

TP Statistiques 1

Génération et sauvegarde de données

On génère un échantillon de 40 observations pour chaque loi.

```
# création des données
r <- data.frame(
  rnorm(40,0,1),
  runif(40,0,1),
  rpois(40,17),
  rexp(40, rate = 1),
  rchisq(40, df = 1.3),
  rbinom(40, 100,0.5),
  rcauchy(40, location = 0, scale = 1)
)

# noms des colonnes
colnames(r) <- c("gaussienne", "uniforme0-1", "poisson, p=17",
  "exp", "chisq (1.3)", "binomiale (40,100,0.5)",
  "cauchy"
)

# écriture du csv, on ne met pas de colonne avec le numéro de la ligne
write.csv(r, file = "data2.csv", row.names = FALSE)
```

```
# relecture
file <- read.csv("data2.csv", header = TRUE)
# voir la valeur de la colonne gaussienne
file$gaussienne
```

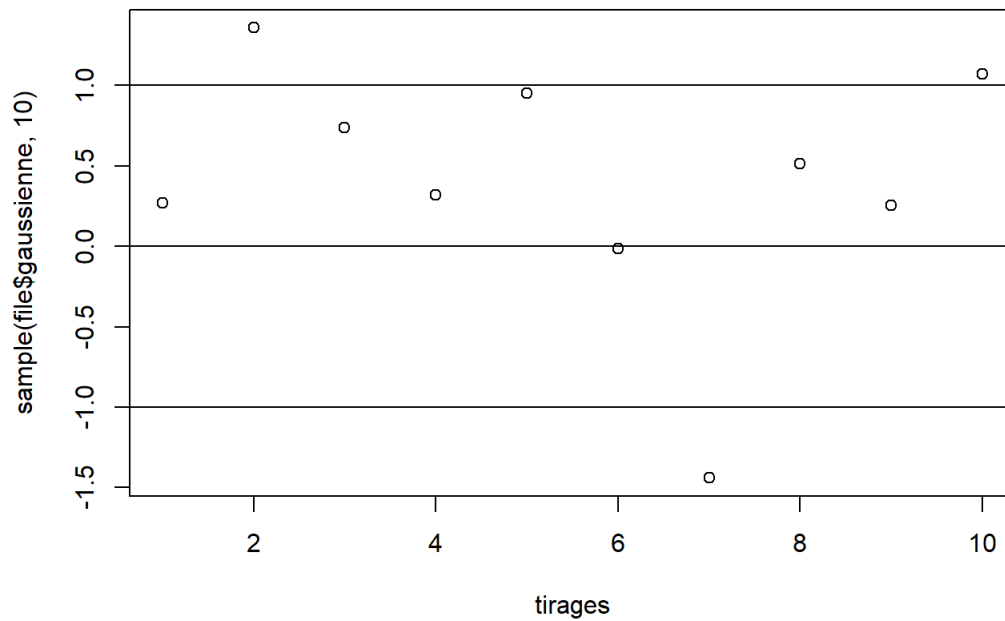
```
## [1] 0.95220706 0.50842435 -0.51483952 0.80485512 0.25514654 1.39911267
## [7] 0.51377897 -1.78506048 -0.43897595 0.90013776 -1.28727059 -0.13700698
## [13] -0.97368986 1.54660584 -1.20982653 -1.43961136 -0.01320967 0.32107157
## [19] 1.35805066 0.25526230 0.59375768 0.73817853 -0.18212854 -0.88836339
## [25] 0.93218566 -1.31795788 1.02849176 -0.08952164 0.27103385 1.20185826
## [31] 0.20887195 0.75254625 0.87752373 1.06937598 -0.81294250 -0.23397687
## [37] 1.30341865 -0.82903639 -0.42670348 1.05012147
```

Tracer les données

On récupère et on affiche 10 valeurs.

```
plot(sample(file$gaussienne, 10), main = "Gaussienne 0-1 (10 tirages)",
  xlab = "tirages")
abline(h=0) # trace une ligne en y=0
abline(h=-1) # trace une ligne en y=-1
abline(h=1) # trace une ligne en y=1
abline(h=2) # trace une ligne en y=2
abline(h=-2) # trace une ligne en y=-2
abline(h=3) # trace une ligne en y=3
abline(h=-3) # trace une ligne en y=-3
```

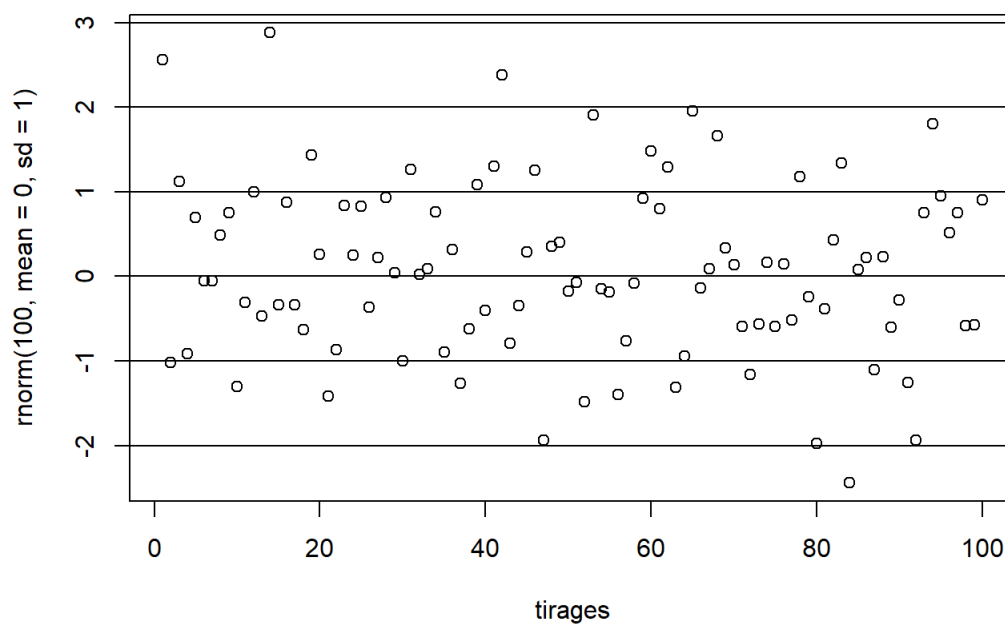
Gaussienne 0-1 (10 tirages)



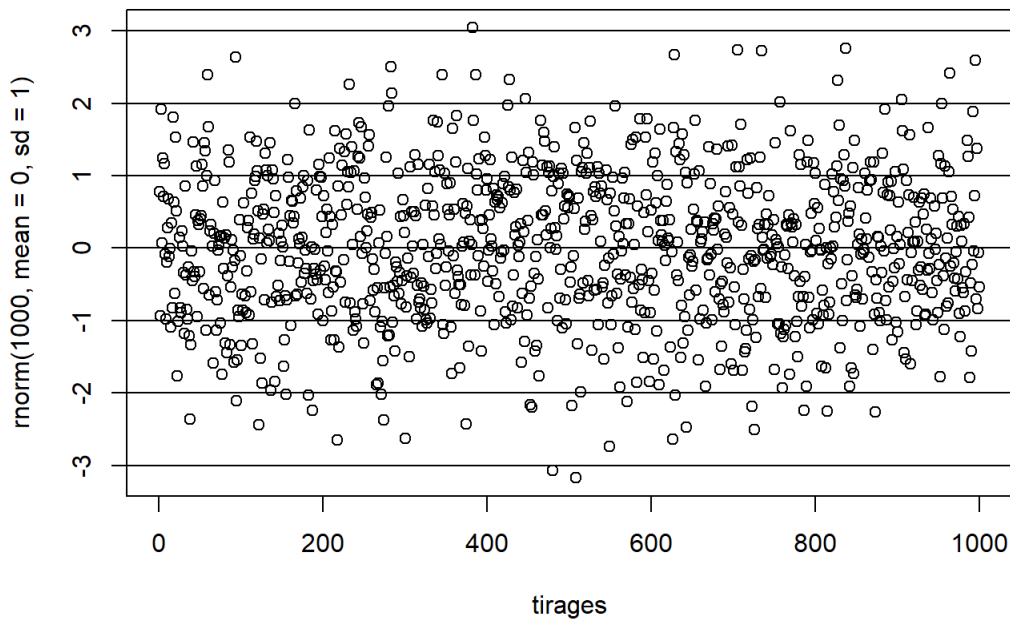
On observe que nous avons 8

valeurs entre -1 et 1 par exemple dans un cas ou de façon générale il y a souvent beaucoup de valeurs autour de 0.

Gaussienne 0-1 (100 tirages)



Gaussienne 0-1 (100 tirages)



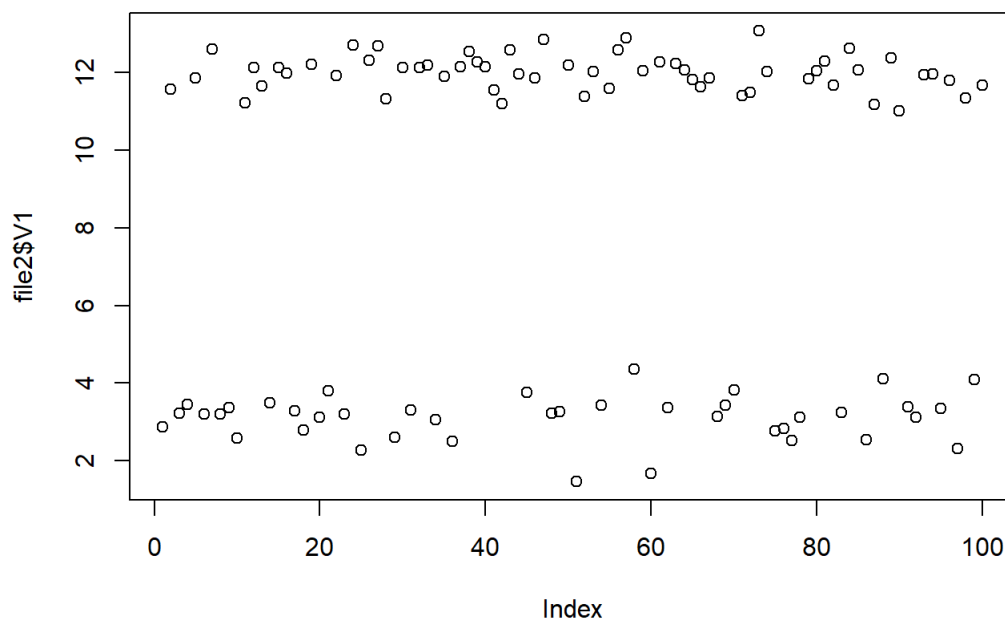
En augmentant le nombre de

tirages, notre hypothèse précédente semble être vraie.

Est-ce que vous pouvez conclure quelque chose sur cette distribution, à partir d'une visualisation ?

```
file2 <- read.csv("distribution_inconnue.csv", header = TRUE)

plot(file2$V1);
```

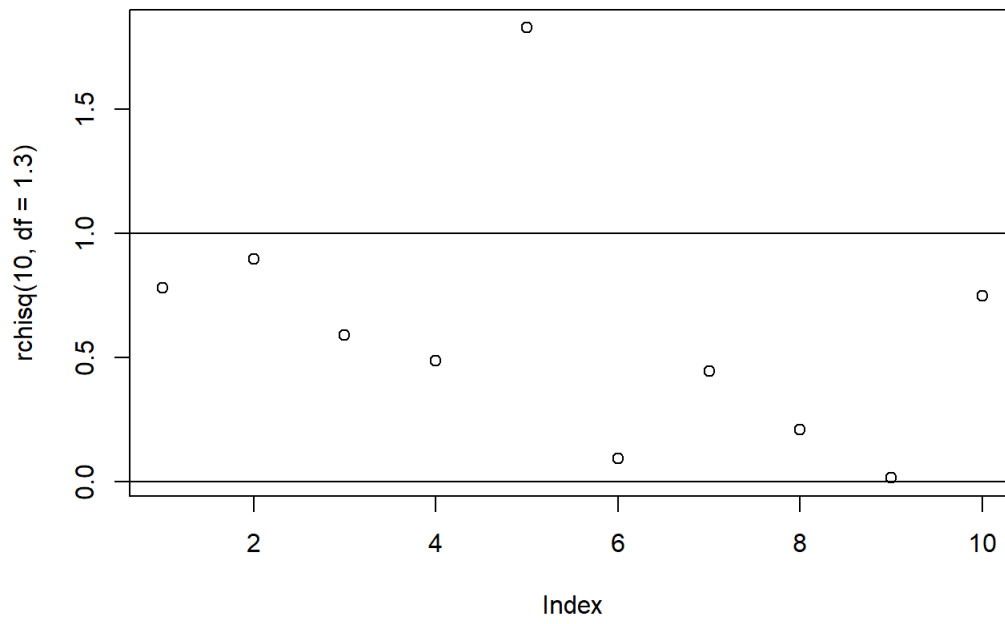


On voit une accumulation

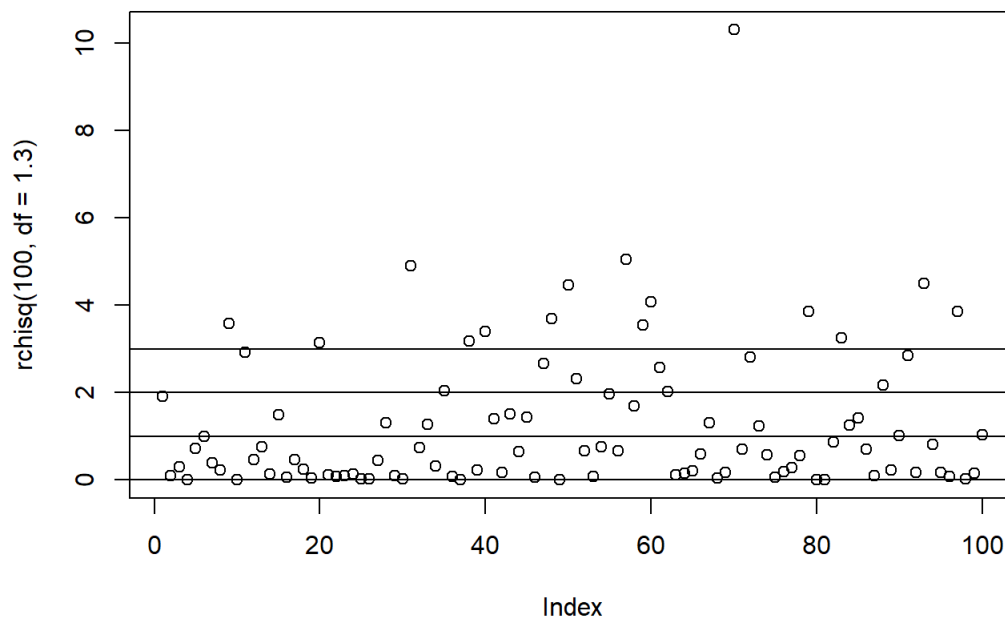
autour de 12 et 3 donc peut être deux lois gaussiennes avec une moyenne de 12 et 3 respectivement. On peut supposer avec l'échelle qu'il y a eu 100 tirages.

On remarque un comportement potentiellement similaire avec Chisq avec presque tous les points étant proches et plus grand ou égal à 0 ici.

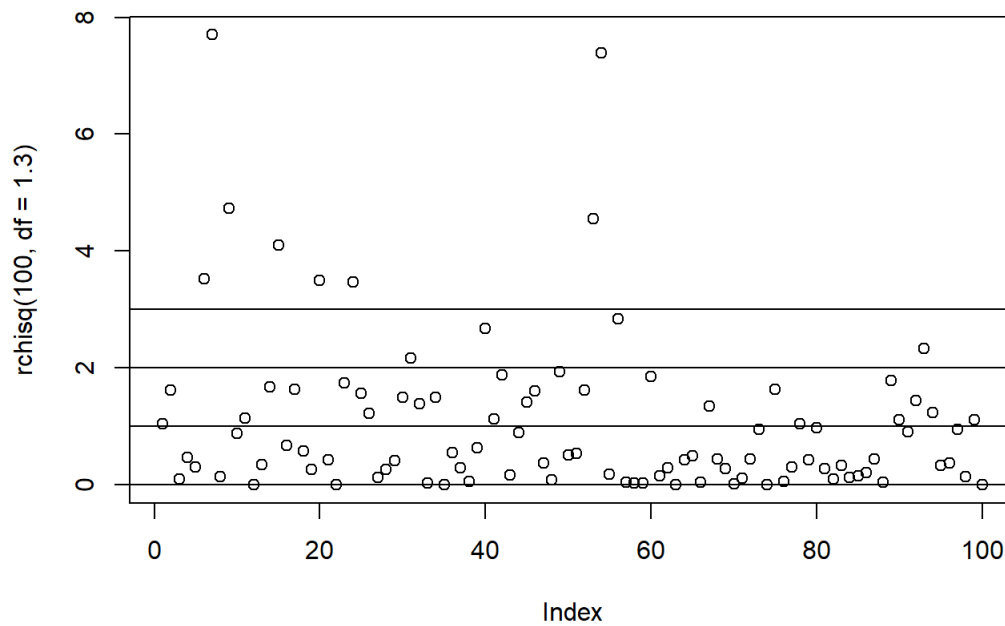
Chisq 1.3 (10 tirages)



Chisq 1.3 (100 tirages)



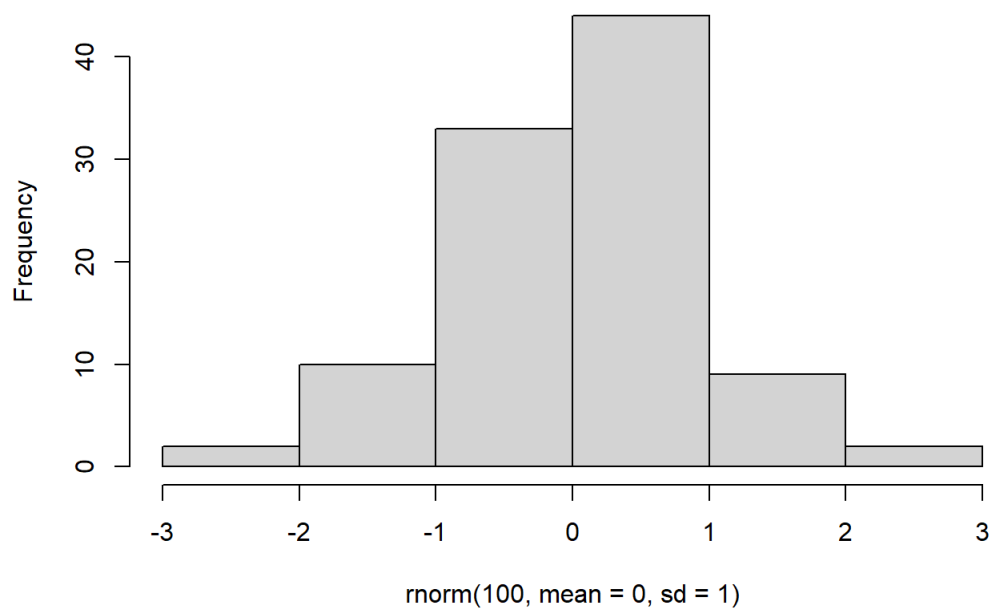
Chisq 1.3 (1000 tirages)



Histogrammes

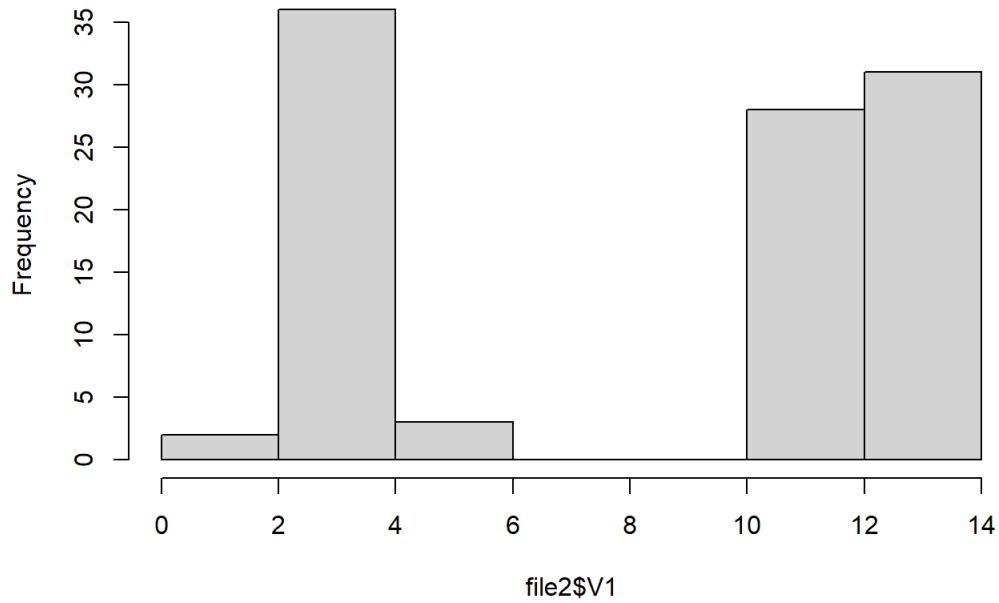
```
hist(rnorm(100, mean = 0, sd = 1),  
     main = "Histogramme gaussienne 0-1 (100 tirages)");
```

Histogramme gaussienne 0-1 (100 tirages)



```
file2 <- read.csv("distribution_inconnue.csv", header = TRUE)  
hist(file2$V1, main = "Histogramme de la distribution inconnue");
```

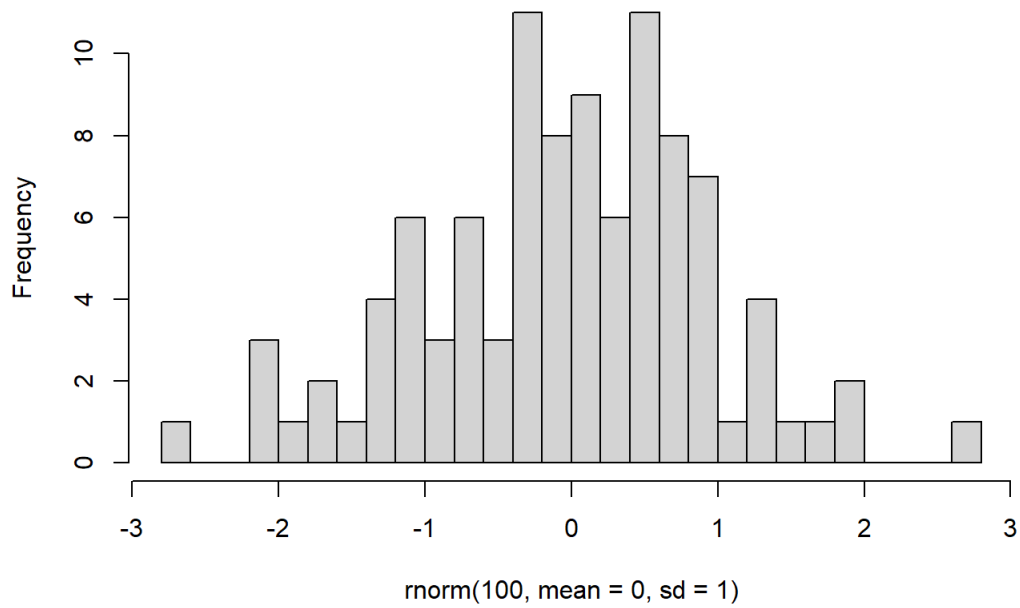
Histogramme de la distribution inconnue



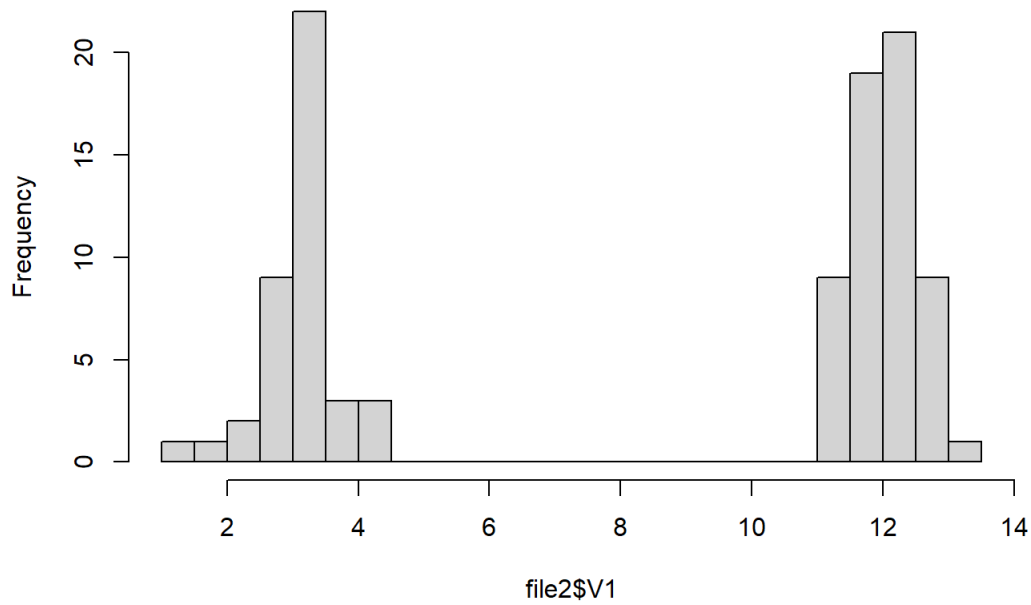
On observe qu'il est possible

que la distribution soit le résultat de trois et non deux lois gaussiennes car il y a deux pics vers 12.

Histogramme gaussienne 0-1 (100 tirages)



Histogramme de la distribution inconnue

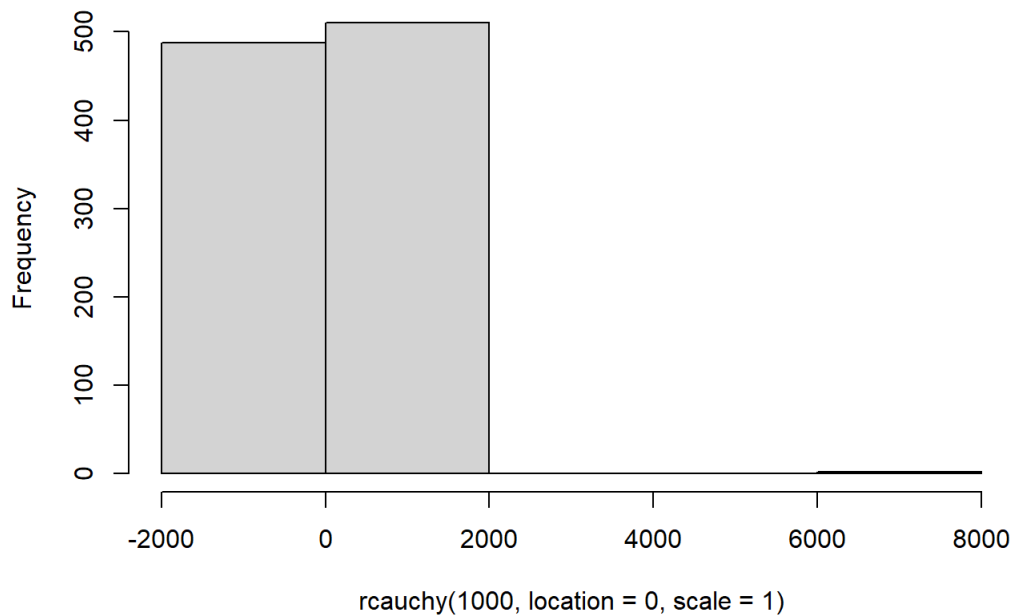


En testant avec Cauchy, on

observe une accumulation autour de la location mais très peu d'écart autour.

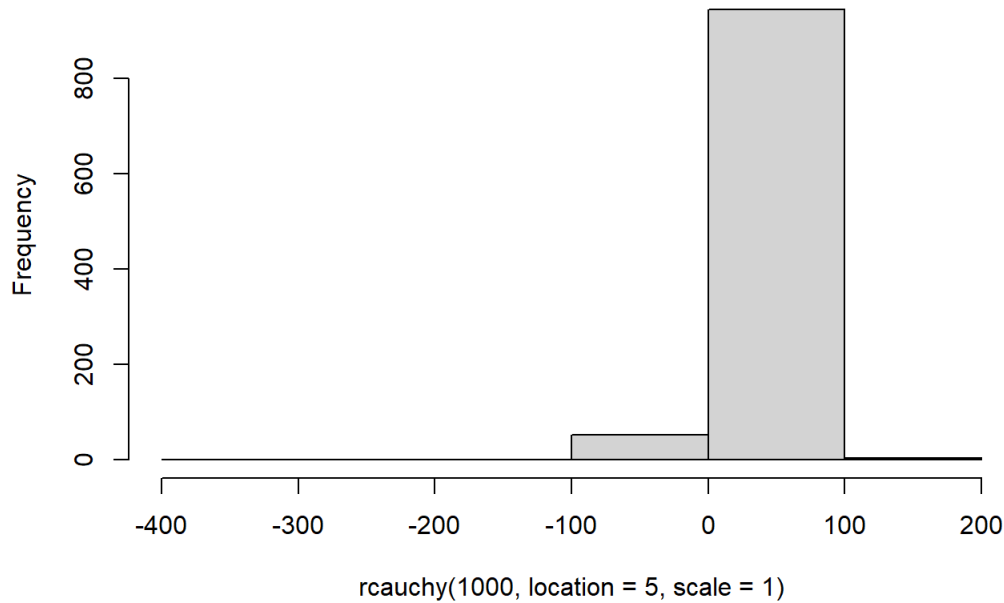
```
hist(rcauchy(1000, location = 0, scale = 1),  
     main = "Histogramme Cauchy",  
     breaks = 5);
```

Histogramme Cauchy



```
hist(rcauchy(1000, location = 5, scale = 1),  
     main = "Histogramme Cauchy",  
     breaks = 5);
```

Histogramme Cauchy



Moments

```
require("e1071") # ou moments
file <- read.csv("data2.csv", header = TRUE)
```

```
m <- t(sapply(file, function(x){
  return(c(mean(x), var(x), skewness(x), kurtosis(x)))
}))
colnames(m) <- c("Moyenne", "Variance", "Skewness", "Kurtosis")
rownames(m) <- c("gaussienne", "uniforme0-1", "poisson, p=17",
"exp", "chisq (1.3)", "binomiale (40,100,0.5)",
"cauchy")
m # print
```

##	Moyenne	Variance	Skewness	Kurtosis
## gaussienne	0.1565474	0.81944275	-0.39489612	-1.016887
## uniforme0-1	0.5392344	0.08082616	-0.14943771	-1.202270
## poisson, p=17	17.4750000	18.92243590	1.22493572	2.597900
## exp	1.2583137	1.38727280	1.94972905	4.594009
## chisq (1.3)	1.1557144	2.30272156	2.19546730	5.259348
## binomiale (40,100,0.5)	49.6000000	20.24615385	0.06698205	-1.045440
## cauchy	-3.8779800	174.58626835	-4.33988690	19.648296

Quantiles

```
q <- t(sapply(file, function(x){
  return(quantile(x, probs = c(0.25,0.5,0.75)))
}))
colnames(q) <- c("Q1", "Q2", "Q3")
cbind(m, q)
```


##	Moyenne	Variance	Skewness	Kurtosis	Q1
## gaussienne	0.1565474	0.81944275	-0.39489612	-1.016887	-0.4579418
## uniforme0-1	0.5392344	0.08082616	-0.14943771	-1.202270	0.3376546
## poisson, p=17	17.4750000	18.92243590	1.22493572	2.597900	14.7500000
## exp	1.2583137	1.38727280	1.94972905	4.594009	0.4698367
## chisq (1.3)	1.1557144	2.30272156	2.19546730	5.259348	0.2604391
## binomiale (40,100,0.5)	49.6000000	20.24615385	0.06698205	-1.045440	46.0000000
## cauchy	-3.8779800	174.58626835	-4.33988690	19.648296	-1.7335137
##	Q2	Q3			
## gaussienne	0.2631481	0.90814973			
## uniforme0-1	0.5700667	0.76493244			
## poisson, p=17	17.0000000	19.25000000			
## exp	0.9450251	1.69046121			
## chisq (1.3)	0.5162036	1.30072500			
## binomiale (40,100,0.5)	49.0000000	54.00000000			
## cauchy	-0.6606289	0.09835571			