Modélisation des Sinistres IARD

Ahmed EL YOUSEFI

IMPORTATION DES DONNÉES :

```
setwd("C:/Users/DELL/Desktop/IARD")
table <- read.csv("nombresinistre16.csv",sep=";",dec=",")
serie <- table[,2]
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)

ANALYSE EXPLORATOIRE :

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

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

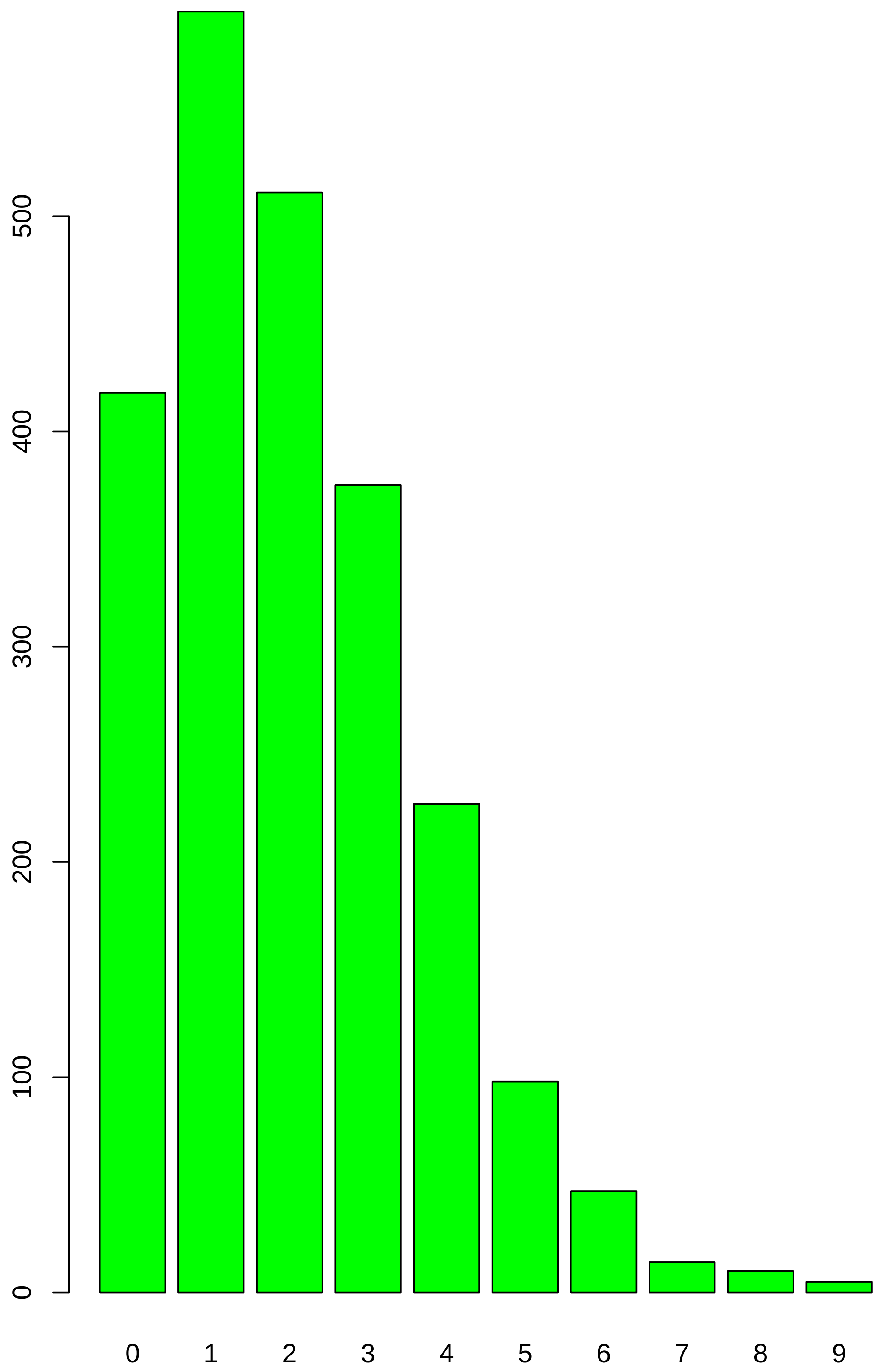
```
## MOYENNE EMPIRIQUE : 2.0195652173913
## VARIANCE EMPIRIQUE : 2.66990846681922
```

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 :

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


0
100
200
300
400
500
0
1
2
3
4
5
6
7
8
9

MODELE PARAMETRIQUE :

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

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

Estimation de parametre :

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

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

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

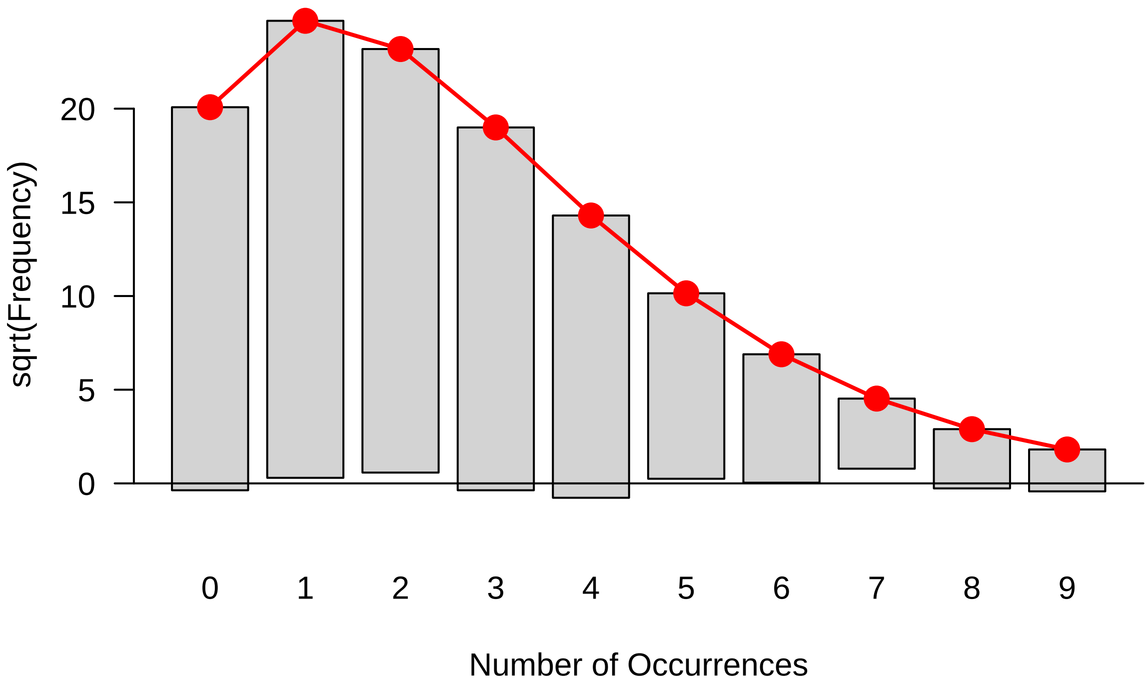
L'estimateur de maximum de vraiesemblance (défini en R) montre que :

$$N \sim \mathcal{B}_N(n = 6, p = 0.74)$$

EVALUATION DE MODELE :

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.

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.

```r
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 :

```r
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
```

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

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

Interpretation :

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