TP1 Statistiques

Gautier Poursin

21 février 2020

Gestion et sauvegarde des données

```
matrice.proba<-data.frame(gaussienne=rnorm(40,mean=0,sd=1),
                          uniforme=runif(40,min=0,max=1),
                          poisson=rpois(40,0.5),
                          exponentielle=rexp(40,rate=1),
                          chi2=rchisq(n=40,df=0.5),
                          binomiale=rbinom(n=40,size=50,prob=0.2),
                          cauchy=rcauchy(n=40,location=0,scale=1))
cauchy<-c(cauchy100=rcauchy(n=100,location=0,scale=1))
cauchy200<-c(cauchy200=rcauchy(n=200, location=0,scale=1))
cauchy300<-c(cauchy300=rcauchy(n=300,location=0,scale=1))


Histcauchy<-write.csv(cauchy,"~/Travail/ecole ingé/Stats/TP1/cauchy")
Matriceproba<-write.csv(matrice.proba,"~/Travail/ecole ingé/Stats/TP1/Matriceproba")
read.csv("~/Travail/ecole ingé/Stats/TP1/Matriceproba",header=TRUE)
```

```
##      X  gaussienne  uniforme poisson exponentielle         chi2 binomiale
## 1    1 -0.47202053 0.5625520       1   0.128593884 3.077224e-04        15
## 2    2 -1.27491147 0.5233676       0   0.598156305 2.435749e+00         7
## 3    3 -0.79629874 0.4223858       0   0.582308347 1.192019e+00        15
## 4    4 -1.47662022 0.9169353       0   1.001450230 1.201149e-01        12
## 5    5 -0.43938646 0.8723673       1   0.956320777 2.887121e-02         8
## 6    6 -0.79271369 0.6200776       0   2.005696183 1.080576e-03        12
## 7    7  0.79362798 0.7224407       1   0.604719904 1.247140e-01         9
## 8    8  0.88243639 0.1415459       0   0.236525425 1.266194e+00         7
## 9    9  2.75201480 0.5661959       0   0.293602759 2.325220e-01        10
## 10  10 -0.13378472 0.7122721       1   0.592320082 1.607922e+00        12
## 11  11  2.05436047 0.7887771       0   1.526894589 7.272668e-04         9
## 12  12 -1.80319548 0.9262013       1   0.617370637 1.138029e-01        10
## 13  13 -0.75537398 0.7774180       2   0.992904099 2.085376e-02         9
## 14  14  0.18110753 0.8566386       2   0.109017339 2.686040e-03        17
## 15  15  0.17591310 0.8068117       0   0.252398851 5.990565e-01        11
## 16  16  0.87851541 0.1819602       0   3.152212195 3.269042e-04         4
## 17  17  0.60019511 0.1335568       2   0.908493171 6.247583e-02         7
## 18  18 -0.06611213 0.5210418       1   0.183769283 2.412032e-02        10
## 19  19 -0.11471758 0.2918336       0   1.039422798 5.741698e-03        13
## 20  20 -1.42011782 0.1211079       0   0.295024509 2.573844e-03        11
## 21  21 -0.07186559 0.3677454       2   0.596386776 1.616157e-05        12
```

```
## 22 22 -0.36329379 0.4113659       0   0.045407405 6.310647e-03         10
## 23 23 -0.59007633 0.0598773       1   1.327168145 8.430020e-02         12
## 24 24 -0.96822665 0.2784146       0   0.172536660 1.037585e-01         12
## 25 25 -0.28556936 0.2512820       1   0.071249888 6.046250e-01          9
## 26 26 -1.40346927 0.3934696       0   0.096670624 1.673369e-04          9
## 27 27 -0.55536869 0.6139431       0   0.178824400 1.607949e-01          6
## 28 28 -1.99280319 0.3002951       2   0.965224857 7.547524e-01          7
## 29 29 -1.55985396 0.6373032       0   0.009056801 2.822647e-08         10
## 30 30 -0.45228947 0.3715850       0   1.958794296 7.596464e-01         10
## 31 31 -1.92564734 0.2369936       0   0.850679363 1.165344e-01         14
## 32 32  0.20052560 0.7925894       1   0.567476015 6.418325e-02          4
## 33 33  0.70920244 0.6396454       1   0.117843384 6.437967e-01          6
## 34 34  0.19432771 0.4283157       2   0.028866341 7.262553e-01         12
## 35 35 -0.28730616 0.1455974       0   0.248576466 1.226151e+00         13
## 36 36  0.30812019 0.1336397       0   0.092511149 8.532449e-01         13
## 37 37 -0.47421443 0.4073324       0   0.276787272 3.730648e-01          8
## 38 38  0.39405694 0.7433177       1   0.570917765 1.808536e-01         10
## 39 39 -1.01073249 0.5122827       1   0.495021928 1.717593e-05         10
## 40 40  0.46156782 0.5629021       2   0.720661474 3.140845e-03          9
##          cauchy
## 1    2.29513969
## 2   -1.94062450
## 3   -3.20182205
## 4   -0.41653866
## 5   -0.11845112
## 6    1.62717976
## 7   -0.68692093
## 8   -1.88987523
## 9    6.85118615
## 10   0.97342374
## 11   0.60213538
## 12   0.04158684
## 13  -0.32308494
## 14  -0.68446447
## 15  -0.87517703
## 16   8.04121704
## 17  -0.20133661
## 18   1.17178874
## 19  -1.53963861
## 20  -1.25202284
## 21  -0.58358107
## 22 -13.33820033
## 23  -3.35220064
## 24  -6.42507297
## 25   1.27491862
## 26   0.48552270
## 27  -0.25172771
## 28  -0.53902598
## 29  -1.17545948
## 30   1.82409168
## 31 -24.70660985
## 32  -2.16512032
## 33  61.85240530
## 34  -0.55695631
```

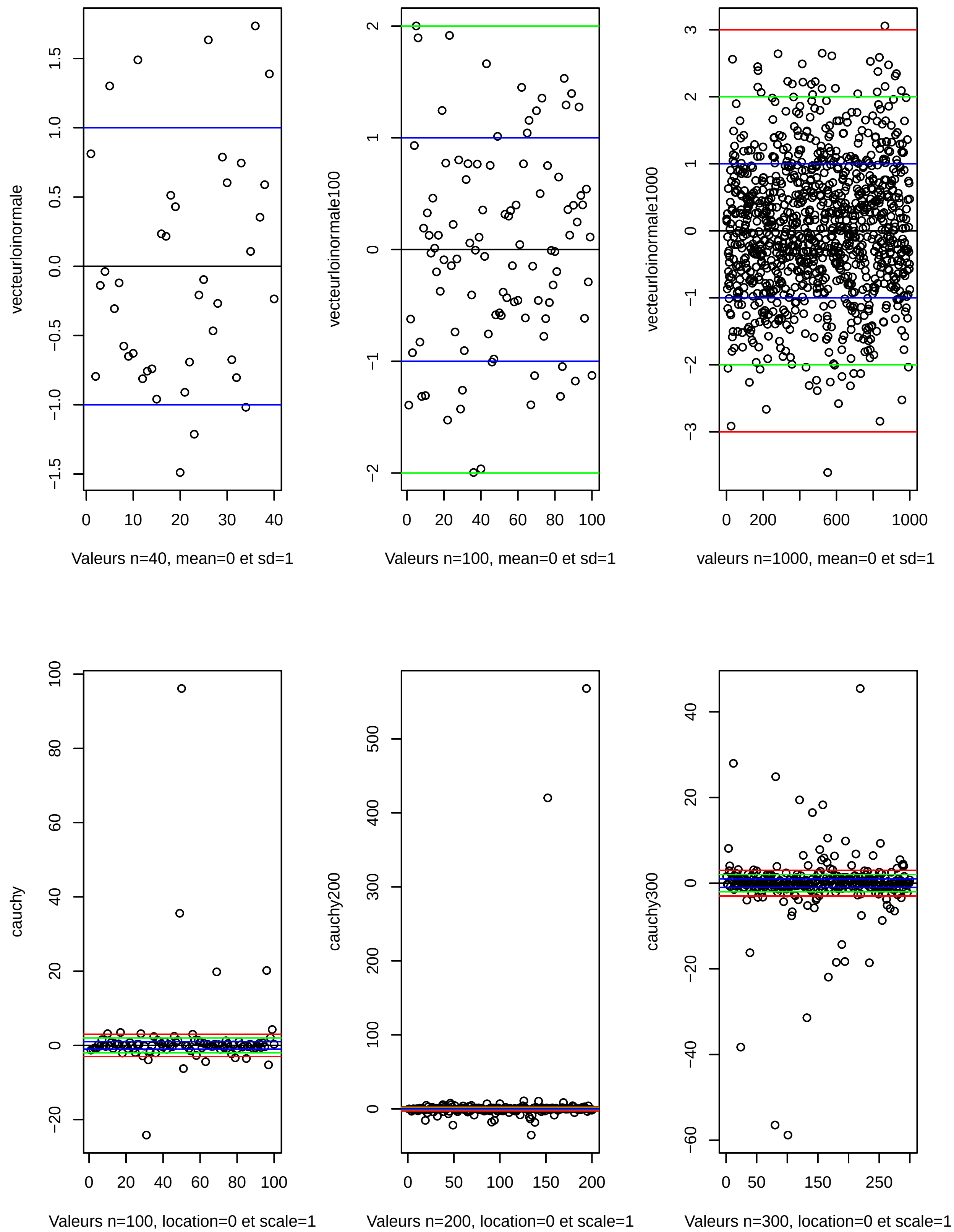
```
## 35   3.68471060
## 36  -0.95437340
## 37  -1.25764727
## 38   0.75028140
## 39   0.12631850
## 40  -5.05863724
```

```r
read.csv("~/Travail/ecole ingé/Stats/TP1/cauchy100",header=TRUE)
```

```
##            X           x
## 1   cauchy11   2.8830930
## 2   cauchy12 -30.2433643
## 3   cauchy13 -11.0585542
## 4   cauchy14  -2.7830620
## 5   cauchy15  -1.1215112
## 6   cauchy16   2.6302934
## 7   cauchy17  10.0733770
## 8   cauchy18  -4.6846856
## 9   cauchy19   0.2687111
## 10 cauchy110  -1.1836058
```

```r
vecteurloinormale<-c(normale=rnorm(40,mean=0,sd=1))
vecteurloinormale100<-c(normale=rnorm(100,mean=0,sd=1))
vecteurloinormale1000<-c(normale=rnorm(1000,mean=0,sd=1))
```

Tracer les données



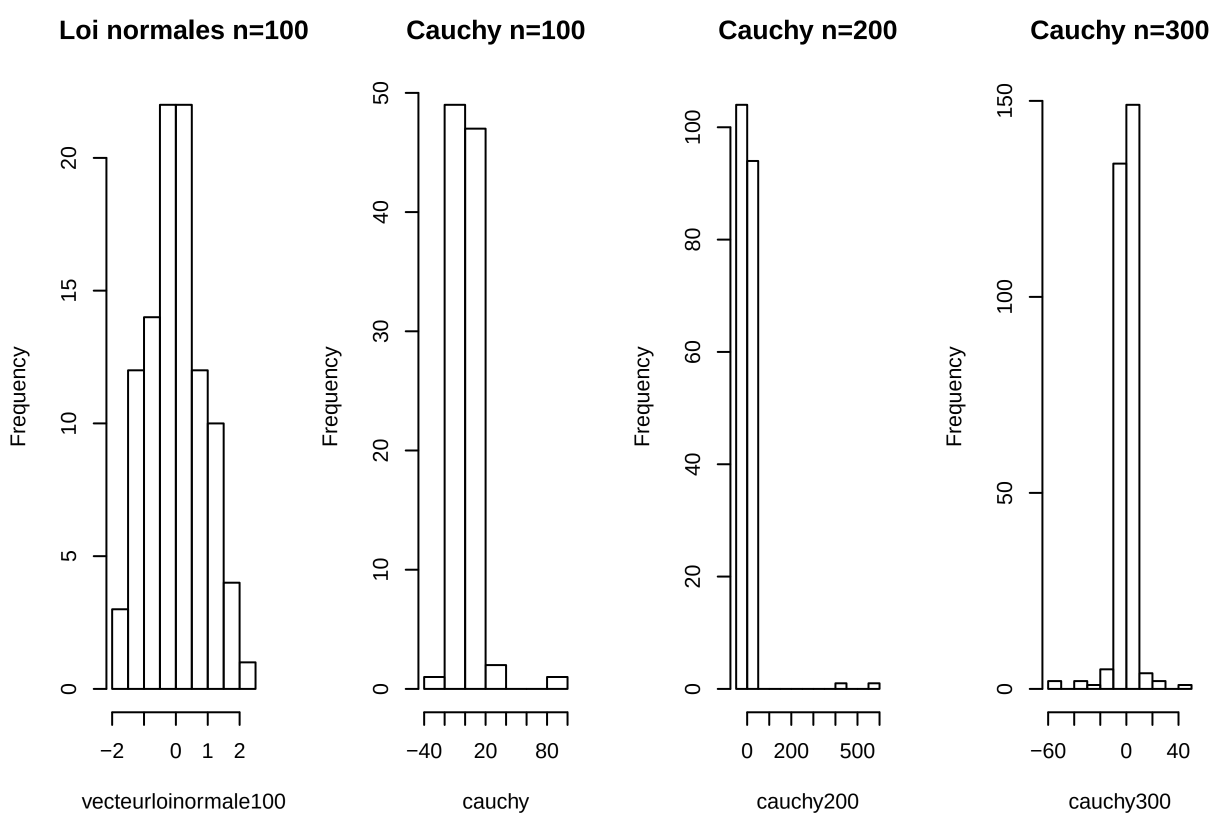
Commentaires

Concernant les lois normales à 40, 100 et 1000 valeurs, l'ensemble des points se situent principalement dans la zone [-1,1] et autour de 0. Cela s'explique par l'hypothèse faite sur la variance, qui se situe entre -1 et 1 etla distribution, qui se centre autour de 0. Plus il y a de valeurs et plus il y a des valeurs qui s'éloignent du centre, ce qui est normal. Cependant, très rares sont les valeurs qui s'éloignent de plus 3 unités du centre.

Concernant la distribution de Cauchy, bien que le nombre de valeurs augmente, il y a très peu de valeur qui s'éloignent du centre. Cependant, contrairement à la distribution de loi normale, dès qu'une valeur s'éloigne, elle peut s'éloigner très loin du centre.

Histogrammes

```r
par(mfcol=c(1,4))
hist(vecteurloinormale100,main=paste("Loi normales n=100"))
hist(cauchy,main=paste("Cauchy n=100"))
hist(cauchy200,main=paste("Cauchy n=200"))
hist(cauchy300,main=paste("Cauchy n=300"))
```




```r
library(moments)
```

Commentaires

Les histogrammes confirment les observations précédentes. Pour les lois normales avec n=100, mean=0 et sd=1 , plus on se rapproche du centre et plus il y a des valeurs. On retrouve bien la forme d'une loi normale.

Concernant les lois de cauchy qui ont toutes pour location 0 et pour échelle 1, toutes les valeurs se situent autour du centre. Cependant, il y a quelques valeurs très éloigné du centre. Il y a donc 1 grand pic de valeurs et plusieurs petits pics contenant 1 ou 2 valeurs très éloigné du centre.

Moments d'ordres supérieurs

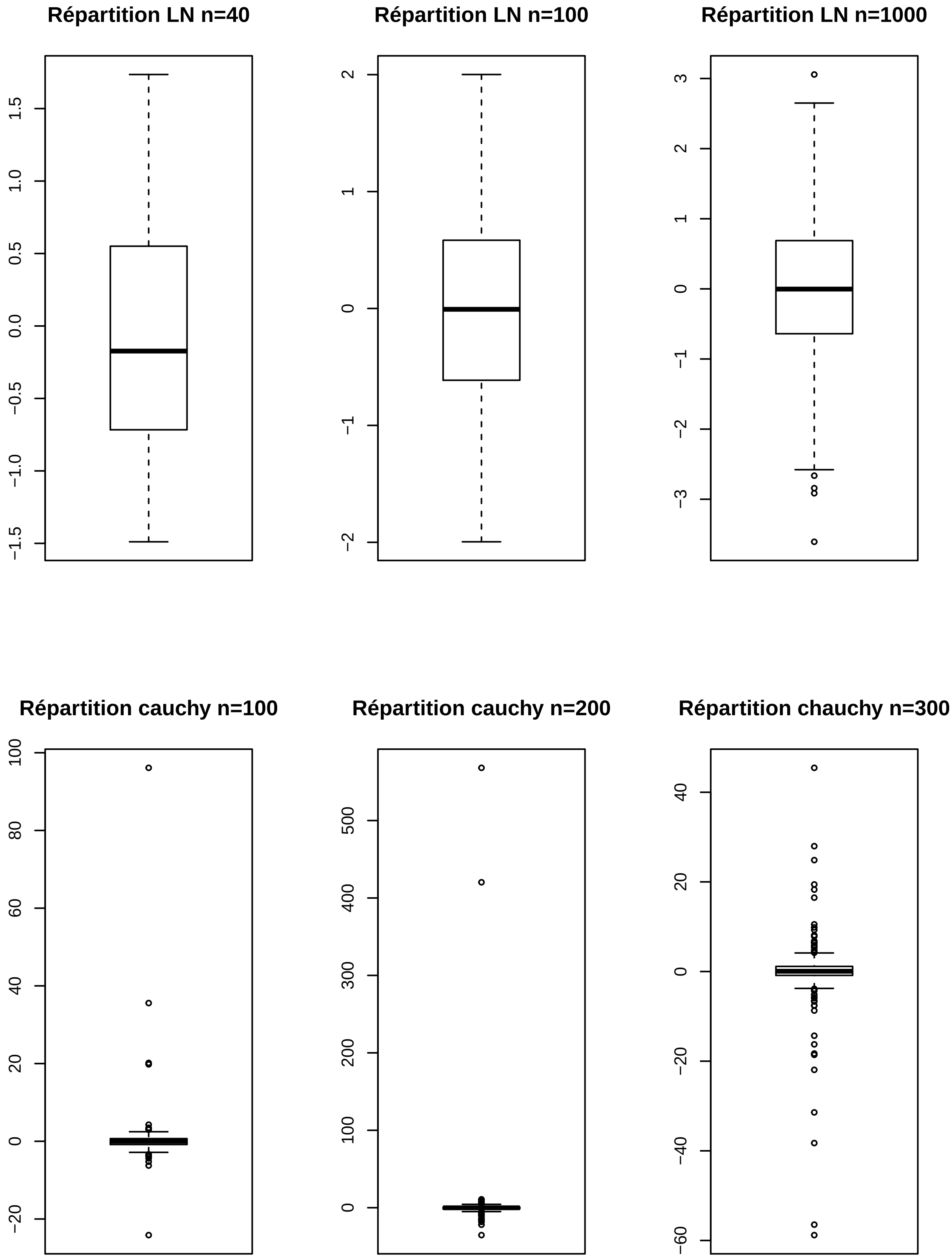
```
momentsordres<-data.frame(
loinormale100.skewness=skewness(vecteurloinormale100),                              loinormale1000.skewness=
loinormale.skewness=skewness(vecteurloinormale),                                loinormale100..kurtosis=kurtos
loinormale1000.moyenne=mean(vecteurloinormale1000),
loinormale.moyenne=mean(vecteurloinormale),
loinormale100.variance=var(vecteurloinormale100), loinormale1000.variance=var(vecteurloinormale1000), lo


Momentsdordres<-write.csv(momentsordres,"~/Travail/ecole ingé/Stats/TP1/Moments d'ordres")
read.csv("~/Travail/ecole ingé/Stats/TP1/Moments d'ordres",header=TRUE)
```

```
##   X loinormale100.skewness loinormale1000.skewness loinormale.skewness
## 1 1              0.0904517              0.02037385            0.525788
##   loinormale100..kurtosis loinormale1000.kurtosis loinormale.kurtosis
## 1                2.542202                3.019051            2.409988
##   loinormale100.moyenne loinormale1000.moyenne loinormale.moyenne
## 1         -0.0006425599             0.01767853        -0.04143595
##   loinormale100.variance loinormale1000.variance loinormale.variance
## 1              0.7937294               0.9895775           0.6880293
```

Commentaires

Les moments d'ordres nous apportent des informations complémentaires sur nos échantillons. Tout d'abord, on obtient l'espérance/moyenne ainsi que la variance de notre loi. VAce la fonction skewness (qui correspond au moment d'ordre 3), si c'est positif, la densité penche vers la droite et vers la gauche si c'est négatif. Avec la fonctions kurtosis (qui correspond au moment d'ordre 4), si c'est positif, la densité à des queues plus épaisses que celles d'une gaussienne. Si c'est négatif, la densité à des queues qui décroissent plus vite qu'une loi normale.

Boxplot et quantiles



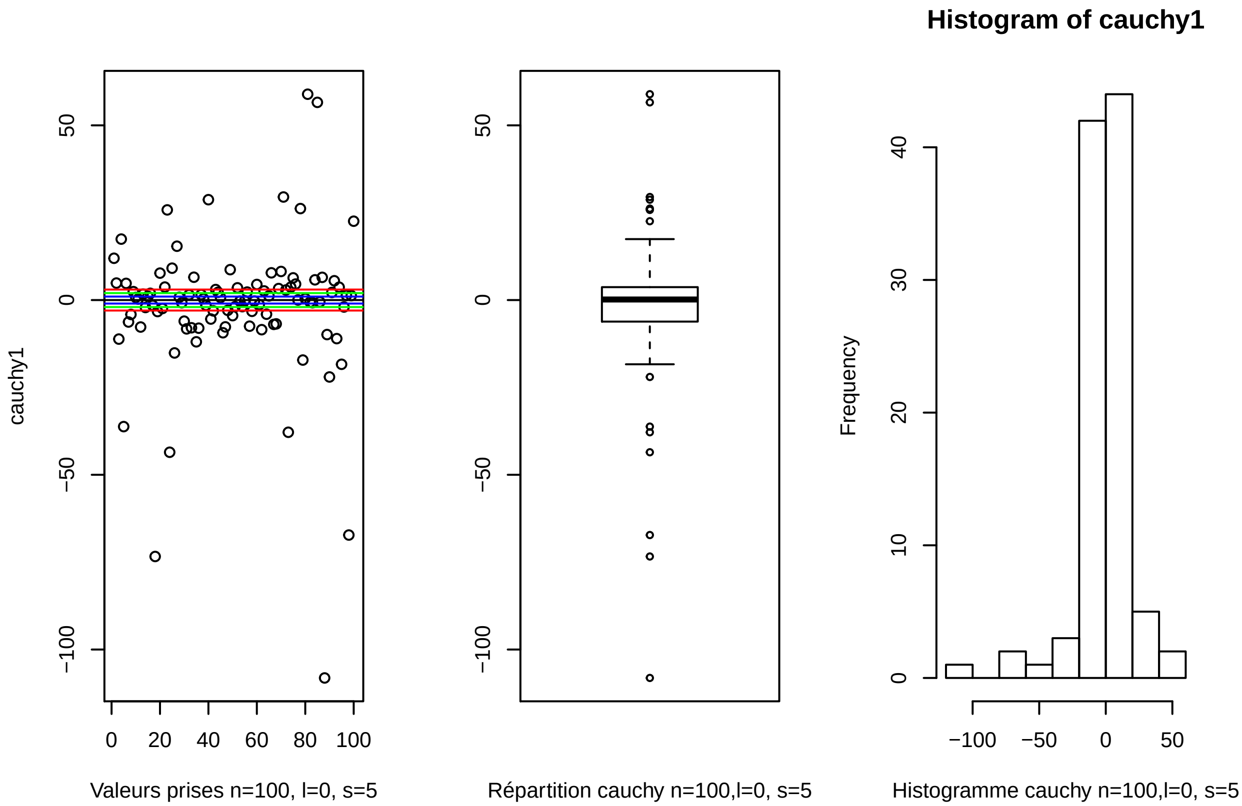
```
##      V1                              V2                               V3
## 1       quantile.vecteurloinormale100. quantile.vecteurloinormale1000.
## 2   0%               -1.99555374129347                -3.60700425764081
## 3  25%              -0.613063895171381               -0.639658921279371
## 4  50%             -0.007053627309438            -0.00314914149779895
## 5  75%               0.562370258273407                0.687061172219842
## 6 100%                2.00116997345081                 3.05692927858283
##                              V4
## 1 quantile.vecteurloinormale.
## 2           -1.48911787262623
## 3          -0.703964965779736
## 4          -0.173164924055108
## 5            0.53148667436974
## 6            1.73517916733608
```

Commentaires

Les boxplots permettent de voir où se situent les valeurs par rapport au centre et aux différentd quantiles. Ce qu'on remarque notamment, c'est que pour les distributions de lois normales, il y peu de valeurs en dehors de la boite et si elles sont en dehors, elles sont proches du premier ou dernier quantile. Pour les lois de Cauchy, il y a aussi des valeurs en dehors des limitations de la boite mais elles sont très loin du premier ou denrier quantile. Cela confirme encore une fois les remarques faites avec les histogrammes.

Interprétation visuelle

```r
cauchy1<-c(cauchy100=rcauchy(n=100,location=0,scale=5))
cauchy2<-c(cauchy200=rcauchy(n=100, location=10,scale=0.5))
cauchy3<-c(cauchy300=rcauchy(n=100,location=-5,scale=3))

par(mfrow=c(1,3))

plot(cauchy1, xlab="Valeurs prises n=100, l=0, s=5")
abline(h=0)
abline(h=-1,col="blue")
abline(h=1,col="blue")
abline(h=2,col="green")
abline(h=-2,col="green")
abline(h=3,col="red")
abline(h=-3,col="red")


boxplot(cauchy1, xlab="Répartition cauchy n=100,l=0, s=5")

hist(cauchy1,xlab="Histogramme cauchy n=100,l=0, s=5")
```




```
momentsordrescauchy1<-data.frame(cauchy1.skewness=skewness(cauchy1), cauchy1.kurtosis=kurtosis(cauchy1)


Momentsdordrescauchy1<-write.csv(momentsordrescauchy1,"~/Travail/ecole ingé/Stats/TP1/Moments d'ordres c
read.csv("~/Travail/ecole ingé/Stats/TP1/Moments d'ordres cauchy n=100, l=0, s=5",header=TRUE)
```

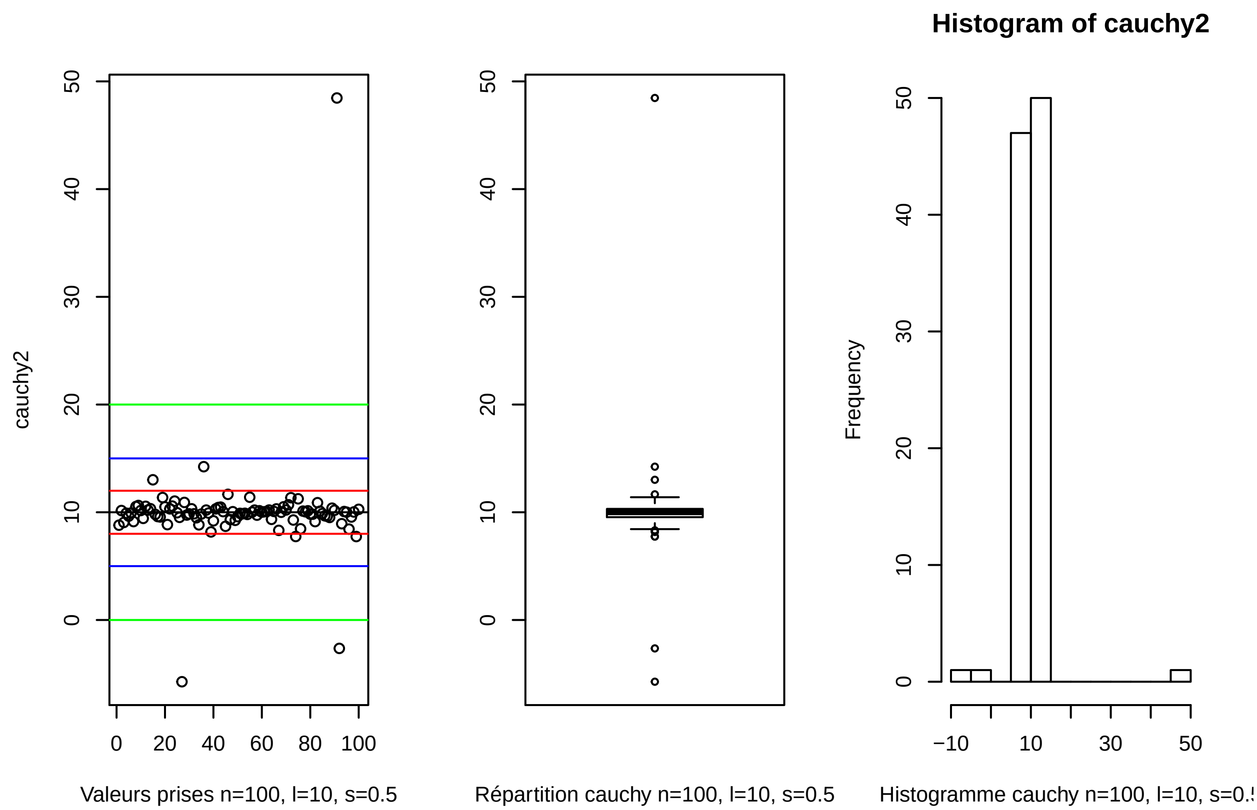
```
##   X cauchy1.skewness cauchy1.kurtosis cauchy1.esperance cauchy1.variance
## 1 1        -1.808997         12.81346         -1.885128         406.3548
```

Commentaires sur cette répartition avec n=100, location=0 et scale=5

Ici, on remarque que l'esperance est supérieure à la location fixée. Cela peut s'expliquer par les quelques valeurs situées loins de la location. Cela augmente(ou diminue) fortement l'espérance ainsi que cela influe énormement sur la variance, d'où cette variance élevée.

```
par(mfrow=c(1,3))
plot(cauchy2, xlab="Valeurs prises n=100, l=10, s=0.5")
abline(h=10)
abline(h=5,col="blue")
abline(h=15,col="blue")
abline(h=0,col="green")
abline(h=20,col="green")
abline(h=8,col="red")
abline(h=12,col="red")
```

```
boxplot(cauchy2, xlab="Répartition cauchy n=100, l=10, s=0.5")

hist(cauchy2,xlab="Histogramme cauchy n=100, l=10, s=0.5")
```




```
momentsordrescauchy2<-data.frame(cauchy2.skewness=skewness(cauchy2), cauchy2.kurtosis=kurtosis(cauchy2)

Momentsdordrescauchy2<-write.csv(momentsordrescauchy2,"~/Travail/ecole ingé/Stats/TP1/Moments d'ordres 
read.csv("~/Travail/ecole ingé/Stats/TP1/Moments d'ordres cauchy n=100, l=10, s=0.5",header=TRUE)
```

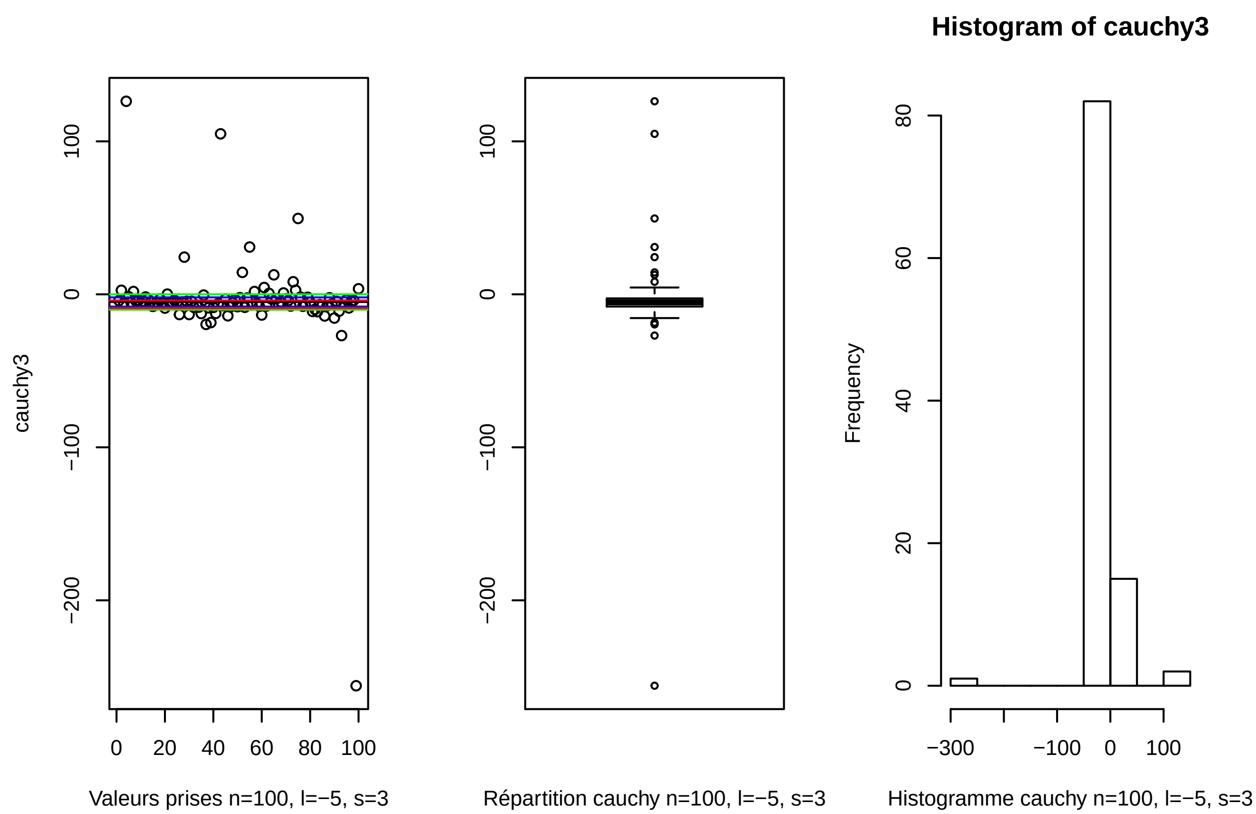
```
##   X cauchy2.skewness cauchy2.kurtosis cauchy2.esperance cauchy2.variance
## 1 1         5.816747         58.58519          10.06324         19.84838
```

Commentaires sur cette répartition avec n=100, location=10 et s=0.5

Ici, l'espérance est proche de la location prévue donc il y a quelques valeurs éloignées de la location mais elles restent assez proches pour ne pas influer sur l'espérance, ce qui explique cette faible variance par rapport à la variance précédente.

```
par(mfrow=c(1,3))
plot(cauchy3, xlab="Valeurs prises n=100, l=-5, s=3")
abline(h=-5)
abline(h=-8,col="blue")
abline(h=-2,col="blue")
```

```
abline(h=0,col="green")
abline(h=-10,col="green")
abline(h=-4,col="red")
abline(h=-9,col="red")

boxplot(cauchy3, xlab="Répartition cauchy n=100, l=-5, s=3")

hist(cauchy3,xlab="Histogramme cauchy n=100, l=-5, s=3")
```




```
momentsordrescauchy3<-data.frame(cauchy3.skewness=skewness(cauchy3), cauchy3.kurtosis=kurtosis(cauchy3)

Momentsdordrescauchy3<-write.csv(momentsordrescauchy3,"~/Travail/ecole ingé/Stats/TP1/Moments d'ordres c
read.csv("~/Travail/ecole ingé/Stats/TP1/Moments d'ordres cauchy n=100, l=-5, s=3",header=TRUE)
```

```
##   X cauchy3.skewness cauchy3.kurtosis cauchy3.esperance cauchy3.variance
## 1 1        -3.805378         43.89783         -4.433083         1015.736
```

Commentaires sur cette répartition avec n=100, location=-5 et scale=3

Comme précédement, on remarque que l'espérance est proche de la location, ce qui est logique. Il y a beaucoup de valeurs proches de la location et les quelques valeurs éloignées influent légèrement sur l'espérance puisqu'elle est supérieure à la location. C'est pourquoi la variance est la aussi élevée.

Ce que l'on peut conclure ici, c'est que plus on impose une échelle élevée, plus il y a des risques d'avoir des valeurs éloignées de la location et donc plus notre variance est importante et plus notre espérance s'éloigne de la location.

Conclusion du TP

Ce TP m'a permis de prendre en main le logiciel R encore inconnu jusqu'ici. Il m'a permis aussi d'apprendre les bases de la statistiques avec des histogrammes ou même des boxplots qui permettent de représenter visuellement une densité de loi et qui permettent d'en déduire des caractéristiques. Malheureusement, je n'ai pu extraire le fichier Data.txt qui m'aurait permis de faire une application directe de ce TP avec des choses de la vie quotidienne.