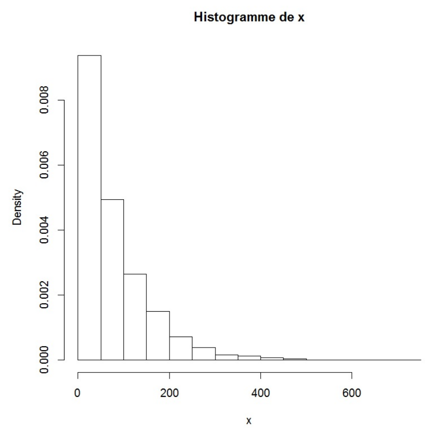
TP n° 7 : Estimateur du maximum de vraisemblance

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

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

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.

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
```

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

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

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])
```

Exercice 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.

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)
```