(\theta; X_1, \ldots, X_n) = \sup(X_1, \ldots, X_n)$$

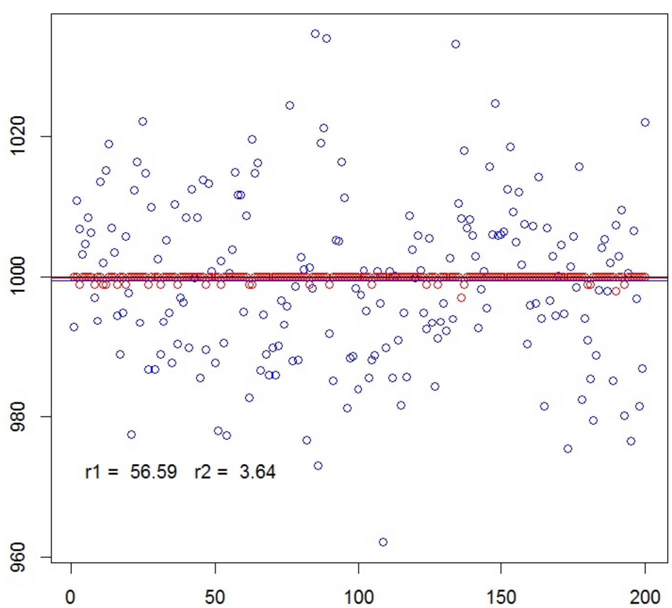
Ainsi, la méthode du maximum de vraisemblance nous donne l'estimateur

$$V_n = \sup(X_1, \ldots, X_n).$$

2. Reproduire et comprendre les enjeux des commandes suivantes :

```
simunif = function(teta, n, k) {
est1 = numeric(k)
est2 = numeric(k)
for (i in 1 : k) {
echant = sample(1 : teta, n, replace = T)
est1[i] = 2 * mean(echant) - 1
est2[i] = max(echant)
}
yrange = range(est1, est2)
plot(1:k, est1, type = "p", col = "blue", ylim = yrange, main = "", sub = "",
xlab = "", ylab = "")
points(1:k, est2, type = "p", col = "red", ylim = yrange, main = "", sub = "",
xlab = "", ylab = "")
abline(h = teta, col = "black")
abline(h = mean(est1), col = "blue")
abline(h = mean(est2), col = "red")
r1 = round(sum(est1 - teta)^2 / k, 2)
r2 = round(sum(est2 - teta)^2 / k, 2)
text(1, min(yrange) + 10, paste("r1 = ",r1,"  r2 = ",r2), pos = 4)
}
simunif(1000, 20, 200)
simunif(1000, 2000, 200)
```





On constate que, quand n est petit, l'estimateur U_n est meilleur, et quand n est grand, V_n est meilleur.