DEVOIR ASSURANCE 1

AHMED EL YOUSEFI

1 EXERCICE 1 :

1.1 IMPORTATION DES DONNÉES :

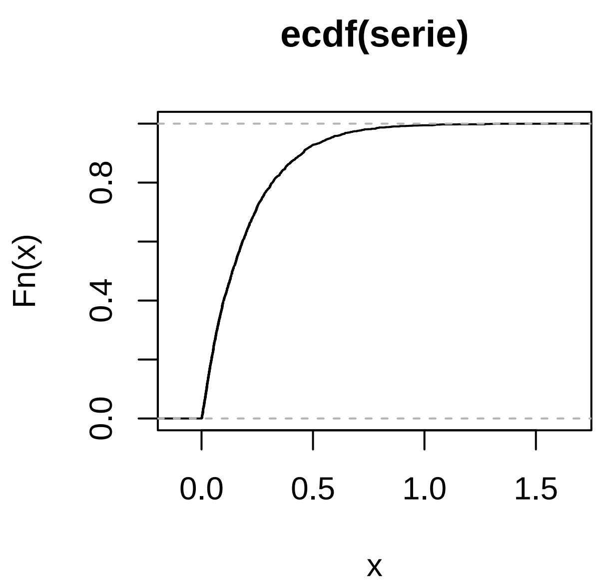
```
setwd("C:/Users/DELL/Documents/R STUDIO/DEVOIR ASSURANCE")
table <- read.csv("montantssinistre16.csv",sep=";",dec=",")
serie <- table[,2]
head(table)
```

```
##   X          x
## 1 1 0.07675949
## 2 2 0.04085351
## 3 3 0.17902609
## 4 4 0.06165967
## 5 5 0.11395614
## 6 6 0.22183424
```

1.2 ANALYSE EXPLORATOIRE :

Faisons le plotage de la fonction de repartition empirique de l'echantillon :

```
plot(ecdf(serie))
```

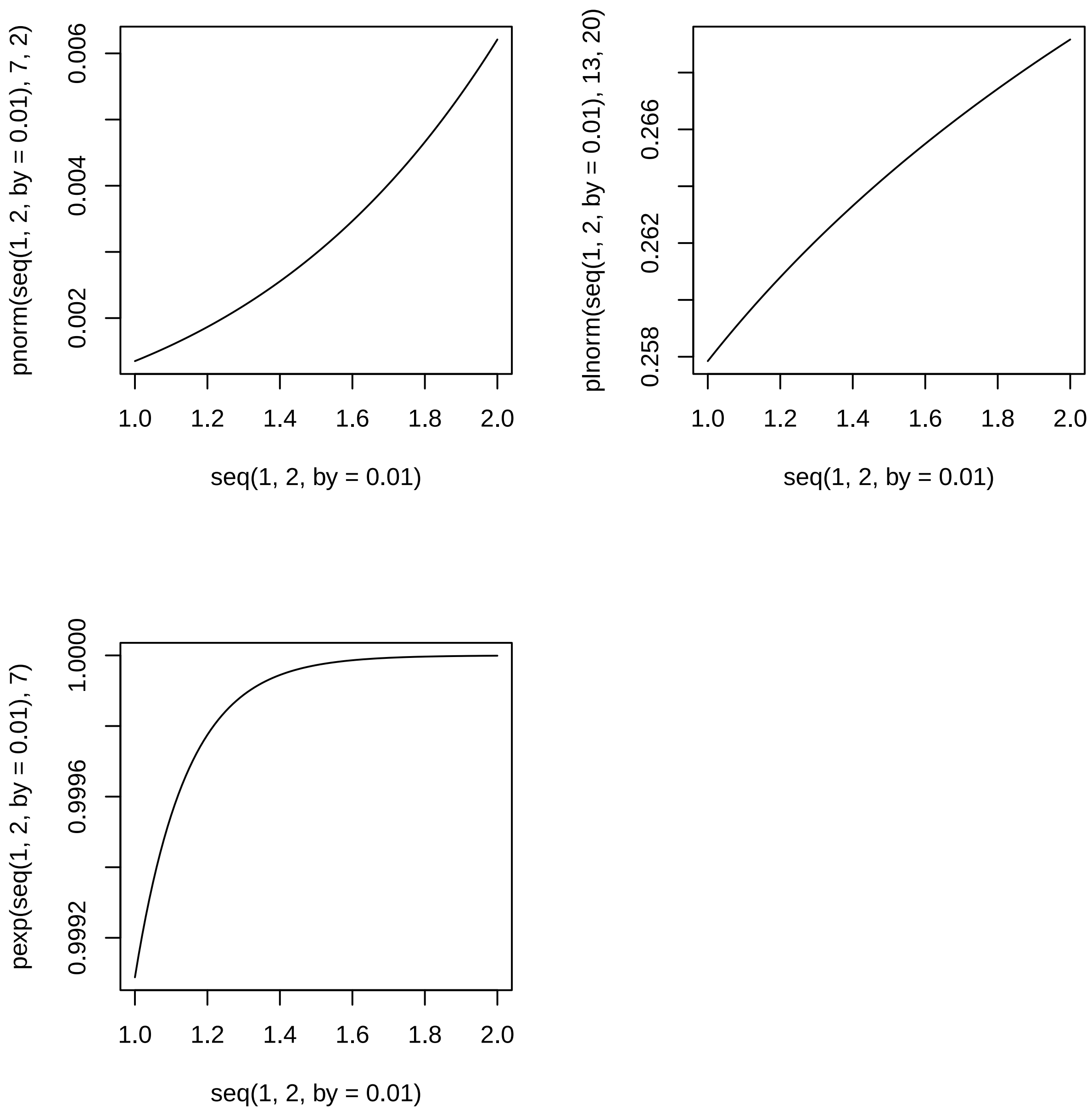


Pour decider une distribution adequate on pourra comparer visualser les fonctions de repartition des trois modeles respectivement de gauche à droite :

- normal
- log normal
- exponentiel

```
par(mfrow=c(2,2))

plot(seq(1,2,by=0.01),pnorm(seq(1,2,by=0.01),7,2),type="l")
plot(seq(1,2,by=0.01),plnorm(seq(1,2,by=0.01),13,20),type="l")
plot(seq(1,2,by=0.01),pexp(seq(1,2,by=0.01),7),type="l")
```



Il est claire que le modele *exponentiel* est le plus adequate (parmis les trois modeles) en fait :

- Le modele normale a une fonction d repartition convexe, ce qui n'est pas le cas pour la fonction de repartition empirique.
- Le modele log-normale a une fonction d repartition faiblement concave, ce qui n'est pas le cas pour la fonction de repartition empirique.

1.3 MODELE PARAMETRIQUE :

En suite on supposera que la variable : M = montant des sinistre suite loi exponentielle, c-à-d :

$$M \sim (\mathfrak{E}(\lambda))_{\lambda>0}$$

NB : il sera aussi possible d'utiliser les modeles : gamma et weibull (le modele exponentiel n'est qu'une cas particulier de ces deux modeles), mais on préfere d'utilisr le modele exponentiele pour sa simplicité.

1.3.1 Estimation de parametre :

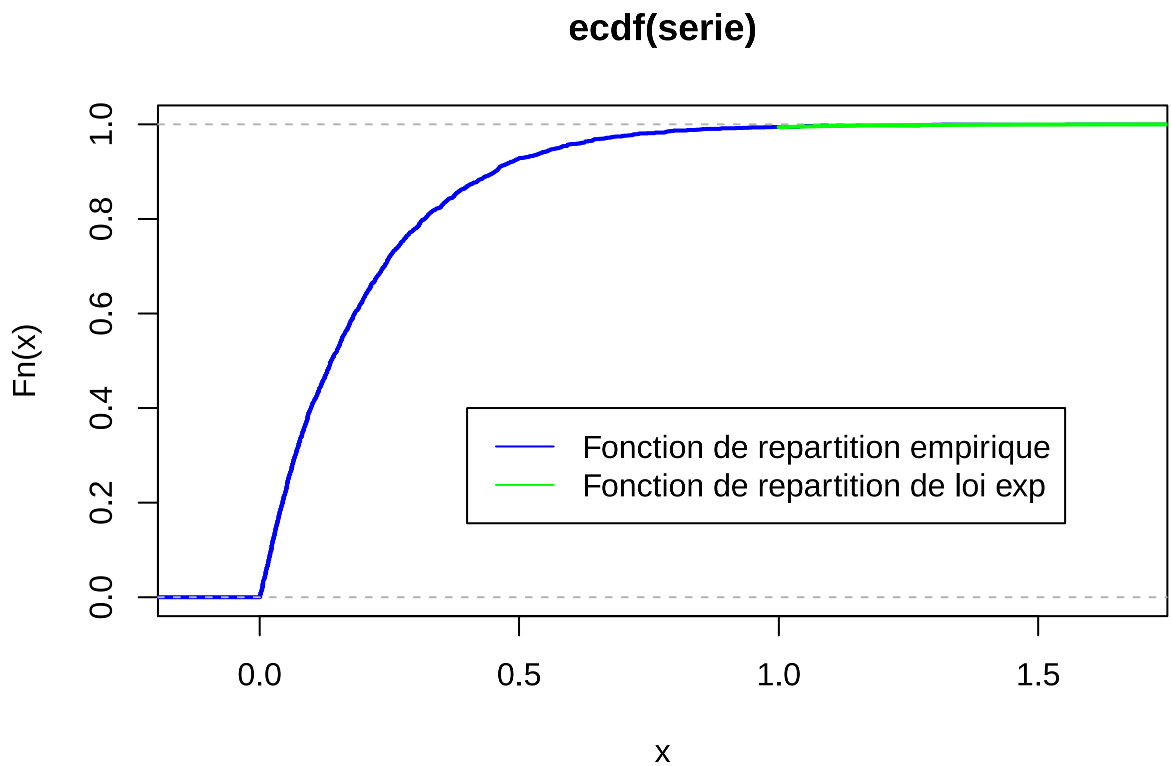
```
library(MASS)
model <- fitdistr(serie , "exponential")
rate <- model$estimate
rate
```

```
##     rate
## 5.116635
```

1.4 EVALUATION DE MODELE :

1.4.1 Evaluation visuelle :

```
plot(ecdf(serie),col="blue",lwd=2)
lines(seq(1,2,by=0.01),pexp(seq(1,2,by=0.01),rate=rate),type="l",col="green",lwd=2)
legend(x=0.4,y=0.4,legend = c("Fonction de repartition empirique","Fonction de repartition de loi exp")
```



Cette graphique montre la quasi-egalité entre les deux foncitons de repartitions, ce qui montre que ce choix etait adequate.

1.4.2 Evaluation statistique :

Pour s'assurer on ferra le teste de *Kolmogrov* :

```
ks.test(serie , pexp,rate)
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  serie
## D = 0.012832, p-value = 0.8433
## alternative hypothesis: two-sided
```

1.5 INTERPRETATION :

On obtient un p-value > 5%, donc on décide d'accepter l'hypothese H_0 avec un risque de 16%. cela signifie que Le modèle proposée est validé statistiquement.

2 EXERCICE 2 :

2.1 IMPORTATION DES DONNÉES :

```
setwd("C:/Users/DELL/Documents/R STUDIO/DEVOIR ASSURANCE")
table <- read.csv("nombresinistre16.csv",sep=";",dec=",")
head(table)
```

```
##   X x
## 1 1 3
## 2 2 1
## 3 3 2
## 4 4 1
## 5 5 4
## 6 6 0
```

```
serie <- table[,2]
summary(serie)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    1.00    2.00    2.02    3.00    9.00
```

2.2 ANALYSE EXPLORATOIRE :

```
glue::glue_collapse(
  c(
    "MOYENNE  EMPIRIQUE : " ,
   mean(serie), "\n" ,
   "VARIANCE EMPIRIQUE : " ,
   var(serie)                 )
)
```

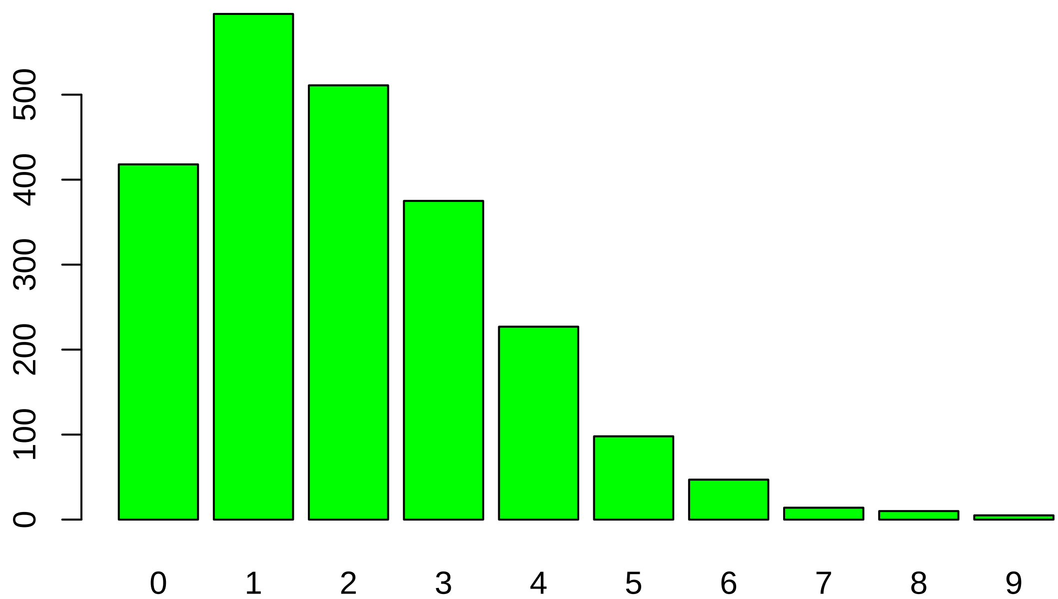
```
## MOYENNE  EMPIRIQUE : 2.0195652173913
## VARIANCE EMPIRIQUE : 2.66990846681922
```

On remarque bien que la moyenne empirique est inferieur à variance empirique, ce qui permet de conclure :

$$\mathbb{E}(N) < \mathbb{V}(N)$$

Cela nous encourage à utiliser le modèle **binomiale négative** pour modeliser la variable : X = **nombre sinistres**. On peut voir aussi l'histogramme de l'echantillon fourni dans la base des données :

```
effectives <- table(serie)
barplot(effectives,col="green")
```



2.3 MODELE PARAMETRIQUE :

En suite on supposera que la variable : N = nombre des sinistre suite une modèle binomiale négative, c-à-d :

$$N \sim (\mathfrak{B}_{\mathfrak{N}}(n,p))_{n\in\mathbb{N}\ ,\ p\in[0,1]}$$

2.3.1 Estimation de parametre :

```
library(vcd)
```

```
## Le chargement a nécessité le package : grid
```

```
model <- goodfit(serie , "nbinom")
model$par
```

```
## $size
## [1] 6.016545
##
## $prob
## [1] 0.7486887
```

```
size <- as.integer(model$par$size)
prob <- model$par$prob
```

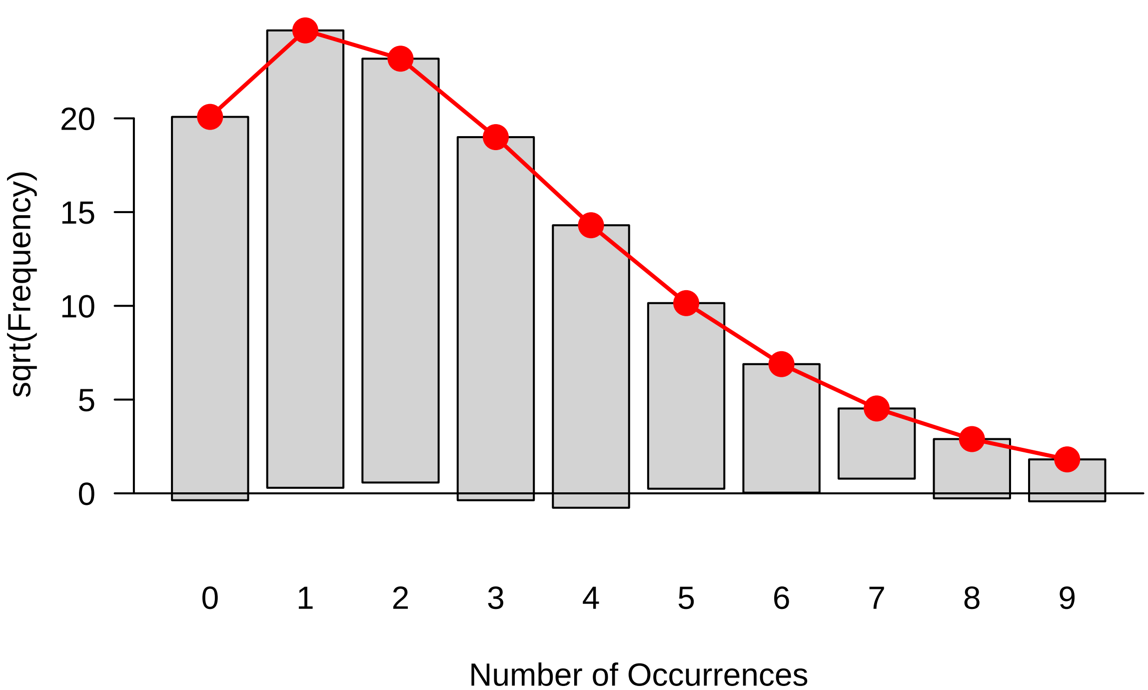
L'estimateur de maximum de vraiesemblance (défini en **R**) montre que :

$$X \sim \mathfrak{B}_{\mathfrak{N}}(6, 0.74)$$

2.4 EVALUATION DE MODELE :

2.4.1 evaluation visuelle :

```
plot(model)
```



Le graphique rouge represente la racine carrée de l'effective des sinistres espérée (en théorie) par le modele binomiale, en outre le **barplot** en gris represente la racine carrée de l'effective experimenté des sinistres.

2.4.2 evaluation statistique :

Visualisons d'abord les valeurs observée vs les valeurs esperée par le modele :

```
library(data.table)
data<-data.table(model$count , model$observed , model$fitted)
data
```

```
##     V1  V2         V3
##  1:  0 418 403.139879
##  2:  1 595 609.557911
##  3:  2 511 537.428060
##  4:  3 375 360.909545
##  5:  4 227 204.451633
##  6:  5  98 102.932040
##  7:  6  47  47.495974
##  8:  7  14  20.490400
##  9:  8  10   8.378538
## 10:  9   5   3.279282
```

Regroupons d'abord les valeurs esperée par le modele qui sont < 5, ici on regroupe les classes 8 et 9.

```
data[9,2:3]<-data[9,2:3]+data[10,2:3]
data<-data[1:9,]
data
```

```
##    V1  V2        V3
## 1:  0 418 403.13988
## 2:  1 595 609.55791
## 3:  2 511 537.42806
## 4:  3 375 360.90954
## 5:  4 227 204.45163
## 6:  5  98 102.93204
## 7:  6  47  47.49597
## 8:  7  14  20.49040
## 9:  8  15  11.65782
```

Effectuant maintenant le test statistique :

```
data<-as.matrix(data[,3])
n<-length(data)

proba <- sapply(1:(n-1), function(n){dnbinom(n,size = size,prob = prob)})
proba <- c(proba , 1-sum(proba))
# The statistical test    :
stat<-chisq.test(data,proba)$statistic
df<-n-2-1
pvalue<-1-pchisq(stat,df)
names(pvalue)<-c("p-value")
pvalue
```

```
##      p-value
## 1.588729e-13
```

```
chisq.test(data,proba)
```

```
##
##  Pearson's Chi-squared test
##
## data:  data and proba
## X-squared = 72, df = 64, p-value = 0.2303
```

2.5 INTERPRETATION :

On voit bien que le p-value $< 5\%$, donc on rejette l'hypothese H_0. Ce qui signifie que le modèle **binomiale négative** n'est pas valide statistiquement.