DELEGUE Hugo - TANIEL Rémi

Compte rendu TP1 : Estimateurs

Introduction

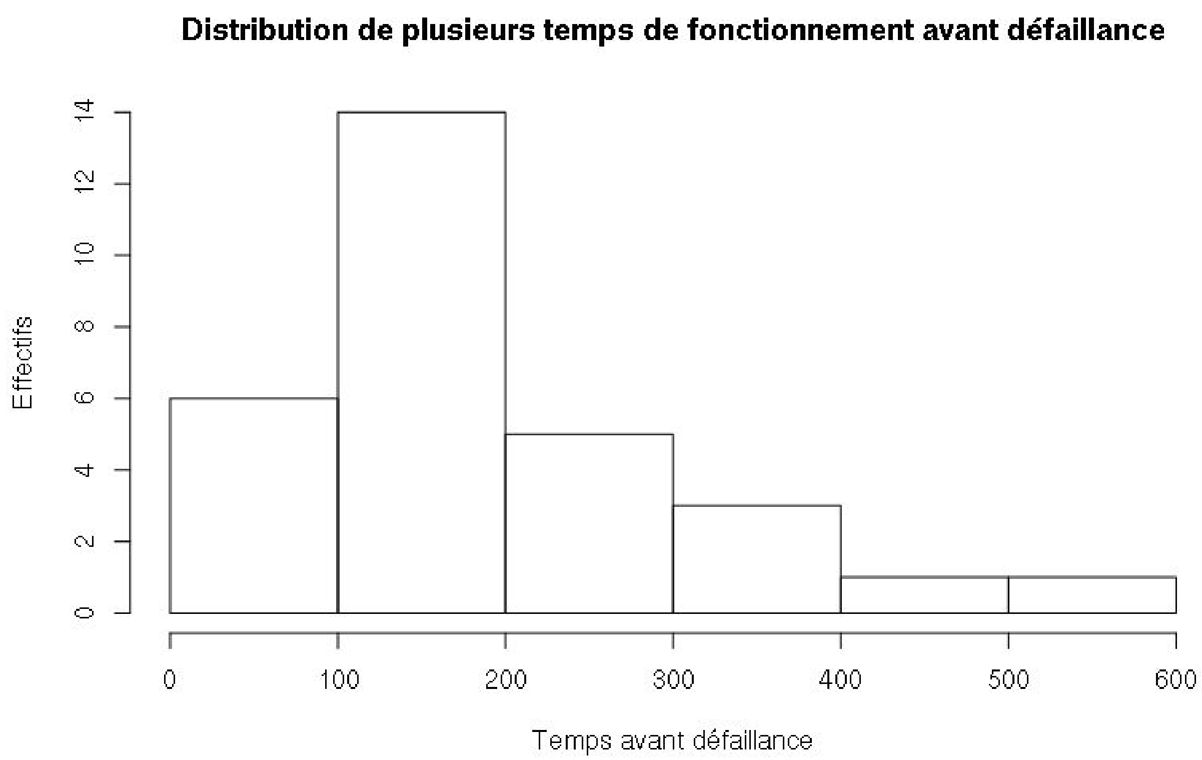
Nous avons un jeu de données représentant le temps de fonctionnement avant défaillance d'un circuit intégré d'ordinateur, on sait également que les données du fichier sont distribués selon une loi log-normale de paramètres μ et σ^2.

L'objectif du TP est donc d'estimer le plus précisément possible ces 2 paramètres.

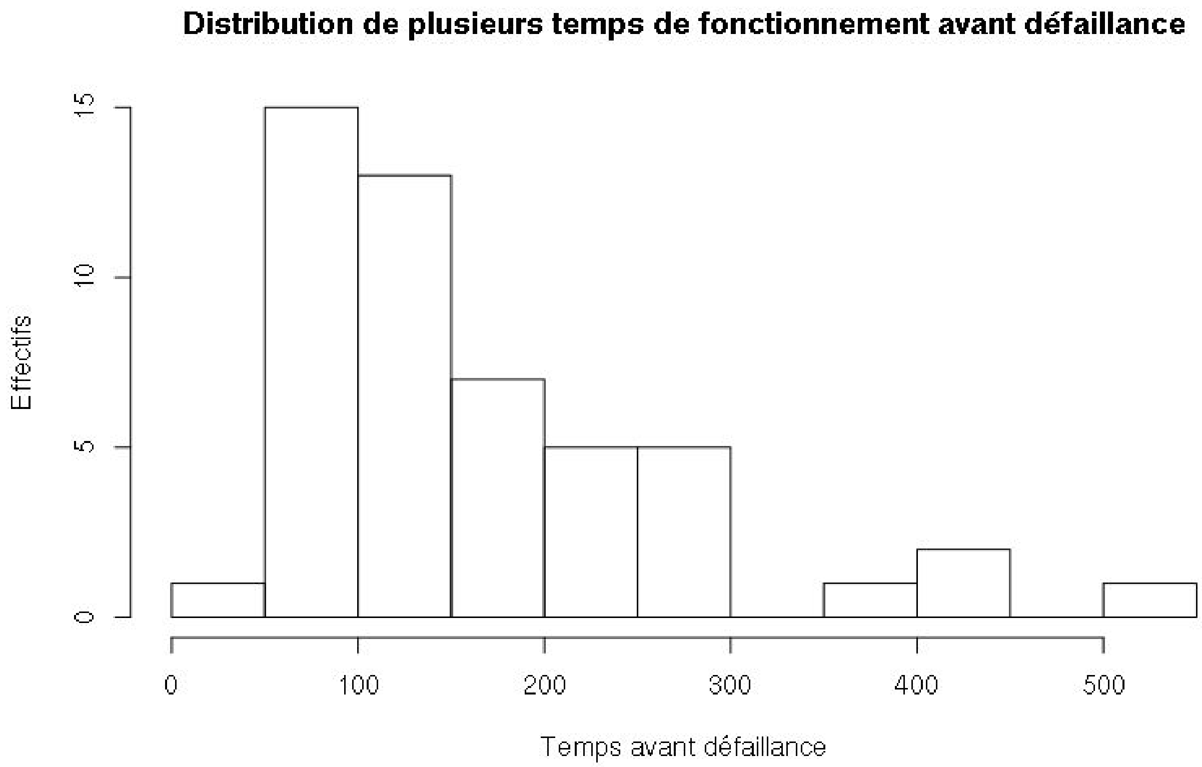
Question 1

On se décide dans un premier temps de générer 3 échantillons aléatoires ($X_1, X_2, ..., X_n$) provenant des données fournis de taille $n = 30, 50, 80$, voici la représentation graphique de ces échantillons :

n = 30

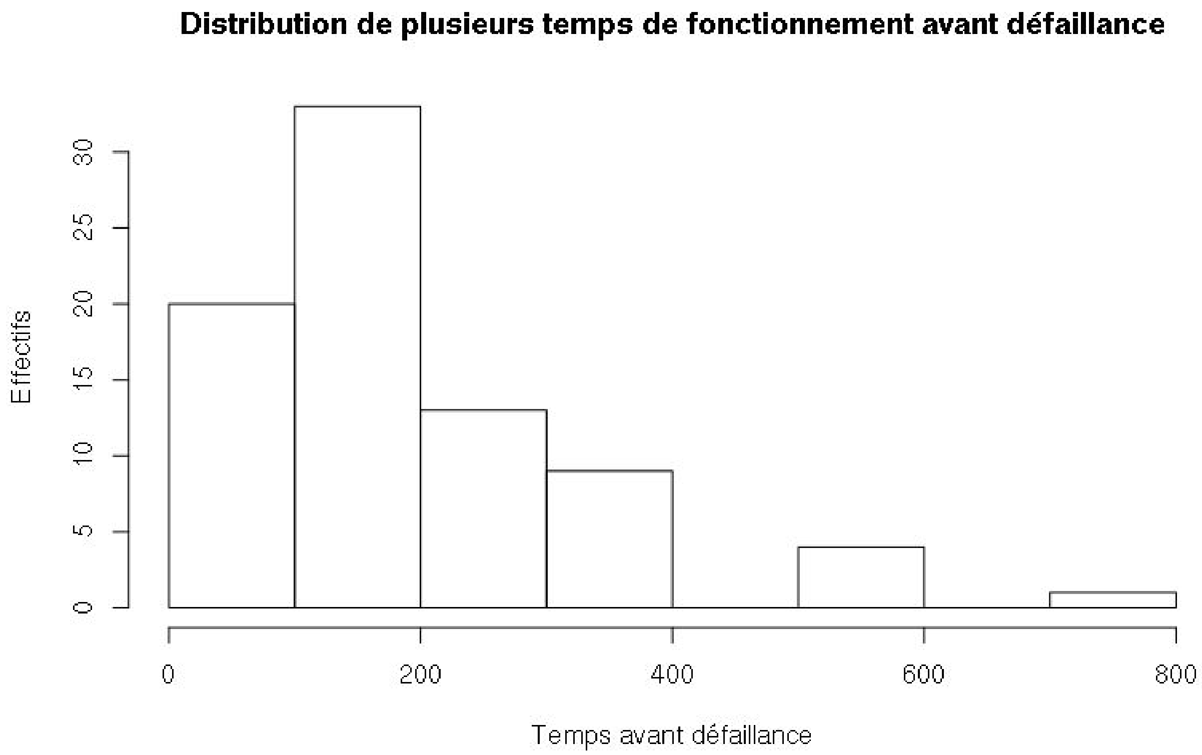


n = 50



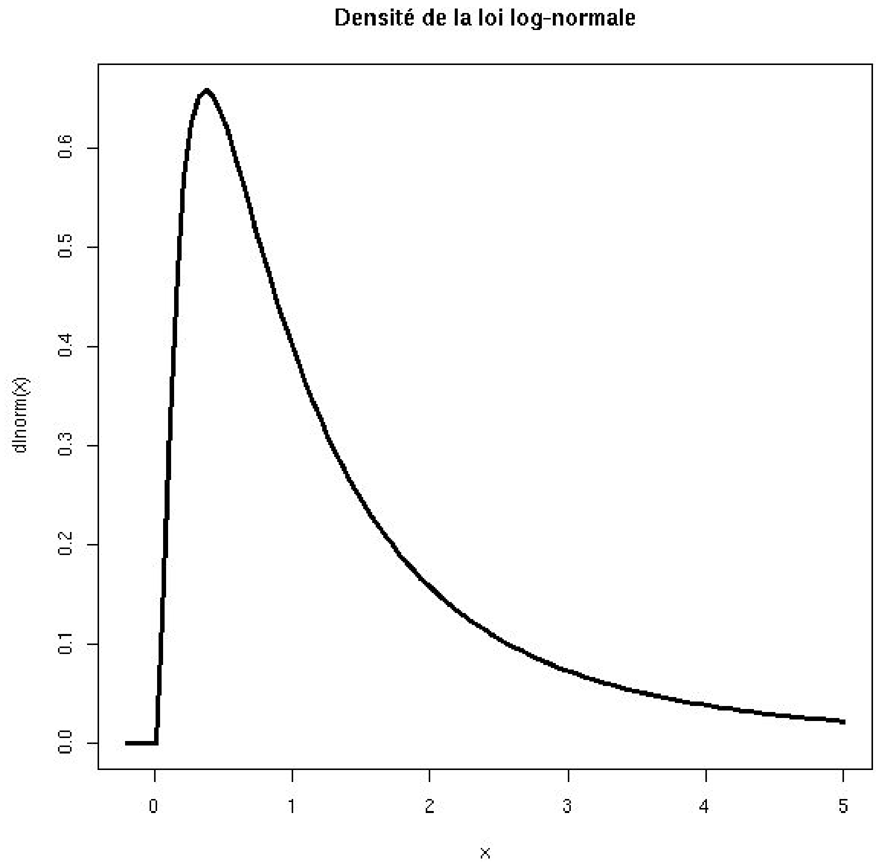
Distribution de plusieurs temps de fonctionnement avant défaillance
Effectifs
0
5
10
15
0
100
200
300
400
500
Temps avant défaillance

n = 80



Distribution de plusieurs temps de fonctionnement avant défaillance
Effectifs
0
5
10
15
20
25
30
0
200
400
600
800
Temps avant défaillance

On remarque que plus la taille de l'échantillon est importante, plus on se rapproche de la courbe de la densité d'une loi log-normale :



Question 2

Le but étant d'estimer μ et σ, nous allons utiliser la méthode du maximum de vraisemblance pour estimer ces 2 paramètres, c'est pourquoi nous devons dans un premier temps, calculer la fonction de log vraisemblance d'un échantillon $(X_1, X_2, ..., X_n)$ en μ et σ

La fonction de répartition d'une loi log-normale de paramètre (μ, σ^2) est la suivante :

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x - \mu)^2}{2\sigma^2}}$$

Pour cela, nous devons maximiser le produit suivant étant donné que $X_1, X_2, ..., X_n$ sont identiquement distribué et indépendants :

$$L = \prod_{i=1}^{n} f(x_i)$$

Pour simplifier le calcul de ce produit, on passe par la log-vraisemblance, on obtient donc la fonction de log-vraisemblance suivante :

$$Log(L) = log(\prod_{i=1}^{n} f(x_i)) = \sum_{i=1}^{n} log(f(x_i)) = \frac{n}{2} log(2\pi\sigma^2) - \Sigma\, log(x_i) - \frac{1}{2}\Sigma\, (\frac{ln(x) - \mu}{\sigma})^2$$

Question 3

Les estimateurs $\widehat{\mu}$ et $\widehat{\sigma}$ sont obtenus grâce à :

$\widehat{\mu} = argmax_{\mu}(Log(L))$

$\widehat{\sigma} = argmax_{\sigma}(Log(L))$

Or $argmax_{\sigma}(Log(L))$ revient à trouver quand s'annule la dérivé partielle de $Log(L)$ par rapport à σ ,c'est à dire résoudre l'équation (et inversement pour μ) :

$\frac{\delta Log(L)}{\delta\sigma} = 0$

Nous obtenons donc les valeurs suivantes pour les estimateurs $\widehat{\mu}$ et $\widehat{\sigma}$ selon la valeur de

n = 30

$\widehat{\mu} = 5,102$

$\widehat{\sigma} = 0,609$

n = 50

$\widehat{\mu} = 4,929$

$\widehat{\sigma} = 0,591$

n = 80

$\widehat{\mu} = 4,992$

$\widehat{\sigma} = 0,675$

On remarque que plus la taille de l'échantillon augmente, plus on se rapproche d'une certaine valeur, qui est la réelle valeur des paramètres μ et σ.

Question 4

On propose de retrouver analytiquement les estimateurs du maximum de vraisemblance puis de comparer ces résultats avec ceux de la question précédente.

On sait que $X \sim LogN(\mu, \sigma^2)$
donc on peut affirmer que
$log(X) \sim N(\mu, \sigma^2)$
Donc
$\mu = E(log(X)) = \frac{1}{n}\sum_{i=1}^{n} log(X_i)$

$\sigma = Var(log(X)) = \frac{1}{n}\sum_{i=1}^{n} (log(X_i) - E(log(X))^2$

En appliquant ces formules à nos échantillons, nous obtenons :

n = 30
$\widehat{\mu} = 5,102$
$\widehat{\sigma} = 0,619$

n = 50
$\widehat{\mu} = 4,929$
$\widehat{\sigma} = 0,597$

n = 80
$\widehat{\mu} = 4,992$
$\widehat{\sigma} = 0,679$

On remarque que les 2 estimateurs de μ sont exactement les mêmes tandis que les estimateurs de σ diffèrent légèrement.

Question 5

Soit $\overline{x}$ la moyenne empirique d'un de nos échantillons, on le calcule grâce à :
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Puis, nous pouvons calculer la moyenne empirique s^2 associé grâce à la formule :
$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$

De plus, on sait que $X \sim LogN(\mu, \sigma^2)$, donc

$$E(X) = e^{\mu+\frac{\sigma^2}{2}}$$
$$Var(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

Nous allons donc exprimer les paramètres μ et σ en fonction de $E(X)$ et $Var(X)$, nous obtenons :

$$\sigma^2 = ln(1 + \frac{Var(X)}{\Xi(X)^2})$$
$$\mu = ln(E(X) - \frac{1}{2} ln(1 + \frac{Var(X)}{E(X)^2})$$

En remplaçant $E(X)$ et $Var(X)$ respectivement par la moyenne empirique $\overline{x}$ et la variance empirique s^2 ,nous obtenons les résultats suivants la taille de l'échantillon :

n = 30

$\mu = 5,118$
$\sigma = 0,566$

n = 50

$\mu = 4,937$
$\sigma = 0,581$

n = 80

$\mu = 5,004$
$\sigma = 0,661$

Cette fois-ci, nos nouveaux estimateurs diffèrent un peu plus de ceux trouvés précédemment.