Probabilités discrètes et statistiques descriptives

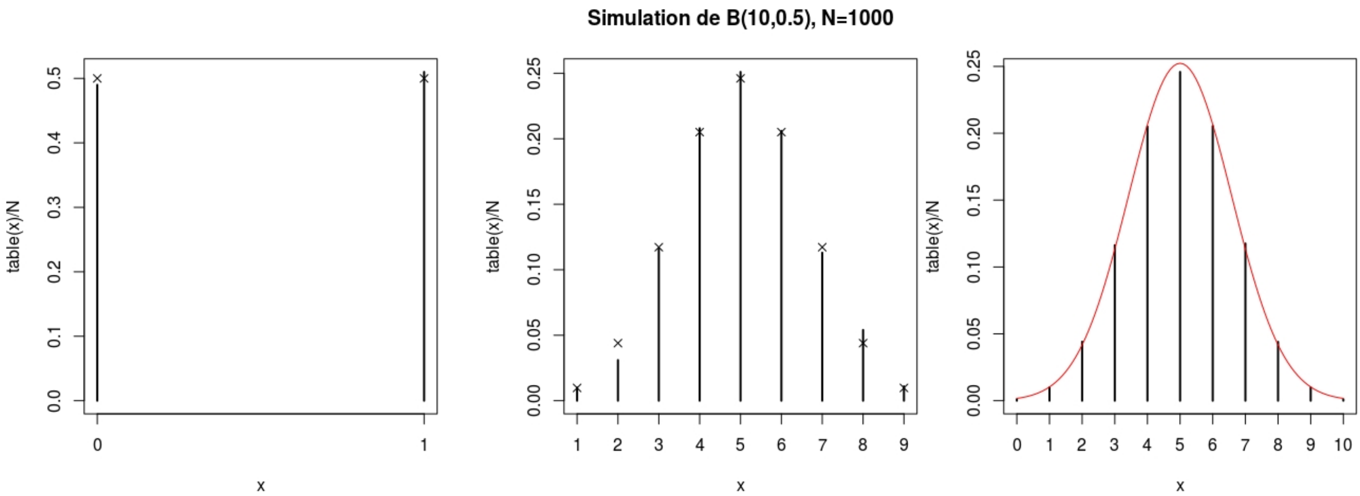
L2 Université Lyon 1 2018-2019

Correction du TP 3 : Simulation avec R

1 Un premier exemple : loi binomiale et loi normale

1. Simuler $N = 100$ réalisations de la loi de Bernoulli $\mathcal{B}(1, 0.5)$ (on pourra utiliser la fonction `rbinom`). Tracer le diagramme en bâton en utilisant la commande `plot(...,...,type='h')`.

```
N=100
x=rbinom(N,1,0.5)
plot(table(x)/N,type='h')
points(0:1,dbinom(0:1,1,0.5),pch=4)
```



2. Simuler $N = 100$ réalisations de la loi binomiale $\mathcal{B}(10, 0.5)$ Tracer le diagramme en bâton correspondant et représenter sur le même graphique les probabilités théoriques correspondantes (on pourra utiliser la fonction `points`).

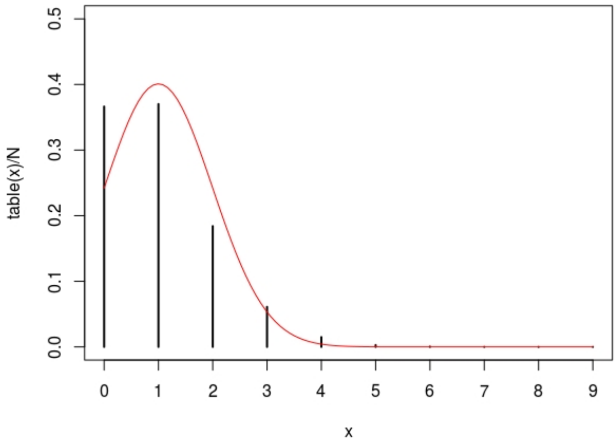
```
N=100 #tester avec des valeurs de N de plus en plus grandes
x=rbinom(N,10,0.5)
plot(table(x)/N,type='h',main="Simulation de B(10,0.5), N=1000")
points(0:10,dbinom(0:10,10,0.5),pch=4)
```

3. Simuler $N = 10^6$ réalisations de la loi binomiale $\mathcal{B}(10, 0.5)$ Tracer le diagramme en bâton de cet échantillon et représenter sur le même graphique la densité de la loi normale adéquate (on pourra utiliser la fonction `curve(dnorm(x,...,...),add=TRUE)`).

```
N=10^6
x=rbinom(N,10,0.5)
plot(table(x)/N,type='h')
curve(dnorm(x,5,sqrt(5/2)),add=TRUE,col='red')
```

4. Simuler $N = 10^6$ réalisations de la loi binomiale $\mathcal{B}(100, 0.01)$ Tracer le diagramme en bâton de cet échantillon et représenter sur le même graphique la densité de la loi normale adéquate.

```
N=10^6
x=rbinom(N,100,0.01)
plot(table(x)/N, type='h',ylim=c(0,0.5))
curve(dnorm(x,1,sqrt(0.99)),add=TRUE, col='red')
```



2 Théorème central limite (suite)

Simuler 10 échantillons de la loi $\mathcal{U}([0,3])$, notés $ech1$, $ech2$,...

Séparer la fenêtre graphique en plusieurs parties en utilisant `par(mfrow=c(2,3))`. Tracer les histogrammes de :

1. $ech1$,
2. $(ech1 + ech2)/2$,
3. $(ech1 + ech2 + ech3)/3$,
4. $(ech1 + ech2 + ech3 + ech4)/4$,
5. $(ech1 + ech2 + ech3 + ech4 + ech5)/5$,
6. $(ech1 + \cdots + ech10)/10$,

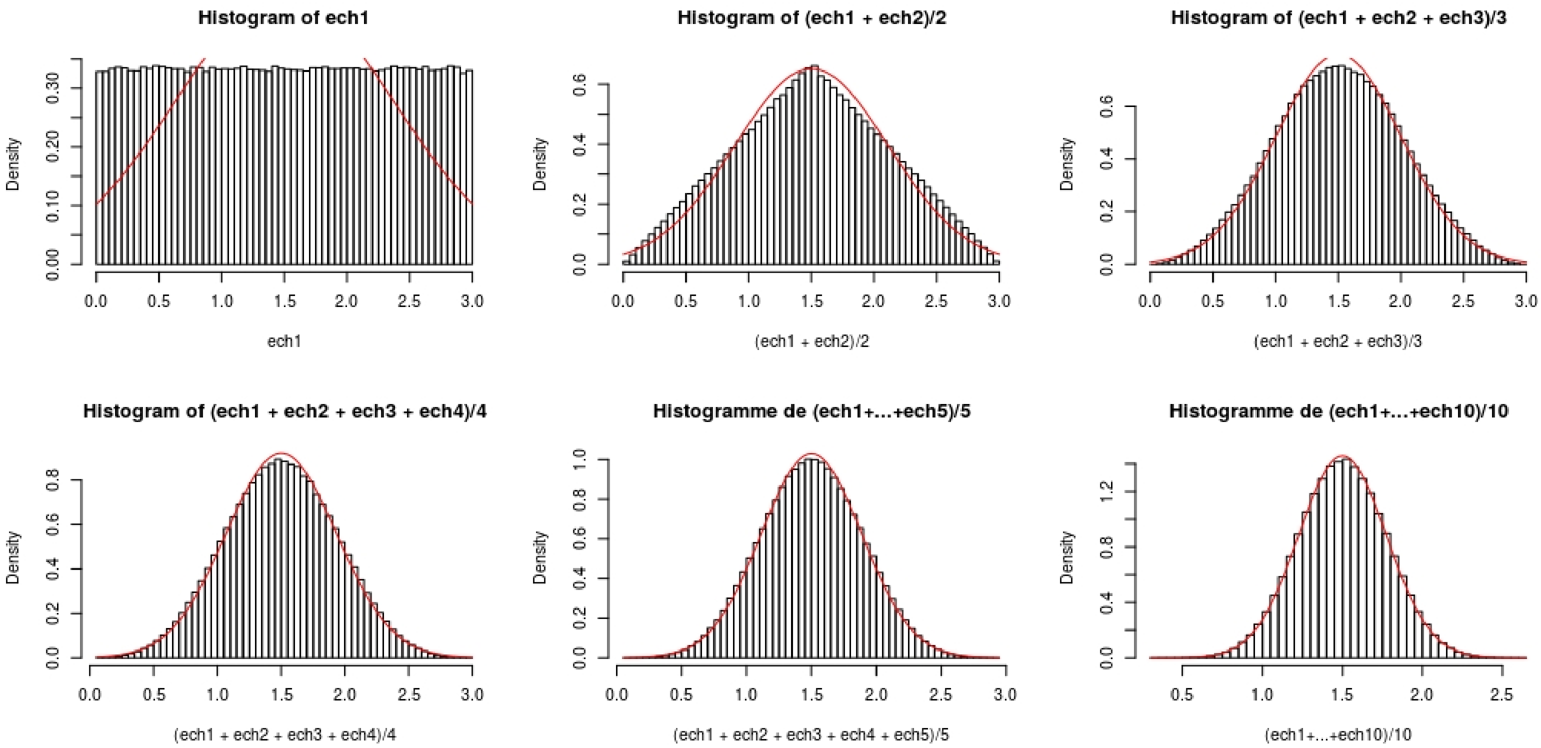
Ajouter sur chaque histogramme la loi normale adéquate. **Attention dans le code R ci-dessous, on a été à la ligne pour des raisons de lisibilité, alors qu'il ne faut pas aller à la ligne dans R, à l'intérieur de l'appel d'une fonction. Le symbole "␣" désigne un espace.**

Pour les paramètres de la loi normale, on utilise que la moyenne d'une loi uniforme sur [a,b] est $(a + b)/2$ (cf cours p 57) et sa variance $(b - a)^2/12$ (exo).

Ensuite la variance de $(X_1 + ...X_n)/n$ pour un échantillon i.i.d est $Var(X_1)n/n^2 = Var(X_1)/n$. (cf demo de la loi faible des grand nombres).

```
N=10^6
ech1=runif(N,0,3)
ech2=runif(N,0,3)
ech3=runif(N,0,3)
ech4=runif(N,0,3)
ech5=runif(N,0,3)
ech6=runif(N,0,3)
ech7=runif(N,0,3)
ech8=runif(N,0,3)
ech9=runif(N,0,3)
ech10=runif(N,0,3)

par(mfrow=c(2,3))
```

```r
hist(ech1,br=50,freq=FALSE)
curve(dnorm(x,3/2,sqrt(9/12)),add=TRUE,col='red')

hist((ech1+ech2)/2,br=50,freq=FALSE)
curve(dnorm(x,3/2,sqrt(9/12)/sqrt(2)),add=TRUE,col='red')

hist((ech1+ech2+ech3)/3,br=50,freq=FALSE)
curve(dnorm(x,3/2,sqrt(9/12)/sqrt(3)),add=TRUE,col='red')

hist((ech1+ech2+ech3+ech4)/4,br=50,freq=FALSE)
curve(dnorm(x,3/2,sqrt(9/12)/sqrt(4)),add=TRUE,col='red')

hist((ech1+ech2+ech3+ech4+ech5)/5,br=50,freq=FALSE,
main="Histogramme de (ech1+...+ech5)/5")
curve(dnorm(x,3/2,sqrt(9/12)/sqrt(5)),add=TRUE,col='red')

hist((ech1+ech2+ech3+ech4+ech5+ech6+ech7+ech8+ech9+ech10)/10,br=50,freq=FALSE,
main="Histogramme de (ech1+...+ech10)/10",xlab="(ech1+...+ech10)/10")
curve(dnorm(x,3/2,sqrt(9/12)/sqrt(10)),add=TRUE,col='red')
```

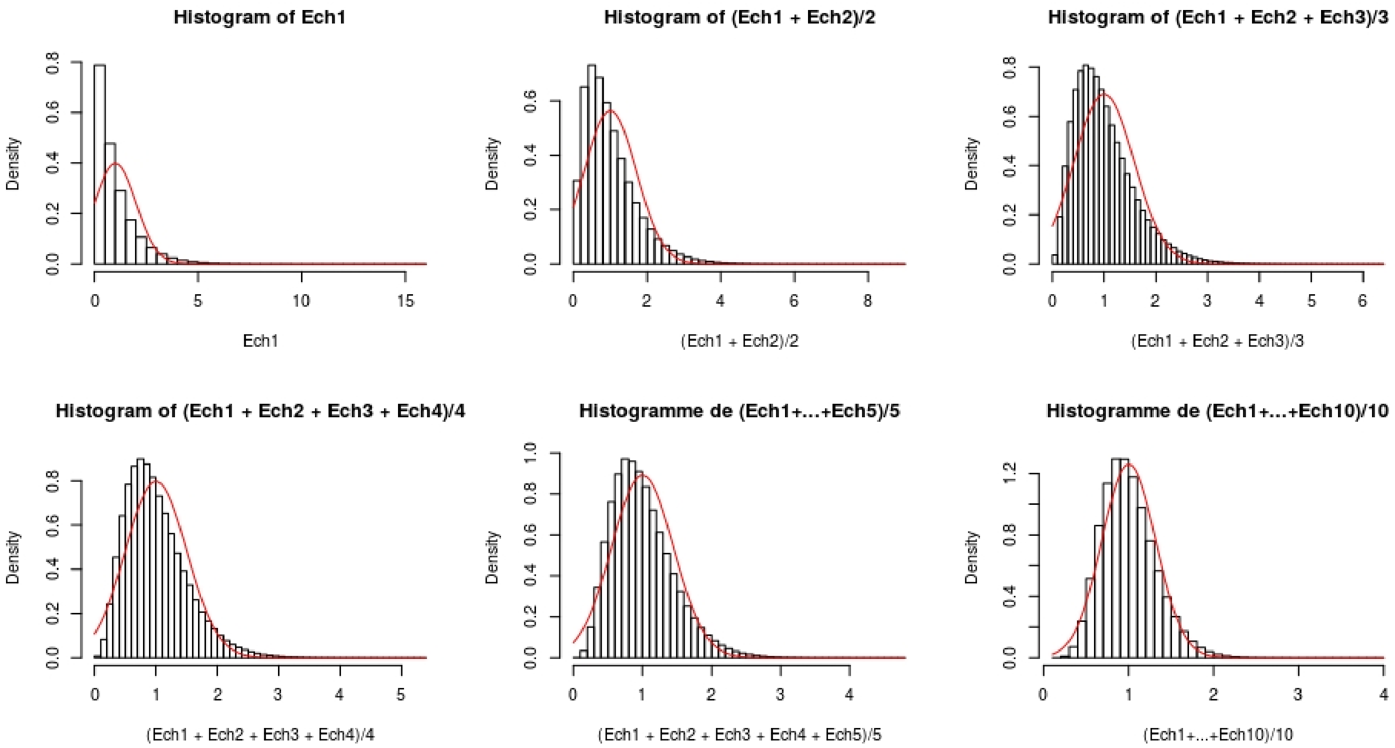


Recommencer avec la loi exponentielle $\mathcal{E}(1)$.

Pour les paramètres de la loi normale, on utilise que la moyenne d'une loi exponentielle $\mathcal{E}(\lambda)$ est $1/\lambda$ (cf cours p 57) et sa variance $1/\lambda^2$ (exo).

```r
Ech1=rexp(N,1)
Ech2=rexp(N,1)
Ech3=rexp(N,1)
Ech4=rexp(N,1)
Ech5=rexp(N,1)
Ech6=rexp(N,1)
Ech7=rexp(N,1)
```

```
Ech8=rexp(N,1)
Ech9=rexp(N,1)
Ech10=rexp(N,1)

par(mfrow=c(2,3))
hist(Ech1, breaks=50,freq=FALSE)
curve(dnorm(x,1,1),add=TRUE, col='red')

hist((Ech1+Ech2)/2,breaks=50,freq=FALSE)
curve(dnorm(x,1,1/sqrt(2)),add=TRUE, col='red')

hist((Ech1+Ech2+Ech3)/3,breaks=50,freq=FALSE)
curve(dnorm(x,1,1/sqrt(3)),add=TRUE, col='red')

hist((Ech1+Ech2+Ech3+Ech4)/4,br=50,freq=FALSE)
curve(dnorm(x,1,1/sqrt(4)),add=TRUE, col='red')

hist((Ech1+Ech2+Ech3+Ech4+Ech5)/5,breaks=50,freq=FALSE
,main="Histogramme de (Ech1+...+Ech5)/5")
curve(dnorm(x,1,1/sqrt(5)),add=TRUE, col='red')

hist((Ech1+Ech2+Ech3+Ech4+Ech5+Ech6+Ech7+Ech8+Ech9+Ech10)/10
,main="Histogramme de (Ech1+...+Ech10)/10",breaks=50,freq=FALSE)
curve(dnorm(x,1,1/sqrt(10)),add=TRUE, col='red')
```



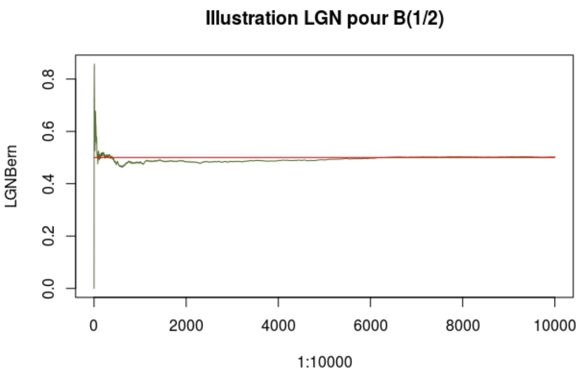
3 Loi des grands nombres

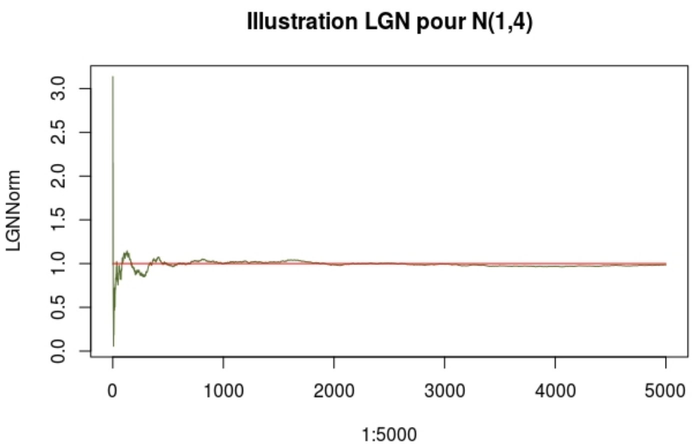
Attention, on revient en affichage normal avec `par(mfrow=c(1,1))`.
Simuler un échantillon $(X_1, X_2, ..., X_N)$ de variables i.i.d. de loi $\mathcal{B}(1, 0.5)$ pour N grand. Illustrer

la loi des grands nombres en traçant $\overline{X}_n = \frac{X_1+X_2+\cdots X_n}{n}$. en fonction de n (compris entre 1 et N). On pourra pour cela utiliser la fonction cumsum.

On pourra pour cela utiliser la fonction cumsum.

```
SBern<-rbinom(10000,1,0.5)

LGNBern<-cumsum(SBern)/(1:10000)

plot(x=1:10000,y=LGNBern,type='l',col="darkolivegreen",
main="Illustration LGN pour B(1/2)")

points(x=1:10000,y=rep(x=0.5,10000),col="red",type='l')
```



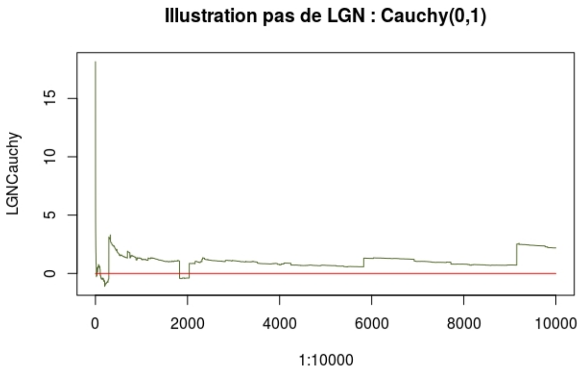


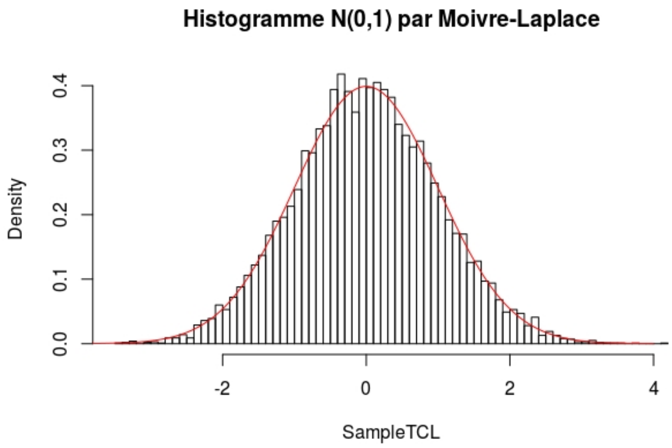
-Procéder de même avec des échantillons de loi $\mathcal{N}(1,4)$ et de Cauchy de paramètre 0 et 1.

```
SNorm14<-rnorm(5000,mean=1,sd=2)

LGNNorm<-cumsum(SNorm14)/(1:5000)

plot(x=1:5000,y=LGNNorm,type='l',col="darkolivegreen",
main="Illustration LGN pour N(1,4)")

points(x=1:5000,y=rep(x=1,5000),col="red",type='l')

Scauchy<-rcauchy(10000,location=0,scale=1)

LGNCauchy<-cumsum(Scauchy)/(1:10000)

plot(x=1:10000,y=LGNCauchy,type='l',col="darkolivegreen",
main="Illustration pas de LGN : Cauchy(0,1)")

points(x=1:10000,y=rep(x=0,10000),col="red",type='l')
```

-Que remarque-t-on ? Comme la loi de Cauchy n'est pas d'ordre 1 ($\int_{-\infty}^{\infty} \frac{|x|}{1+x^2}dx = +\infty$) le théorème ne s'applique, pas de convergence vers la moyenne probabiliste (qui n'existe pas). En fait la variable

$(X_1 + \cdots + X_n)/n$ est toujours de même loi de Cauchy, et donc autant aléatoire que X_1 !





4 Théorème de Moivre-Laplace

Pour un échantillon $(X_1, X_2, ..., X_N)$ de variables i.i.d. de loi $\mathcal{B}(1,p)$, on définit la variable aléatoire V_N par

$$V_N = \sqrt{N}\frac{\overline{X}_N - p}{\sqrt{p(1-p)}}$$

avec $\overline{X}_N = \frac{X_1+X_2+\cdots X_N}{N}$.

Illustrer le Théorème de Moivre-Laplace (Théorème centrale limite) en simulant k réalisations indépendantes de V_N (on pourra utiliser une boucle for), et en traçant dans le même graphique l'histogramme correspondant et la densité de la loi $\mathcal{N}(0,1)$. On prend $p = 0.5$

```
SampleTCL<-rep(x=1,times=10000)

for(i in 1:10000) {

  SBernLarge<-rbinom(10000,1,0.5)

  LGNBernLarge<-sum(SBernLarge)/10000

  SampleTCL[i]<-sqrt(40000)*(LGNBernLarge-1/2)

}

hist(SampleTCL, freq=FALSE, breaks=100,
main="Histogramme N(0,1) par Moivre-Laplace")
x=seq(from=-4, to=4, by=0.01);y=dnorm(x,mean=0,sd=1)

lines(x,y,col="red")
```