# TP 6 : Méthode des moments VS méthode du maximum de vraisemblance

Ce TP étudie l'estimation par la méthode des moments (que l'on a déjà traitée sans la nommer) et l'estimation par maximum de vraisemblance, en montrant que ces deux méthodes peuvent conduire à des estimateurs différents.

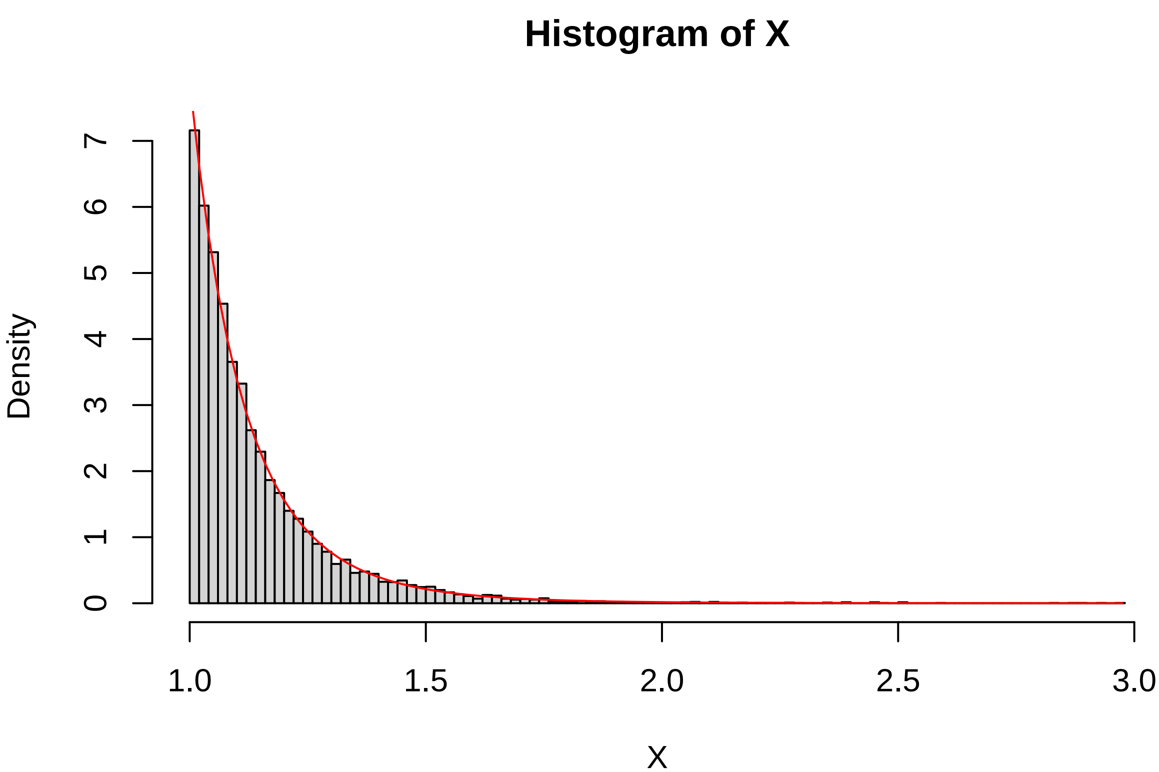
**Estimation d'un paramètre d'une power law.**

**Première partie : quelques simulations**

**1.** Voici le code :

$F_a(x) = \int_1^x (a-1)t^{-a}dt = 1 - x^{-(a-1)}$. Donc si $U$ est de loi uniforme sur $[0,1]$ alors $\left(\frac{1}{1-U}\right)^{1/a-1}$ est de loi de densité $f_a$. Le code suivant simule l'échantillon comme souhaité : on remarque que les courbes coincident bien.

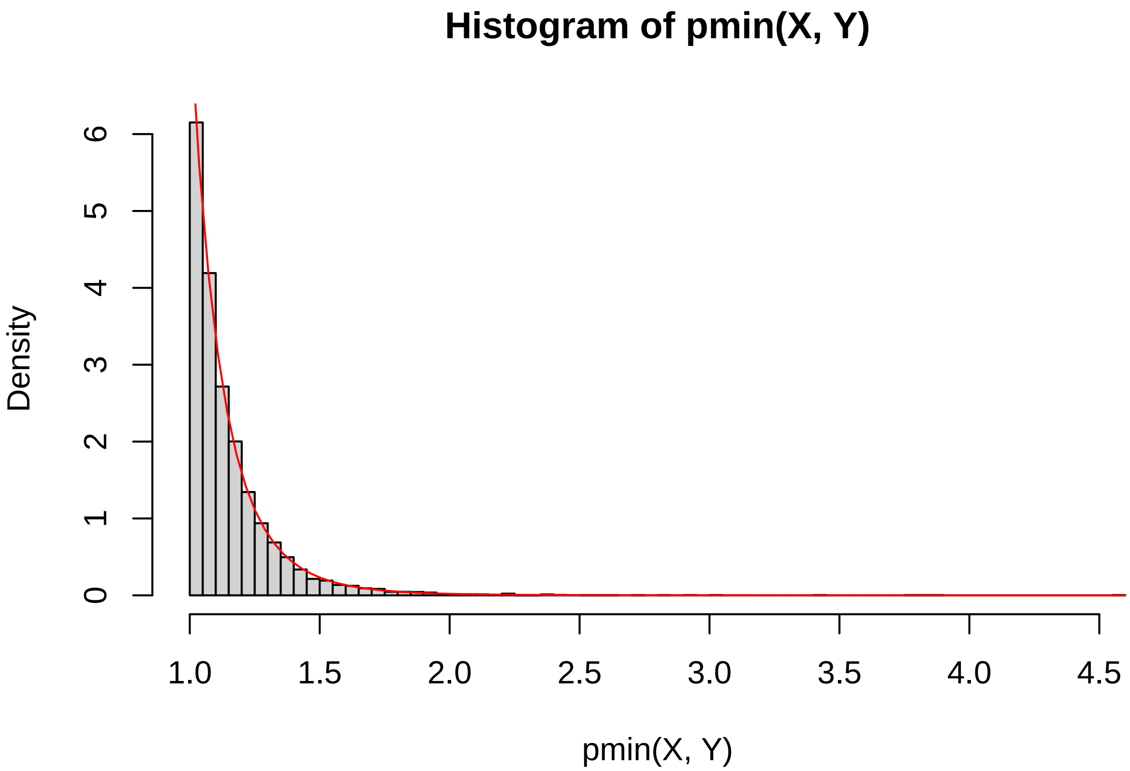
```
a = 8.9
n = 10^4
U = runif(n)
X = (1/(1-U))^{1/(a-1)}
hist(X,freq=FALSE,breaks=120)
curve((a-1)/x^a,from=1,to=max(X),col="red", add=TRUE)
```




On remarque que la densité théorique coincide bien avec l'histogramme.

**2.** Voici le code :

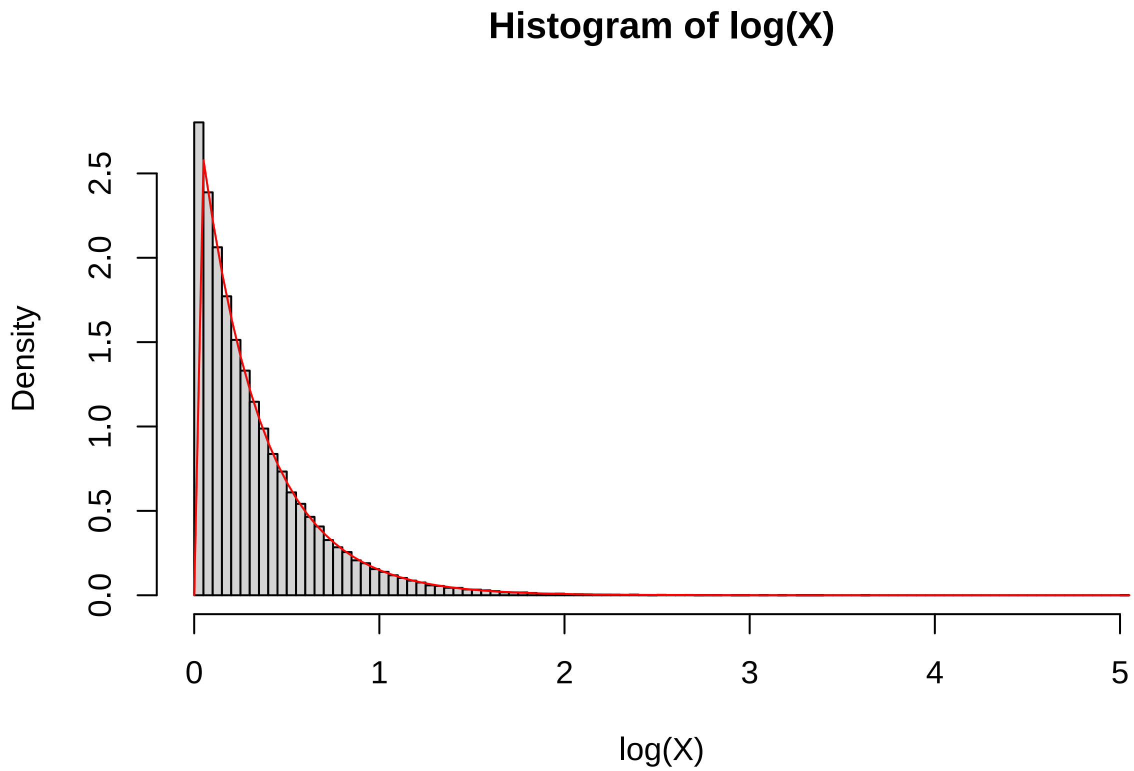
```
a = 3.4
b = 6.2
n = 10^4
U1 = runif(n)
X = (1/(1-U1))^{1/(a-1)}
U2 = runif(n)
Y = (1/(1-U2))^{1/(b-1)}

hist(pmin(X,Y),freq=FALSE,breaks=120)
curve((a+b-2)/x^(a+b-1),col="red", add=TRUE)
```




On remarque que la densité théorique coincide bien avec l'histogramme.

**3.** Voici le code :

```
a = 4
n = 10^5
U1 = runif(n)
X = (1/(1-U1))^{1/(a-1)}

hist(log(X),freq=FALSE,breaks=120)
curve(dexp(x,rate=a-1),col="red", add=TRUE)
```




La loi de $log(X)$ est en fait la loi exponentielle de paramètre $a - 1$. Cela est illustré en superposant à l'histogramme la fonction de la densité voulue.

**Deuxième partie : deux estimateurs**

**4.** Par la LFGN, $\bar{X}_n$ converge p.s. vers $\mathbb{E}_{X \sim PL(a)}[X]$ lorsque cette dernière est finie, ce qui n'arrive que si $a > 2$. Dans ce cas $\bar{X}_n$ converge p.s. vers $\frac{a-1}{a-2}$. On déduit donc l'estimateur

$$\hat{a}_{MM} = \frac{2\bar{X}_n - 1}{\bar{X}_n - 1}.$$

**5.** La vraisemblance s'écrit ici

$$L(a, X_1, \dots, X_n) = (a-1)^n \prod_{i=1}^{n} X_i^{-a},$$

et en passant au log on a

$$\log L(a, X_1, \dots, X_n) = n \log(a-1) - a \sum_{i=1}^{n} \log X_i.$$

La vraisemblance est dérivable, hourra ! Etudions donc les extremas de cette fonction de $a$ en dérivant :

$$\frac{d}{da} \log L(a, X_1, \dots, X_n) = \frac{n}{a-1} - \sum_{i=1}^{n} \log X_i.$$

On résout ensuite

$$\frac{d}{da} \log L(\hat{a}, X_1, \dots, X_n) = 0$$

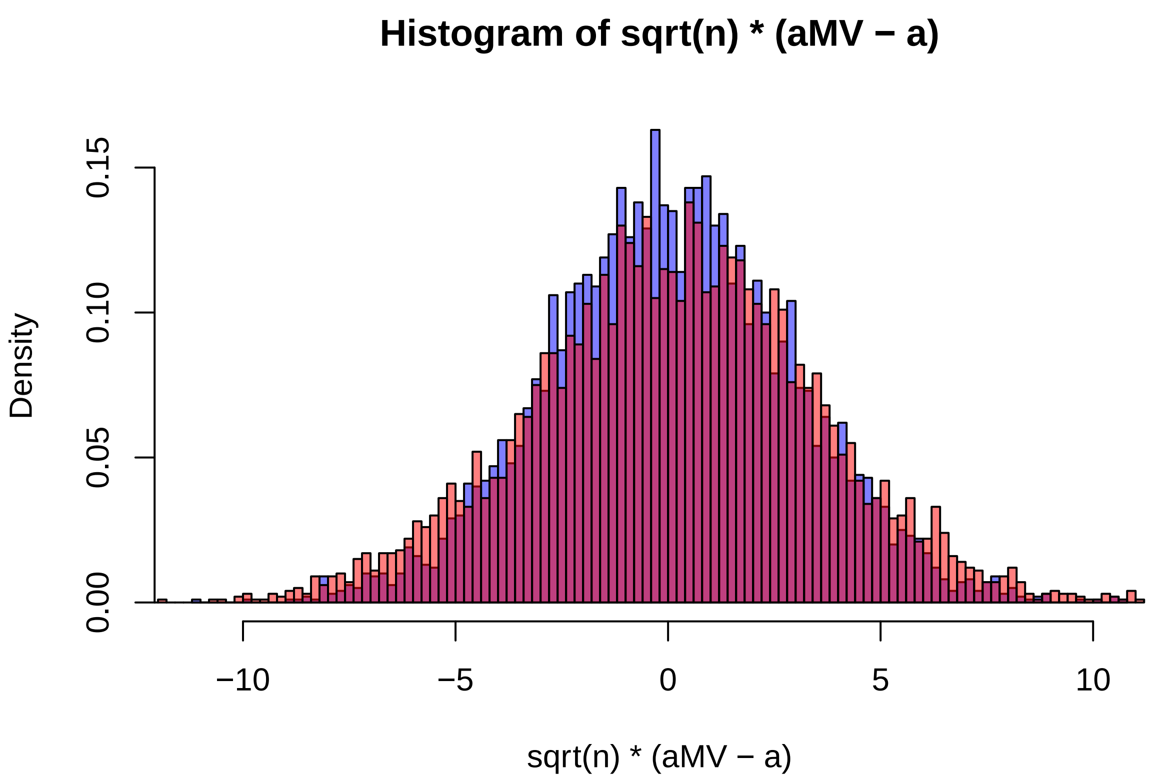
ce qui donne

$$\hat{a}_{MV} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\log(X_i)} + 1.$$

Au passage, on voit que c'est un estimateur bien différent de $\hat{a}_{MM}$.

**6.** Si la conjecture faite à la question 3 est vraie, alors les variables $log(X_i)$ sont i.i.d. de loi $\mathcal{E}(a-1)$ d'espérance finie pour tout $a > 1$ : la loi des grands nombres s'applique, $\sum_{i=1}^{n}\log(X_i)$ converge p.s. vers $\frac{1}{a-1}$ ce qui donne $\hat{a}_{MV} \to a$ p.s.

```
a = 3.9
n = 10^4
N = 5000

aMM = replicate(N,0)
aMV = replicate(N,0)

for (i in 1:N){
  X = (1/(1-runif(n)))^(1/(a-1))  # je simule mon échantillon de taille n de loi PL(a)
  aMM[i] = (2*mean(X)-1)/(mean(X)-1) # je calcule AMM
  aMV[i] = 1/mean(log(X)) + 1 # ja calcule AMV
}

hist(sqrt(n)*(aMV-a),freq=FALSE,breaks=100,
     col = rgb(red = 0, green = 0, blue = 1, alpha = 0.5))
hist(sqrt(n)*(aMM-a),freq=FALSE,breaks=100,
     col = rgb(red = 1, green = 0, blue = 0, alpha = 0.5),add=TRUE)
```




**7.**

On voit bien la normalité asymptotique de ces estimateurs, et on voit de plus que la Gaussienne associée à $\hat{a}_{MM}$ (en rouge) a une variance plus grande que celle associée à $\hat{a}_{MV}$ (en bleu). Autrement dit, elle est plus étalée. C'est donc l'estimateur du maximum de vraisemblance qui semble être le meilleur.

```
mean((aMM-a)^2)
```

**8.**

```
## [1] 0.001140567
```

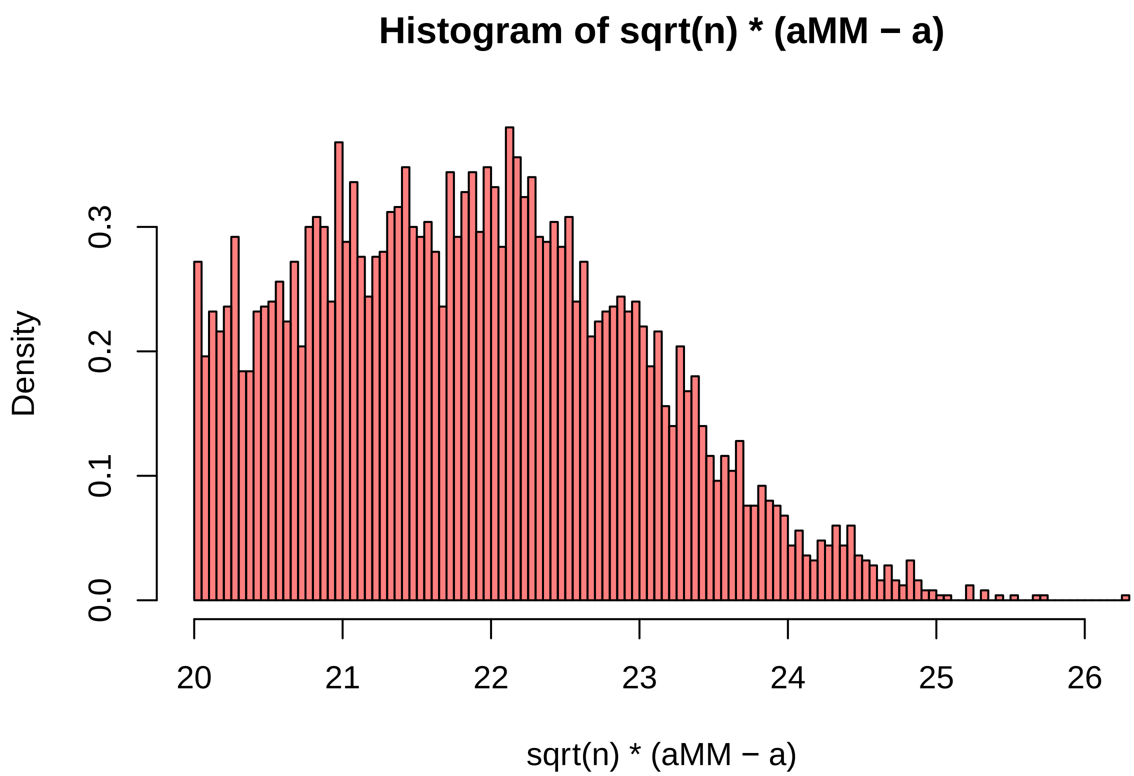
```
mean((aMV-a)^2)
```

```
## [1] 0.000832503
```

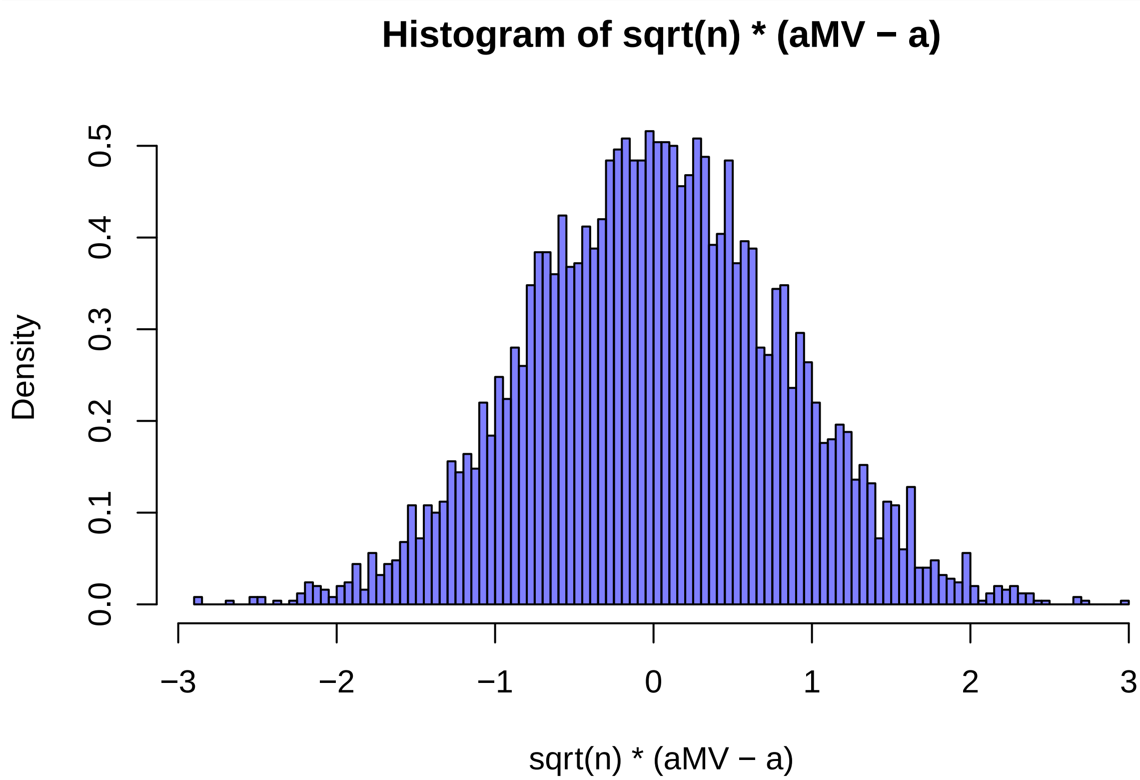
Réponse : on voit nettement que le risque est bien meilleur pour l'estimateur du MV, ce qui confirme notre résultat

**9.** Recopier et relancer le code de la question 7 cette fois-ci pour $a = 1.8$. Que se passe-t-il, et pourquoi ?

```
a = 1.8
n = 10^4
N = 5000

aMM = replicate(N,0)
aMV = replicate(N,0)

for (i in 1:N){
  X = (1/(1-runif(n)))^{1/(a-1)}
  aMM[i] = (2*mean(X)-1)/(mean(X)-1)
  aMV[i] = 1/mean(log(X)) + 1
}

hist(sqrt(n)*(aMM-a),freq=FALSE,breaks=100,
     col = rgb(red = 1, green = 0, blue = 0, alpha = 0.5))
```




```
hist(sqrt(n)*(aMV-a),freq=FALSE,breaks=100,
     col = rgb(red = 0, green = 0, blue = 1, alpha = 0.5))
```



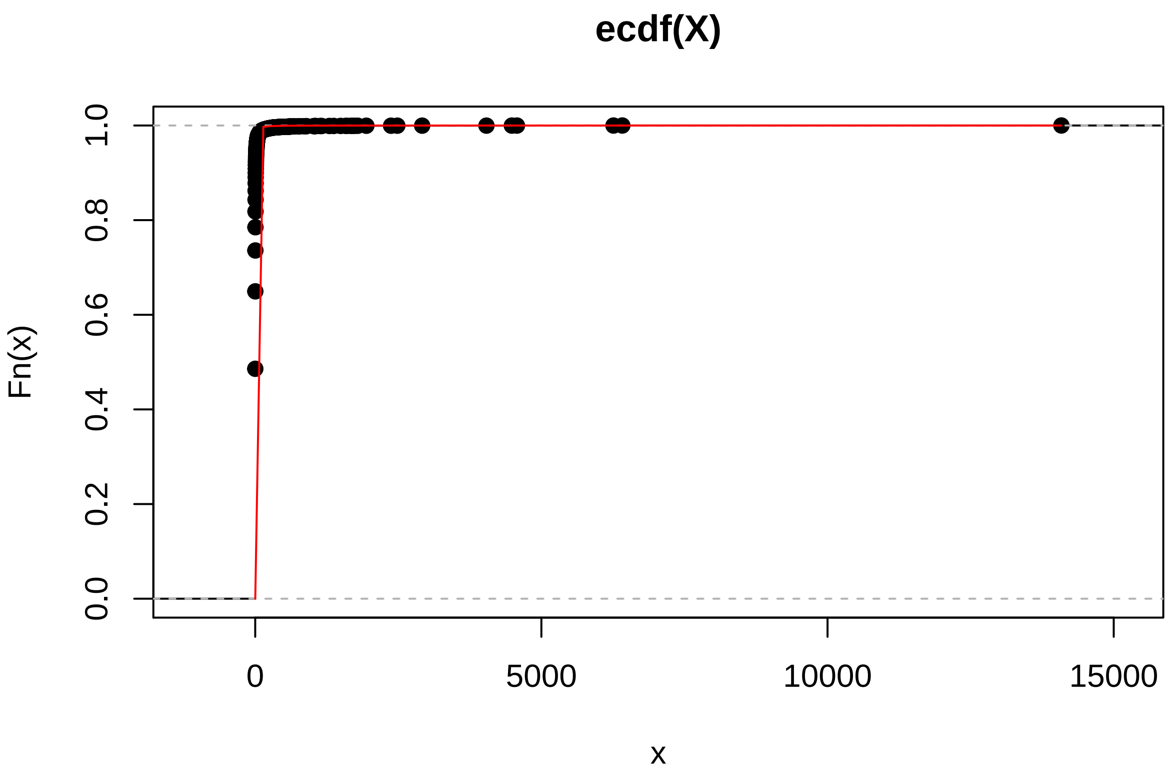

Les deux courbes ne s'affichent pas bien. En regardant de plus près, c'est la rouge qui pose problème : il n'y a pas convergence de l'estimateur des moments et pas de loi limite, car $a < 2$ !

**10.** Un premier argument est naturellement que le risque est plus faible pour l'estimateur du MV, et un deuxième argument est que l'estimateur des moments n'est pas convergent pour toute la gamme de paramètres $a > 1$ !

**Troisième partie : application à des données réelles**

**11.** On voit qu'un mot apparaît plus de 14000 fois, ce qui est beaucoup comparé aux autres. Comme c'est écrit en langue anglaise, il s'agit en fait du mot "and".

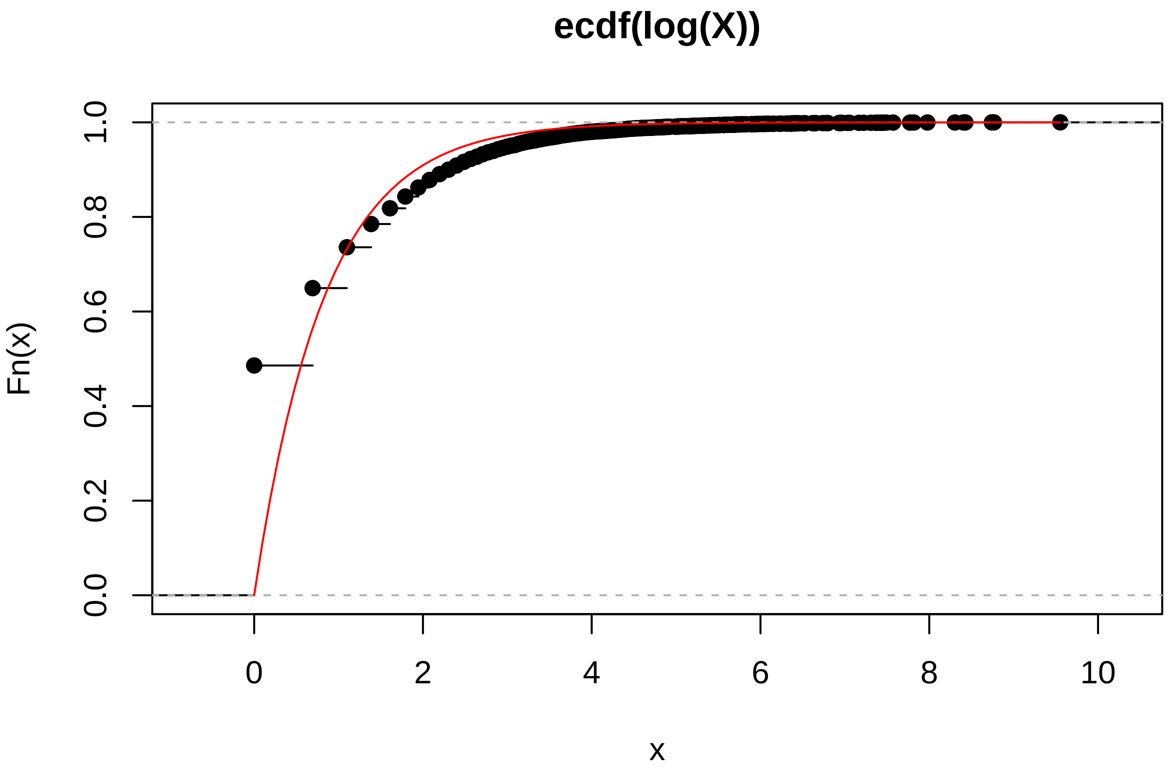
```
data = read.csv("moby_dick.csv")
X = data$occurrences
a = 1/mean(log(X)) + 1
plot(ecdf(X))
curve(1-x^(1.-a),from=1,to=max(X),add=TRUE,col="red")
```




**12.**

**13.** Le graphique ci-dessus n'est pas satisfaisant, la courbe est trop écrasée. Une autre façon de représenter la proximité en loi est d'utiliser le fait que le logarithme d'une variable de loi $PL(a)$ suit une loi $\mathcal{E}(a-1)$. Visualisons cette proximité en loi avec la fonction de répartition du log-échantillon, et celle d'une loi exponentielle.

```
a = 1/mean(log(X)) + 1
plot(ecdf(log(X)))
curve(pexp(x,rate=a-1),from=0,to=log(max(X)),add=TRUE,col="red")
```

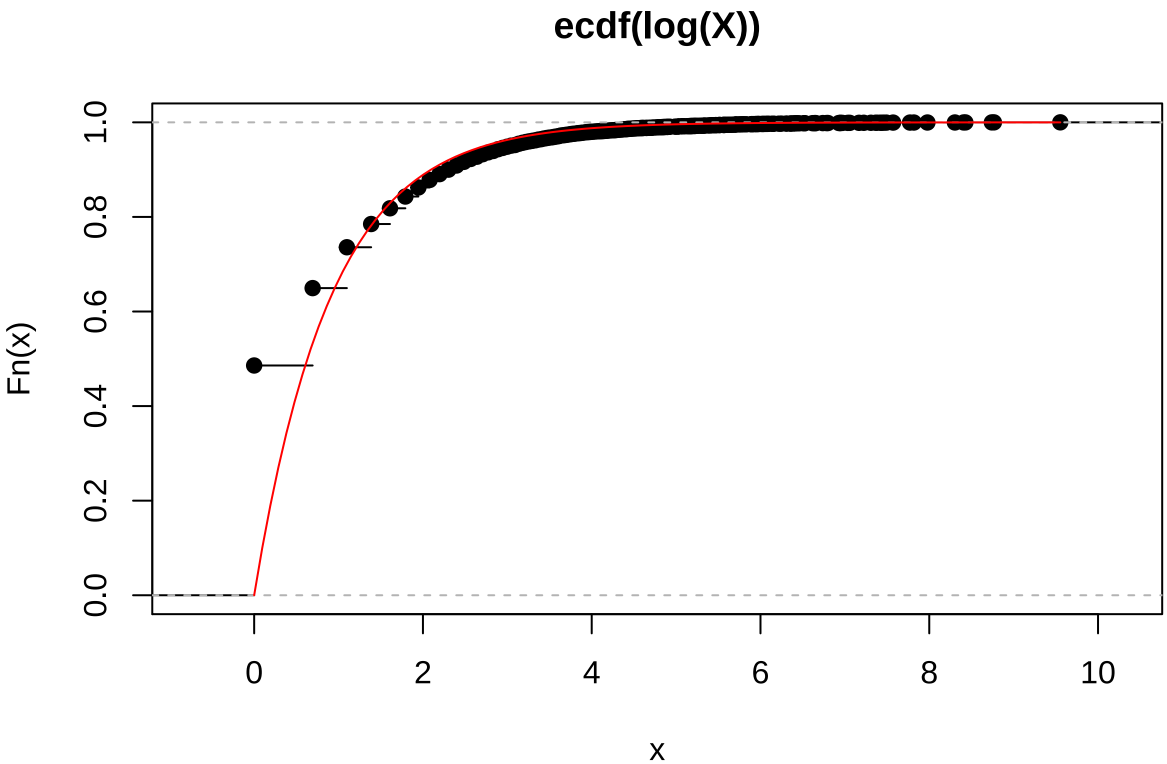



```
a
```

```
## [1] 2.200105
```

**14.** Voici le code :

```
a = (2*mean(X)-1)/(mean(X)-1)
plot(ecdf(log(X)))
curve(pexp(x,rate=a-1),from=0,to=log(max(X)),add=TRUE,col="red")
```




```
a
```

```
## [1] 2.098645
```

Le fit est visuellement meilleur, alors que l'estimateur est théoriquement moins bon ! Cela est du au jeu de données, qui de toute façon ne suit pas vraiment une loi $PL(a)$. Les résultats théoriques servent à aiguiller nos choix, mais face à un modèle par définition limité, et un nombre de données non asymptotique, ils peuvent parfois ne pas être en accord avec la pratique. D'où la phrase suivante :

*"In theory, there is no difference between theory and practice. But in practice, there is."*