TP1 : Estimateurs du maximum de vraisemblance

Nous observons les données provenant du fichier `fiabilites.csv` qui présente le temps de fonctionnement avant défaillance d'un circuit intégré d'ordinateur.
On observe alors 1000 données, que nous avons importé dans un data frame sur R. Nous avons utilisé la fonction read.csv pour enregistrer ces données.
Tel que :
```
read.csv(data <- read.csv("~/Linux/SI/TP1/fiabilites.csv", " ",
                    header = FALSE, dec=",",sep=" ")
```

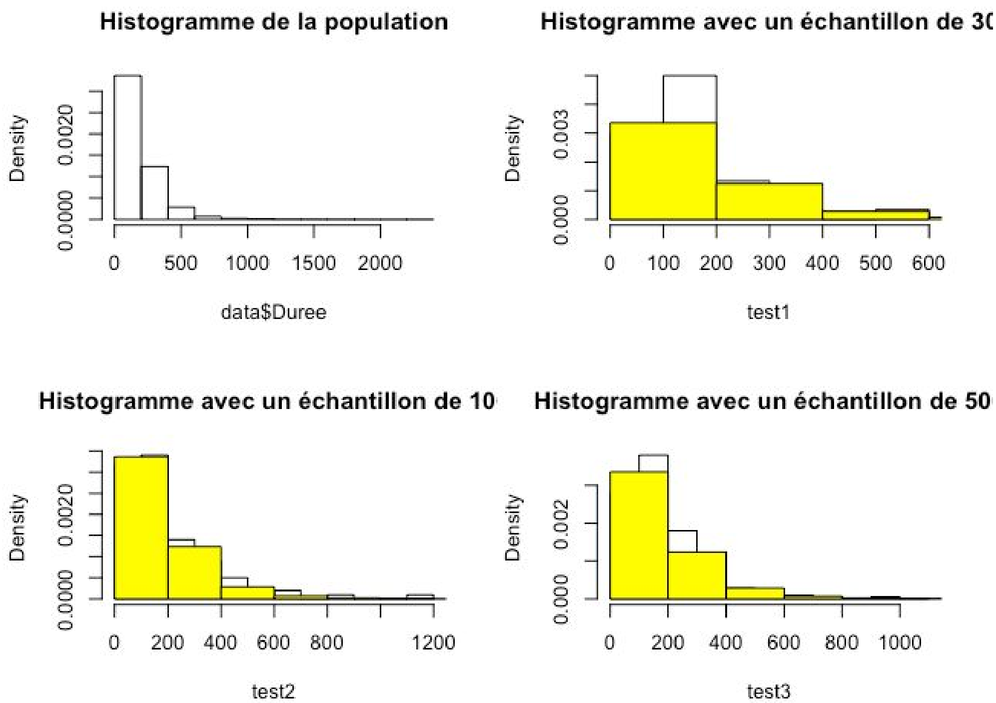
On peut alors observer alors (en terme de statistique descriptive) les indicateurs suivants :

```
Duree
 Min.   :   7.96
 1st Qu.:  92.31
 Median : 146.06
 Mean   : 189.48
 3rd Qu.: 236.15
 Max.   :2325.58
```

De même, nous avons généré différents échantillons aléatoire de taille n=30, 100 et 500. On utilise la fonction sample de R.

On remarque que, sur le code, nous avons nommé la colonne des données : Duree.
Pour représenter les distributions des différents échantillons, nous représentons cela par des histogrammes :

Avec l'attribut prob=TRUE, on retrouve l'histogramme de la populations que l'on "greffe" sur les histogrammes des échantillons. On observe ici :



On remarque alors les échantillons suivent la même loi que la population.

Nous nous penchons, maintenant, sur le log- vraisemblance, car notre but est de trouver les estimateurs des paramètres de la loi avec le maximum de vraisemblance. On crée alors une fonction sous R, qui permet de réaliser le log de vraisemblance.

En effet nous avons cette fonction :

$$f(x\,;\,\mu,\sigma^2) = \frac{1}{\sigma x\sqrt{2\pi}}\, e^{\frac{-(ln(x)-\mu)^2}{2\sigma^2}} \,.\, 1_{x>0}$$

Vraisemblance de { X1, ... , Xn } :

$$= \prod_{i=1}^{n} f(Xi\,;\,\mu,\sigma^2) = \frac{1}{\prod_{i=1}^{n} Xi} \,.\, \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \,.\, e^{-\frac{1}{2\sigma^2}\,.\,\sum_{i=1}^{n}(ln(Xi-\mu)^2} \,.\, \prod_{i=1}^{n} 1_{Xi>0}$$

Ainsi le log de vraisemblance { X1,...,Xn } :

$$= -\sum_{i=a}^{n} log(Xi) - n\, log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(ln(Xi) - \mu)^2 + log(\prod_{i=1}^{n} 1_{Xi>0}$$

Ainsi X → Log X(μ, σ^2)

Voici les résultats de la vraisemblance :
Pour n = 30 : on a 409.611
Pour n = 100 : on a 1420.763
Pour n = 500 : on a 7092.51
Et pour la population on a : 141375.1

On a alors créer les fonctions suivantes LV1 mais aussi LV2, LV1 utilise les fonctions de R, tel que la fonction de densité de la loi normale (dlnorm).

On veut maintenant, estimer en utilisant la fonction optim mu et sigma.
On trouve pour mu : 4.9941
Et pour sigma : 0.7043
Ceci sont les réponses pour la population totale.

De plus, on sait que X → Log X (μ, σ^2)

On sait que l'estimateur de mu est $\overline{X} = \frac{1}{n}\sum Xi$, Mais sur cette loi l'estimateur est

$\overline{X} = \frac{1}{n}\sum ln(Xi)$. On trouve alors pour mu 4,9941 pour la population, ce qui est alors identique à la fonction optim.
De même pour la variance, on l'a calculé avec la fonction var de R. On trouve alors : 0.704347 pour la population.

On remarque donc d'estimer les paramètres par le maximum de vraisemblance est puissant, en effet, on retrouve les mêmes résultats avec le log de vraisemblance et avec les estimateurs "connu" des paramètres.

On peut trouver de nouveau estimateur σ et μ en fonction de l'espérance et de la variance des Xi. Pour cela nous utilisons la formule de l'espérance et la variance d'une loi log-normale :
Espérance d'une loi log-normal : $e^{m + \sigma^2/2}$
Variance d'une loi log-normal : $e^{2m + \sigma^2} (e^{\sigma^2} - 1)$

Ici, on remarque que, en passant par la loi log-normale, on retrouve des estimateurs différents. En effet, pour mu on trouve : 4.988786 et pour sigma on trouve : 0.714821.
Ce sont alors d'autres estimateurs de ces paramètres.