

TP R Statistique Mathématiques 1

Mohamed FOFANA | L2 MIASHS - Université Grenoble Alpes

Contents

1 Séance TP 1: Lois usuelles continues - Intervalles de confiance	1
1.1 De la loi uniforme à la loi exponentielle	1
1.2 Théorème Central Limit	4
1.3 Lois normales, loi du χ^2 , loi de Fisher	8
2 Intervalles de confiance	19
2.1 Construction de l'intervalle de confiance	20
2.2 Simulations	21

1 Séance TP 1: Lois usuelles continues - Intervalles de confiance

1.1 De la loi uniforme à la loi exponentielle

Soit X suivant la loi uniforme $\mathcal{U} [0; 1]$. On va illustrer par simulation que la variable

$$y = -\frac{1}{3} \ln(1 - X)$$

\$ suit une loi exponentielle de paramètre 3.

- (i) On souhaite générer un échantillon de X de taille $n = 500$ que l'on stock dans un vecteur x . On utilise la fonction `runif(n,min=0,max=1)` (où `r` dans `runif` renvoie à `run` pour “tirer”).

```
x = runif(500, min=0,max=1)
```

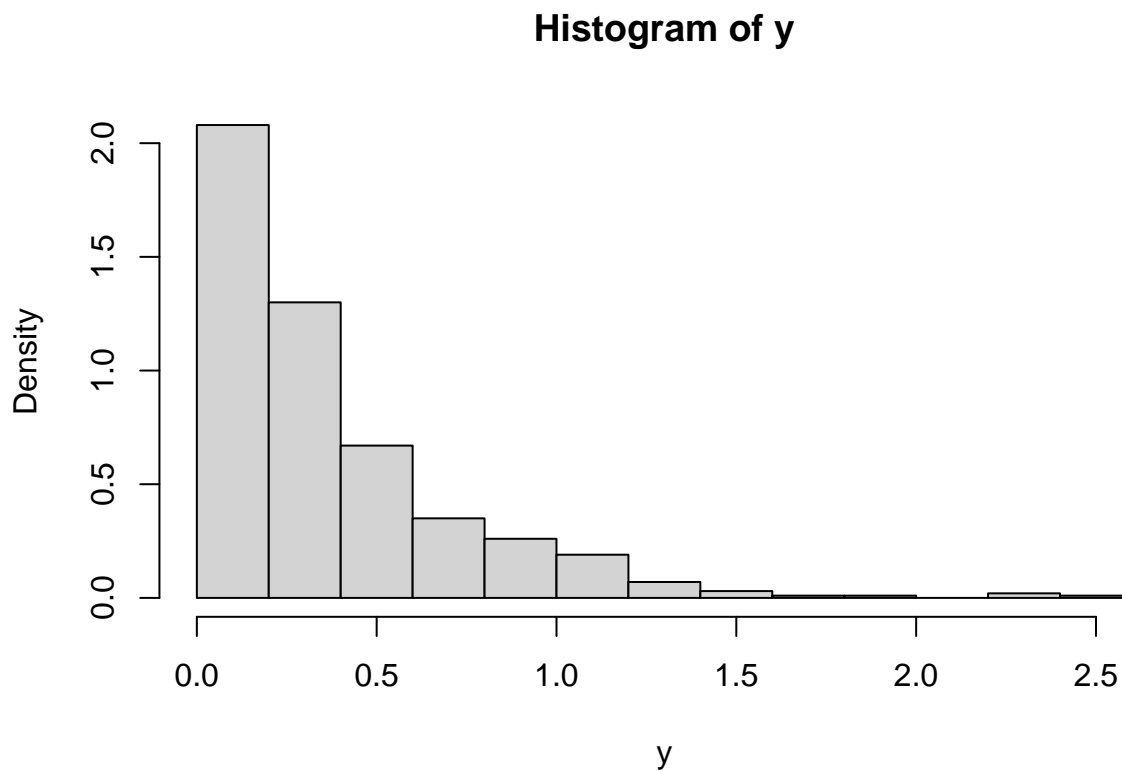
- (ii) On crée ensuite le vecteur y contenant les valeurs de Y associées à x . R fait l'opération suivante termes à termes

```
y = -log(1-x)/3
```

- (iii) On souhaite superposer l'histogramme des tirages de Y avec la courbe de la densité de la loi exponentielle de paramètre 3.

- On trace l'histogramme de y :

```
hist (y, probability=TRUE)
```



(on utilise le paramètre `probability=T` pour se ramener à une densité et non à l'histogramme en terme d'effectif).

- Traçons la densité de loi exponentielle de paramètre 3. Commençons par discrétiser les l'axe des abscisses sur $[0; 2]$ avec un pas de 0,1 :

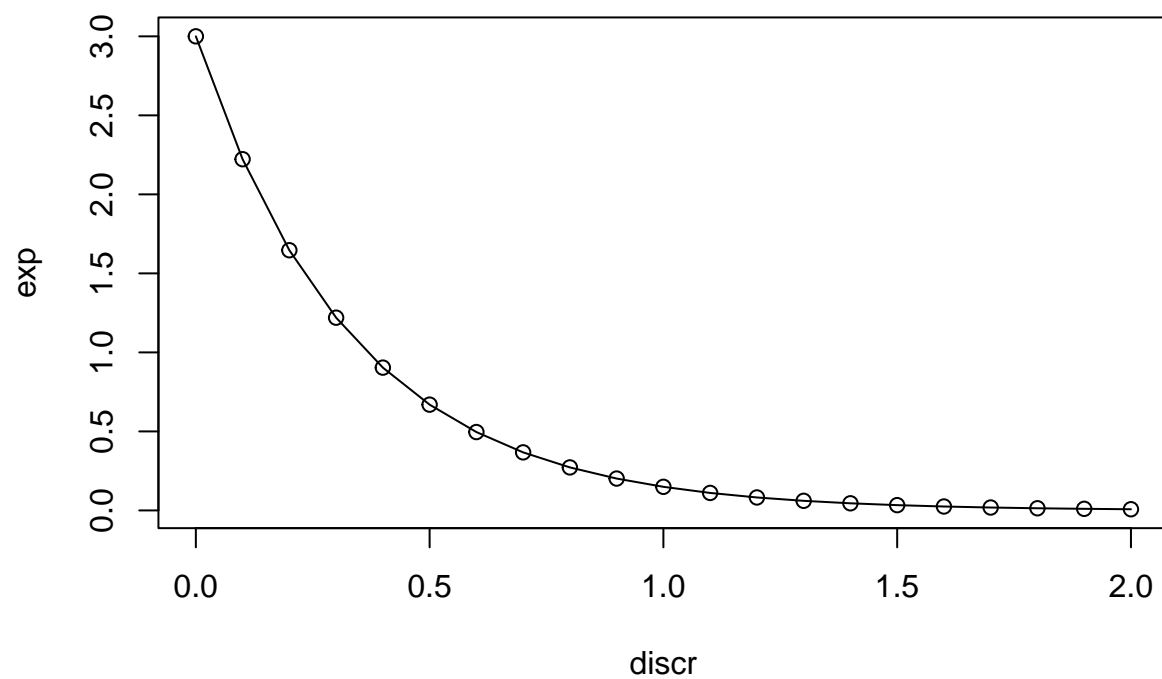
```
discr=seq(0,2,0.1)
```

On calcule les valeurs de la densité de la loi exponentielle de paramètre 3 sur cette discrétisation :

```
exp=dexp (discr,rate=3) # (d dans dexp() signifie density).
```

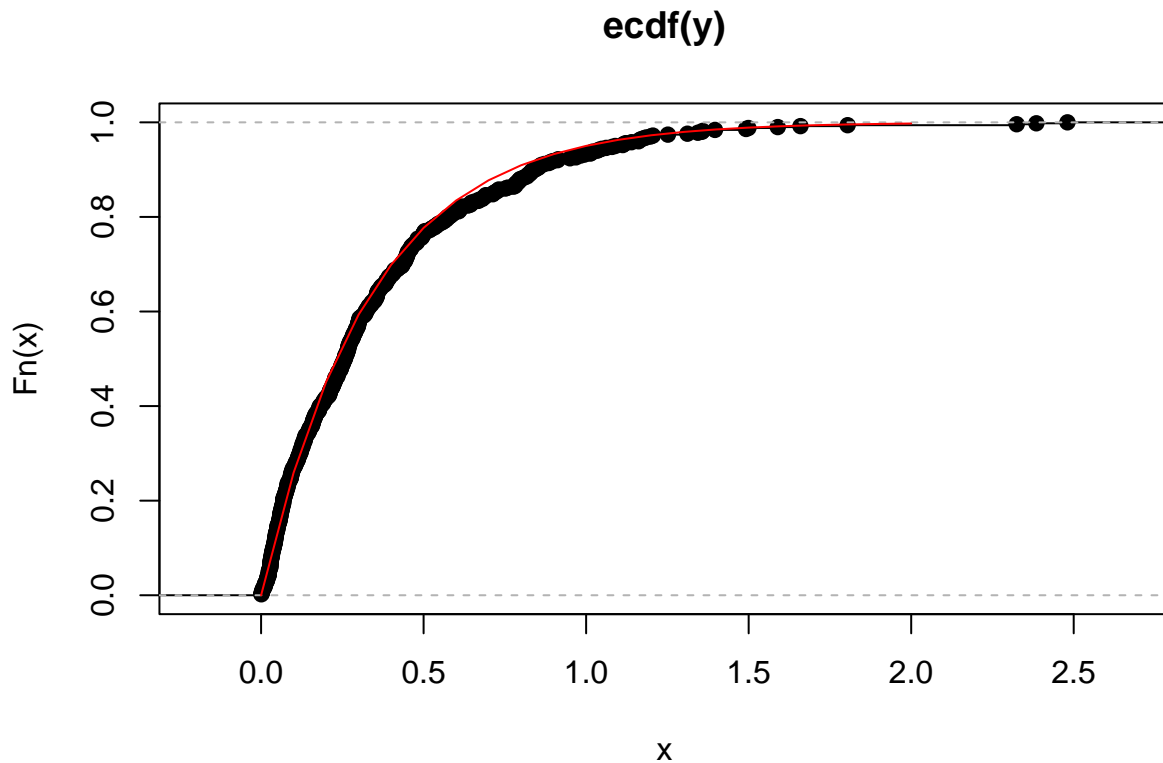
- On ajoute la courbe de la densité de l'exponentielle :

```
plot(discr,exp)  
lines (discr, exp)
```



(iv) On pourra également comparer les fonctions de répartitions :

```
plot (ecdf (y))
lines (discr, pexp (discr, 3), col="red")
```



```
# (ecdf empirical cumulative distribution function, p dans pexp pour probability function.
# (fonction de répartition)).
```

Remarque 1.1 On aurait, bien entendu, pu utiliser directement la commande $y = \text{rexp}(500, 3)$, mais il n'y aurait pas eu d'exercice !

1.2 Théorème Central Limit

On se propose d'observer la convergence en loi de la loi binômiale vers la loi normale :

$$\frac{\mathcal{B}(n, p) - np}{\sqrt{np(1-p)}} \rightarrow \mathcal{N}(0, 1).$$

On fixe $n = 10$, $p = 0,25$.

(a) On effectue 50000 tirages selon la loi binômiale $B(10, 0, 25)$:

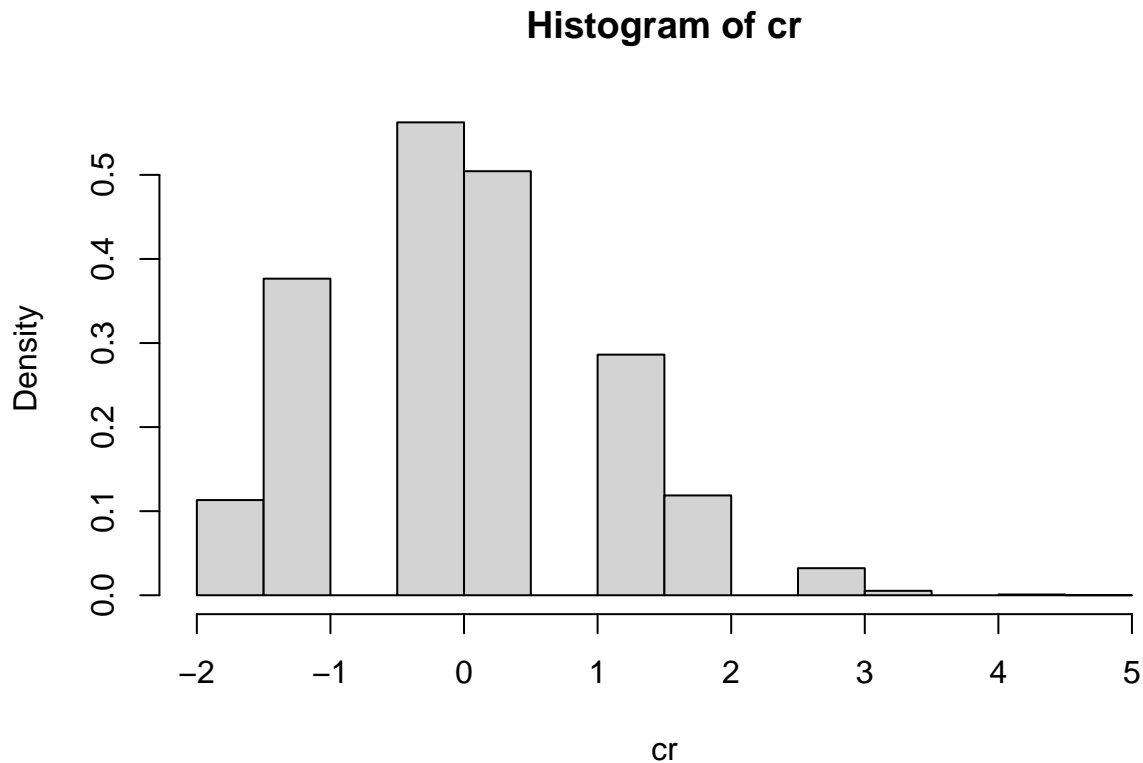
```
bin=rbinom (50000, 10,0.25)
```

(b) On centre et on réduit dans la variable **cr** :

```
cr=(bin-0.25*10)/sqrt (10*0.25*0.75)
```

(c) On trace:

```
hist(cr,probability=TRUE)
```



(d) On discrétise l'intervalle $[-3, 3]$ pour tracer la densité de la loi normale centrée réduite :

```
hist(cr,probability=TRUE)
discr=seq(-3,3,0.01)
```

```
## Warning in discr - seq(-3, 3, 0.01): la taille d'un objet plus long n'est pas
## multiple de la taille d'un objet plus court
```

```
## [1] 3.000000e+00 3.090000e+00 3.180000e+00 3.270000e+00 3.360000e+00
## [6] 3.450000e+00 3.540000e+00 3.630000e+00 3.720000e+00 3.810000e+00
## [11] 3.900000e+00 3.990000e+00 4.080000e+00 4.170000e+00 4.260000e+00
## [16] 4.350000e+00 4.440000e+00 4.530000e+00 4.620000e+00 4.710000e+00
## [21] 4.800000e+00 2.790000e+00 2.880000e+00 2.970000e+00 3.060000e+00
## [26] 3.150000e+00 3.240000e+00 3.330000e+00 3.420000e+00 3.510000e+00
## [31] 3.600000e+00 3.690000e+00 3.780000e+00 3.870000e+00 3.960000e+00
## [36] 4.050000e+00 4.140000e+00 4.230000e+00 4.320000e+00 4.410000e+00
## [41] 4.500000e+00 4.590000e+00 2.580000e+00 2.670000e+00 2.760000e+00
## [46] 2.850000e+00 2.940000e+00 3.030000e+00 3.120000e+00 3.210000e+00
## [51] 3.300000e+00 3.390000e+00 3.480000e+00 3.570000e+00 3.660000e+00
## [56] 3.750000e+00 3.840000e+00 3.930000e+00 4.020000e+00 4.110000e+00
## [61] 4.200000e+00 4.290000e+00 4.380000e+00 2.370000e+00 2.460000e+00
## [66] 2.550000e+00 2.640000e+00 2.730000e+00 2.820000e+00 2.910000e+00
## [71] 3.000000e+00 3.090000e+00 3.180000e+00 3.270000e+00 3.360000e+00
## [76] 3.450000e+00 3.540000e+00 3.630000e+00 3.720000e+00 3.810000e+00
## [81] 3.900000e+00 3.990000e+00 4.080000e+00 4.170000e+00 2.160000e+00
## [86] 2.250000e+00 2.340000e+00 2.430000e+00 2.520000e+00 2.610000e+00
## [91] 2.700000e+00 2.790000e+00 2.880000e+00 2.970000e+00 3.060000e+00
```

```

## [96] 3.150000e+00 3.240000e+00 3.330000e+00 3.420000e+00 3.510000e+00
## [101] 3.600000e+00 3.690000e+00 3.780000e+00 3.870000e+00 3.960000e+00
## [106] 1.950000e+00 2.040000e+00 2.130000e+00 2.220000e+00 2.310000e+00
## [111] 2.400000e+00 2.490000e+00 2.580000e+00 2.670000e+00 2.760000e+00
## [116] 2.850000e+00 2.940000e+00 3.030000e+00 3.120000e+00 3.210000e+00
## [121] 3.300000e+00 3.390000e+00 3.480000e+00 3.570000e+00 3.660000e+00
## [126] 3.750000e+00 1.740000e+00 1.830000e+00 1.920000e+00 2.010000e+00
## [131] 2.100000e+00 2.190000e+00 2.280000e+00 2.370000e+00 2.460000e+00
## [136] 2.550000e+00 2.640000e+00 2.730000e+00 2.820000e+00 2.910000e+00
## [141] 3.000000e+00 3.090000e+00 3.180000e+00 3.270000e+00 3.360000e+00
## [146] 3.450000e+00 3.540000e+00 1.530000e+00 1.620000e+00 1.710000e+00
## [151] 1.800000e+00 1.890000e+00 1.980000e+00 2.070000e+00 2.160000e+00
## [156] 2.250000e+00 2.340000e+00 2.430000e+00 2.520000e+00 2.610000e+00
## [161] 2.700000e+00 2.790000e+00 2.880000e+00 2.970000e+00 3.060000e+00
## [166] 3.150000e+00 3.240000e+00 3.330000e+00 1.320000e+00 1.410000e+00
## [171] 1.500000e+00 1.590000e+00 1.680000e+00 1.770000e+00 1.860000e+00
## [176] 1.950000e+00 2.040000e+00 2.130000e+00 2.220000e+00 2.310000e+00
## [181] 2.400000e+00 2.490000e+00 2.580000e+00 2.670000e+00 2.760000e+00
## [186] 2.850000e+00 2.940000e+00 3.030000e+00 3.120000e+00 1.110000e+00
## [191] 1.200000e+00 1.290000e+00 1.380000e+00 1.470000e+00 1.560000e+00
## [196] 1.650000e+00 1.740000e+00 1.830000e+00 1.920000e+00 2.010000e+00
## [201] 2.100000e+00 2.190000e+00 2.280000e+00 2.370000e+00 2.460000e+00
## [206] 2.550000e+00 2.640000e+00 2.730000e+00 2.820000e+00 2.910000e+00
## [211] 9.000000e-01 9.900000e-01 1.080000e+00 1.170000e+00 1.260000e+00
## [216] 1.350000e+00 1.440000e+00 1.530000e+00 1.620000e+00 1.710000e+00
## [221] 1.800000e+00 1.890000e+00 1.980000e+00 2.070000e+00 2.160000e+00
## [226] 2.250000e+00 2.340000e+00 2.430000e+00 2.520000e+00 2.610000e+00
## [231] 2.700000e+00 6.900000e-01 7.800000e-01 8.700000e-01 9.600000e-01
## [236] 1.050000e+00 1.140000e+00 1.230000e+00 1.320000e+00 1.410000e+00
## [241] 1.500000e+00 1.590000e+00 1.680000e+00 1.770000e+00 1.860000e+00
## [246] 1.950000e+00 2.040000e+00 2.130000e+00 2.220000e+00 2.310000e+00
## [251] 2.400000e+00 2.490000e+00 4.800000e-01 5.700000e-01 6.600000e-01
## [256] 7.500000e-01 8.400000e-01 9.300000e-01 1.020000e+00 1.110000e+00
## [261] 1.200000e+00 1.290000e+00 1.380000e+00 1.470000e+00 1.560000e+00
## [266] 1.650000e+00 1.740000e+00 1.830000e+00 1.920000e+00 2.010000e+00
## [271] 2.100000e+00 2.190000e+00 2.280000e+00 2.700000e-01 3.600000e-01
## [276] 4.500000e-01 5.400000e-01 6.300000e-01 7.200000e-01 8.100000e-01
## [281] 9.000000e-01 9.900000e-01 1.080000e+00 1.170000e+00 1.260000e+00
## [286] 1.350000e+00 1.440000e+00 1.530000e+00 1.620000e+00 1.710000e+00
## [291] 1.800000e+00 1.890000e+00 1.980000e+00 2.070000e+00 6.000000e-02
## [296] 1.500000e-01 2.400000e-01 3.300000e-01 4.200000e-01 5.100000e-01
## [301] 6.000000e-01 6.900000e-01 7.800000e-01 8.700000e-01 9.600000e-01
## [306] 1.050000e+00 1.140000e+00 1.230000e+00 1.320000e+00 1.410000e+00
## [311] 1.500000e+00 1.590000e+00 1.680000e+00 1.770000e+00 1.860000e+00
## [316] -1.500000e-01 -6.000000e-02 3.000000e-02 1.200000e-01 2.100000e-01
## [321] 3.000000e-01 3.900000e-01 4.800000e-01 5.700000e-01 6.600000e-01
## [326] 7.500000e-01 8.400000e-01 9.300000e-01 1.020000e+00 1.110000e+00
## [331] 1.200000e+00 1.290000e+00 1.380000e+00 1.470000e+00 1.560000e+00
## [336] 1.650000e+00 -3.600000e-01 -2.700000e-01 -1.800000e-01 -9.000000e-02
## [341] 1.110223e-16 9.000000e-02 1.800000e-01 2.700000e-01 3.600000e-01
## [346] 4.500000e-01 5.400000e-01 6.300000e-01 7.200000e-01 8.100000e-01
## [351] 9.000000e-01 9.900000e-01 1.080000e+00 1.170000e+00 1.260000e+00
## [356] 1.350000e+00 1.440000e+00 -5.700000e-01 -4.800000e-01 -3.900000e-01
## [361] -3.000000e-01 -2.100000e-01 -1.200000e-01 -3.000000e-02 6.000000e-02

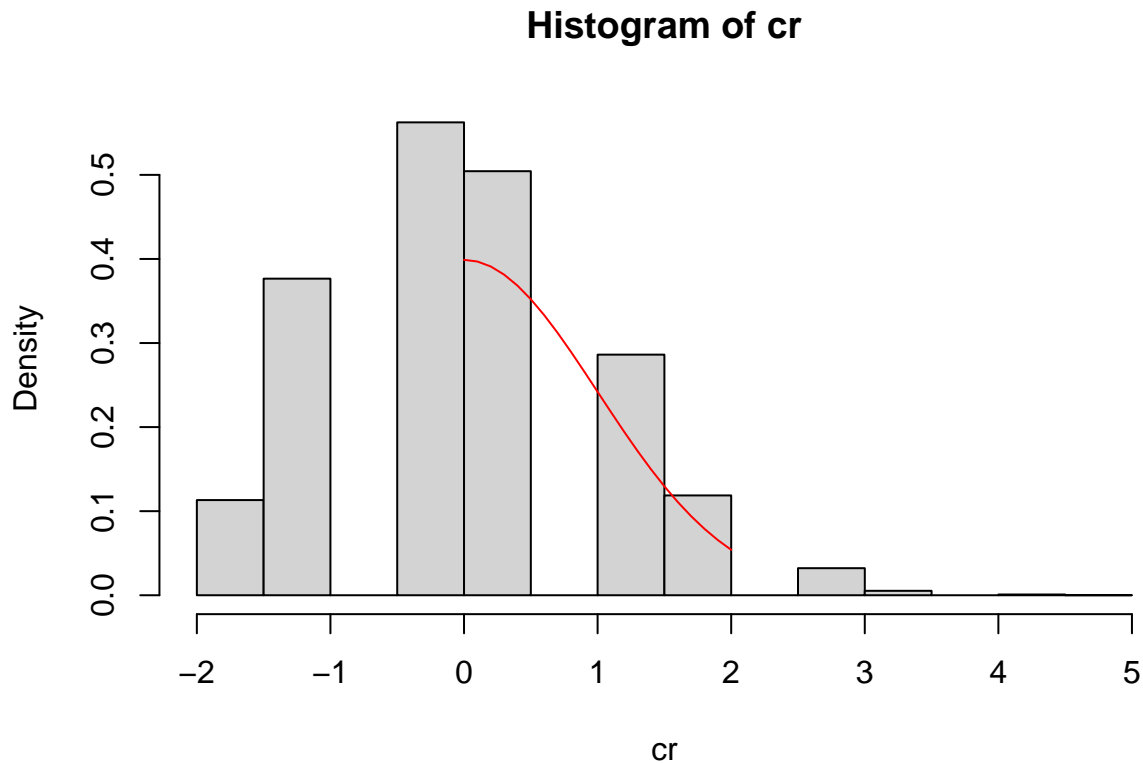
```

```

## [366] 1.500000e-01 2.400000e-01 3.300000e-01 4.200000e-01 5.100000e-01
## [371] 6.000000e-01 6.900000e-01 7.800000e-01 8.700000e-01 9.600000e-01
## [376] 1.050000e+00 1.140000e+00 1.230000e+00 -7.800000e-01 -6.900000e-01
## [381] -6.000000e-01 -5.100000e-01 -4.200000e-01 -3.300000e-01 -2.400000e-01
## [386] -1.500000e-01 -6.000000e-02 3.000000e-02 1.200000e-01 2.100000e-01
## [391] 3.000000e-01 3.900000e-01 4.800000e-01 5.700000e-01 6.600000e-01
## [396] 7.500000e-01 8.400000e-01 9.300000e-01 1.020000e+00 -9.900000e-01
## [401] -9.000000e-01 -8.100000e-01 -7.200000e-01 -6.300000e-01 -5.400000e-01
## [406] -4.500000e-01 -3.600000e-01 -2.700000e-01 -1.800000e-01 -9.000000e-02
## [411] 4.440892e-16 9.000000e-02 1.800000e-01 2.700000e-01 3.600000e-01
## [416] 4.500000e-01 5.400000e-01 6.300000e-01 7.200000e-01 8.100000e-01
## [421] -1.200000e+00 -1.110000e+00 -1.020000e+00 -9.300000e-01 -8.400000e-01
## [426] -7.500000e-01 -6.600000e-01 -5.700000e-01 -4.800000e-01 -3.900000e-01
## [431] -3.000000e-01 -2.100000e-01 -1.200000e-01 -3.000000e-02 6.000000e-02
## [436] 1.500000e-01 2.400000e-01 3.300000e-01 4.200000e-01 5.100000e-01
## [441] 6.000000e-01 -1.410000e+00 -1.320000e+00 -1.230000e+00 -1.140000e+00
## [446] -1.050000e+00 -9.600000e-01 -8.700000e-01 -7.800000e-01 -6.900000e-01
## [451] -6.000000e-01 -5.100000e-01 -4.200000e-01 -3.300000e-01 -2.400000e-01
## [456] -1.500000e-01 -6.000000e-02 3.000000e-02 1.200000e-01 2.100000e-01
## [461] 3.000000e-01 3.900000e-01 -1.620000e+00 -1.530000e+00 -1.440000e+00
## [466] -1.350000e+00 -1.260000e+00 -1.170000e+00 -1.080000e+00 -9.900000e-01
## [471] -9.000000e-01 -8.100000e-01 -7.200000e-01 -6.300000e-01 -5.400000e-01
## [476] -4.500000e-01 -3.600000e-01 -2.700000e-01 -1.800000e-01 -9.000000e-02
## [481] 2.220446e-16 9.000000e-02 1.800000e-01 -1.830000e+00 -1.740000e+00
## [486] -1.650000e+00 -1.560000e+00 -1.470000e+00 -1.380000e+00 -1.290000e+00
## [491] -1.200000e+00 -1.110000e+00 -1.020000e+00 -9.300000e-01 -8.400000e-01
## [496] -7.500000e-01 -6.600000e-01 -5.700000e-01 -4.800000e-01 -3.900000e-01
## [501] -3.000000e-01 -2.100000e-01 -1.200000e-01 -3.000000e-02 -2.040000e+00
## [506] -1.950000e+00 -1.860000e+00 -1.770000e+00 -1.680000e+00 -1.590000e+00
## [511] -1.500000e+00 -1.410000e+00 -1.320000e+00 -1.230000e+00 -1.140000e+00
## [516] -1.050000e+00 -9.600000e-01 -8.700000e-01 -7.800000e-01 -6.900000e-01
## [521] -6.000000e-01 -5.100000e-01 -4.200000e-01 -3.300000e-01 -2.400000e-01
## [526] -2.250000e+00 -2.160000e+00 -2.070000e+00 -1.980000e+00 -1.890000e+00
## [531] -1.800000e+00 -1.710000e+00 -1.620000e+00 -1.530000e+00 -1.440000e+00
## [536] -1.350000e+00 -1.260000e+00 -1.170000e+00 -1.080000e+00 -9.900000e-01
## [541] -9.000000e-01 -8.100000e-01 -7.200000e-01 -6.300000e-01 -5.400000e-01
## [546] -4.500000e-01 -2.460000e+00 -2.370000e+00 -2.280000e+00 -2.190000e+00
## [551] -2.100000e+00 -2.010000e+00 -1.920000e+00 -1.830000e+00 -1.740000e+00
## [556] -1.650000e+00 -1.560000e+00 -1.470000e+00 -1.380000e+00 -1.290000e+00
## [561] -1.200000e+00 -1.110000e+00 -1.020000e+00 -9.300000e-01 -8.400000e-01
## [566] -7.500000e-01 -6.600000e-01 -2.670000e+00 -2.580000e+00 -2.490000e+00
## [571] -2.400000e+00 -2.310000e+00 -2.220000e+00 -2.130000e+00 -2.040000e+00
## [576] -1.950000e+00 -1.860000e+00 -1.770000e+00 -1.680000e+00 -1.590000e+00
## [581] -1.500000e+00 -1.410000e+00 -1.320000e+00 -1.230000e+00 -1.140000e+00
## [586] -1.050000e+00 -9.600000e-01 -8.700000e-01 -2.880000e+00 -2.790000e+00
## [591] -2.700000e+00 -2.610000e+00 -2.520000e+00 -2.430000e+00 -2.340000e+00
## [596] -2.250000e+00 -2.160000e+00 -2.070000e+00 -1.980000e+00 -1.890000e+00
## [601] -1.800000e+00

```

```
lines (discr, dnorm (discr), col="red")
```



(ii) On se propose de faire varier les valeurs de n et p par une fonction. Reprenons toutes les lignes précédentes :

```
binnorm=function (n,p) {
q=1-p
bin=rbinom (50000,n,p)
cr=(bin-n*p)/sqrt(n*p*q)
hist (cr, probability=TRUE)
discr=seq(-3,3,0.01)
lines (discr,dnorm(discr), col="red")
}
```

On peut alors faire tourner la fonction avec différentes valeurs de n et p

(iii) On peut recommencer avec les fonctions de répartition :

```
binnormFDR=function (n,p){
q=1-p
bin=rbinom (50000,n,p)
cr=(bin-n*p)/sqrt(n*p*q)
plot (ecdf (cr))
discr=seq(-3,3,0.01)
lines (discr,pnorm(discr),col="red")
}
```

1.3 Lois normales, loi du χ^2 , loi de Fisher

On souhaite illustrer les liens entre les lois normales, la loi du χ^2 et la loi de Fisher.

- (i) On souhaite simuler quatre lois normales centrées réduites indépendantes X_1, X_2, X_3 et X_4 . On génère quatre tirages de 500 :

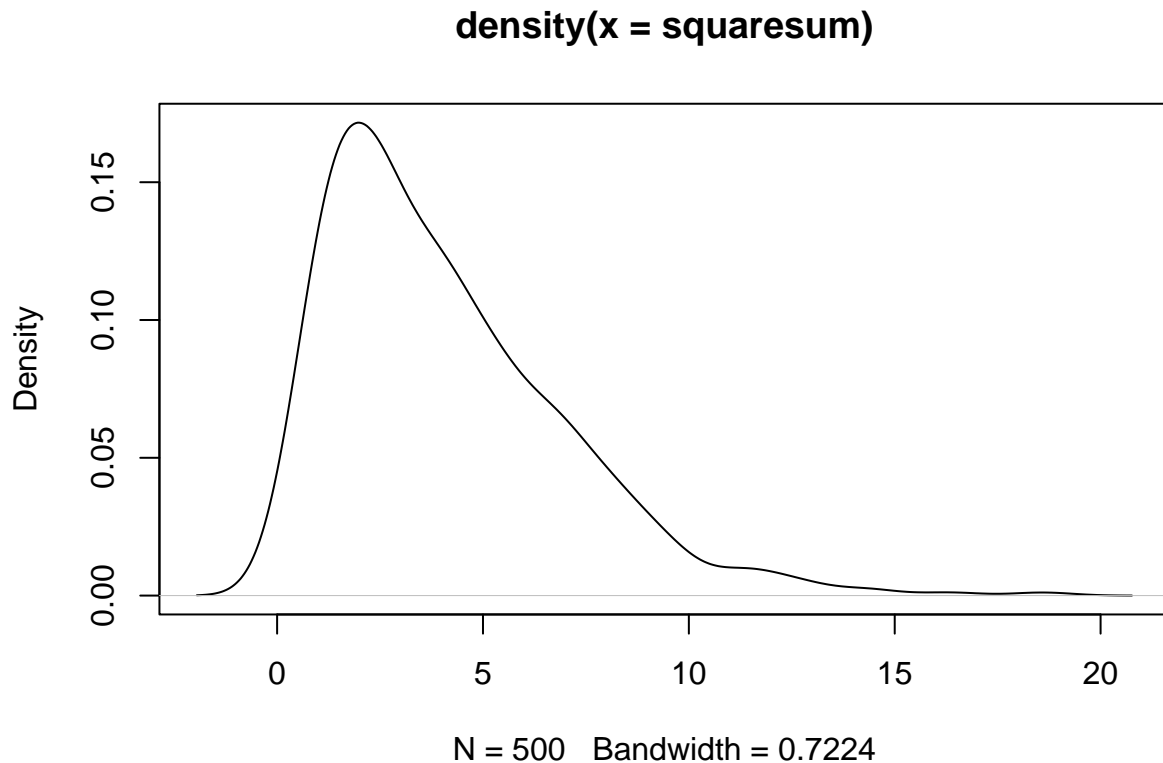
```
x1=rnorm(500)
x2=rnorm(500)
x3=rnorm(500)
x4=rnorm(500)
```

- (ii) Le théorème du cours annonce que $X_1^2 + X_2^2 + X_3^2 + X_4^2$ suit une loi du χ^2 à 4 degrés de liberté. La simulation de $X_1^2 + X_2^2 + X_3^2 + X_4^2$ est donnée par

```
squaresum=x1^2+x2^2+x3^2+x4^2
```

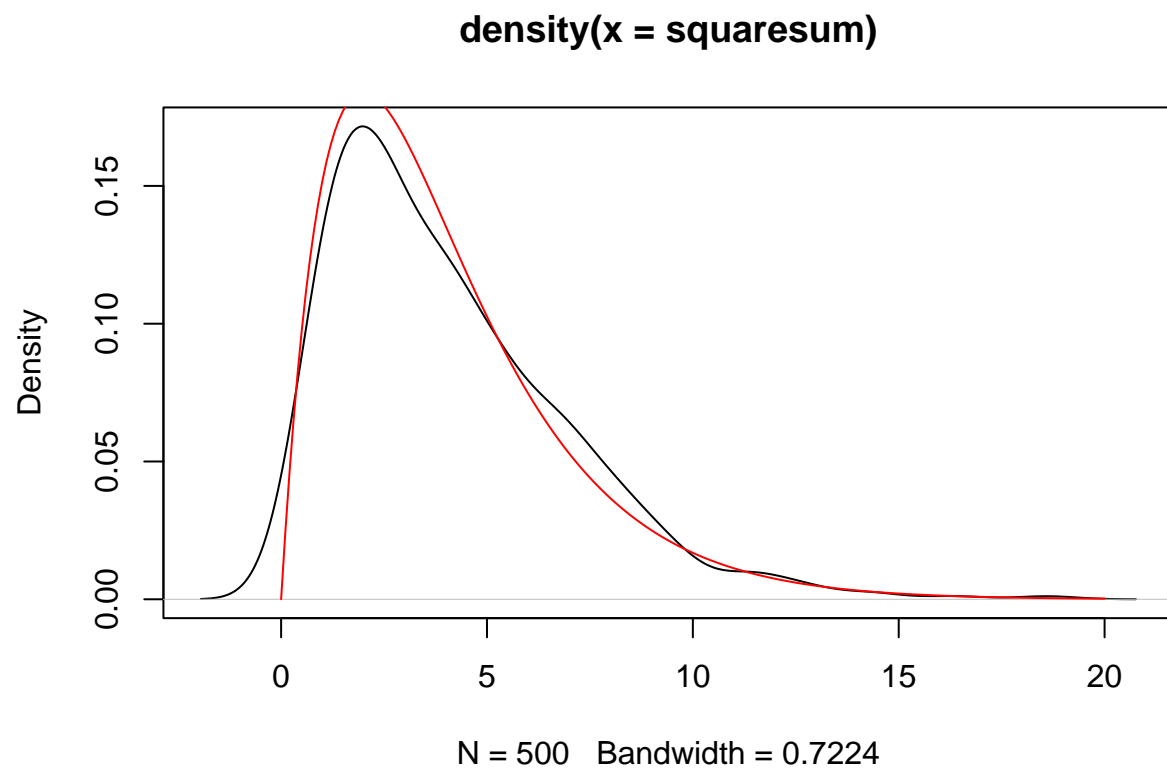
- (iii) On trace la densité de squaresum :

```
plot(density(squaresum))
```



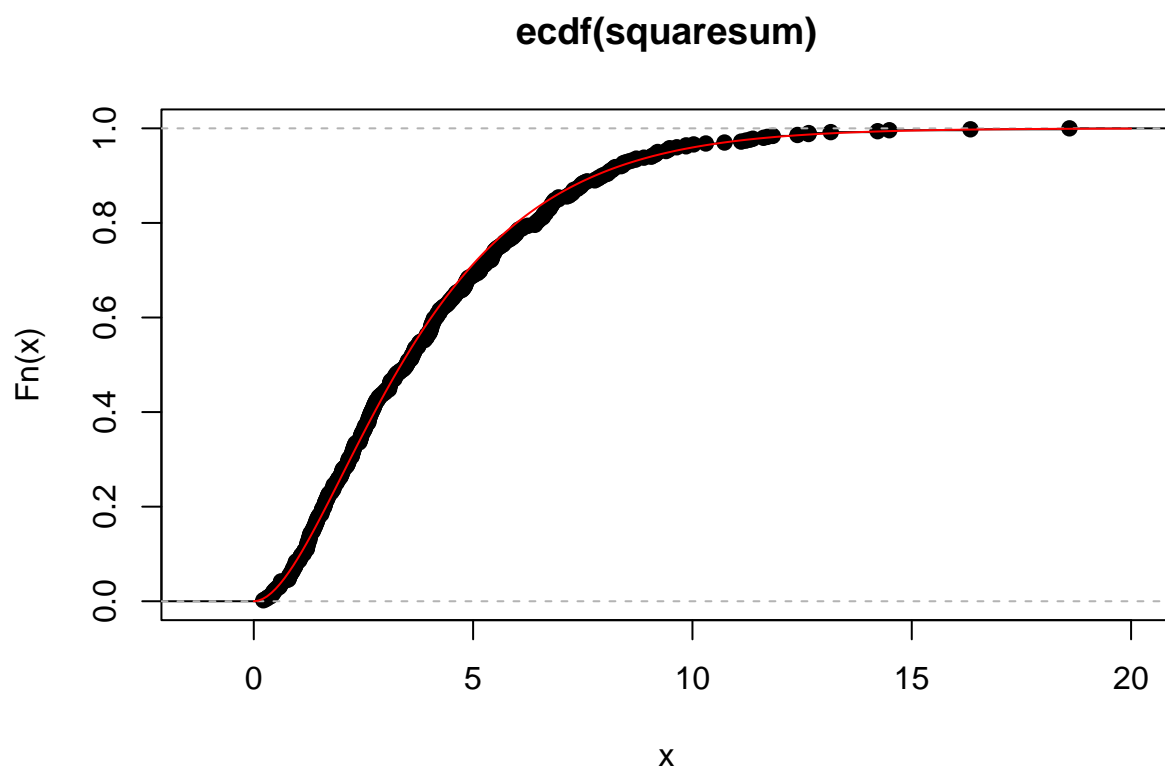
- (iv) La densité observée est non nulle sur $[0,20]$. On discrétise et on compare la densité observée à celle de la loi du χ^2_4 :

```
plot(density(squaresum))
discr=seq(0,20,0.1)
chi=dchisq(discr,4)
lines(discr, chi, col="red")
```



On peut également comparer les fonctions de répartition :

```
plot(ecdf (squaresum))  
lines(discr,pchisq (discr,4), col="red")
```



(v) Un autre résultat dit que le rapport

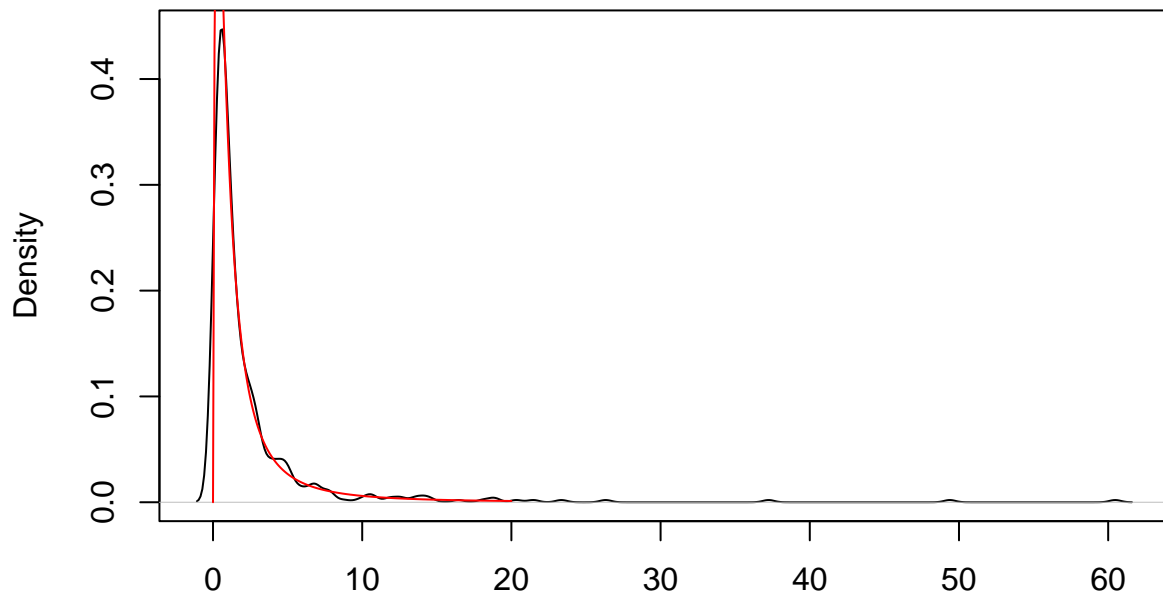
$$\frac{\chi_{d_1}^2/d_1}{\chi_{d_2}^2/d_2} \sim \mathcal{F}_{d_1, d_2}$$

=

On se propose de vérifier cette propriété pour $d_1 = 4$ en utilisant les 500 tirages de squaresum et de prendre les valeurs de d_2 entre 3 et 10. On initialise.

```
chi=rchisq (500,3)
plot (density ((squaresum/4)/(chi/3)))
lines(discr, df (discr,4,3), col="red")
```

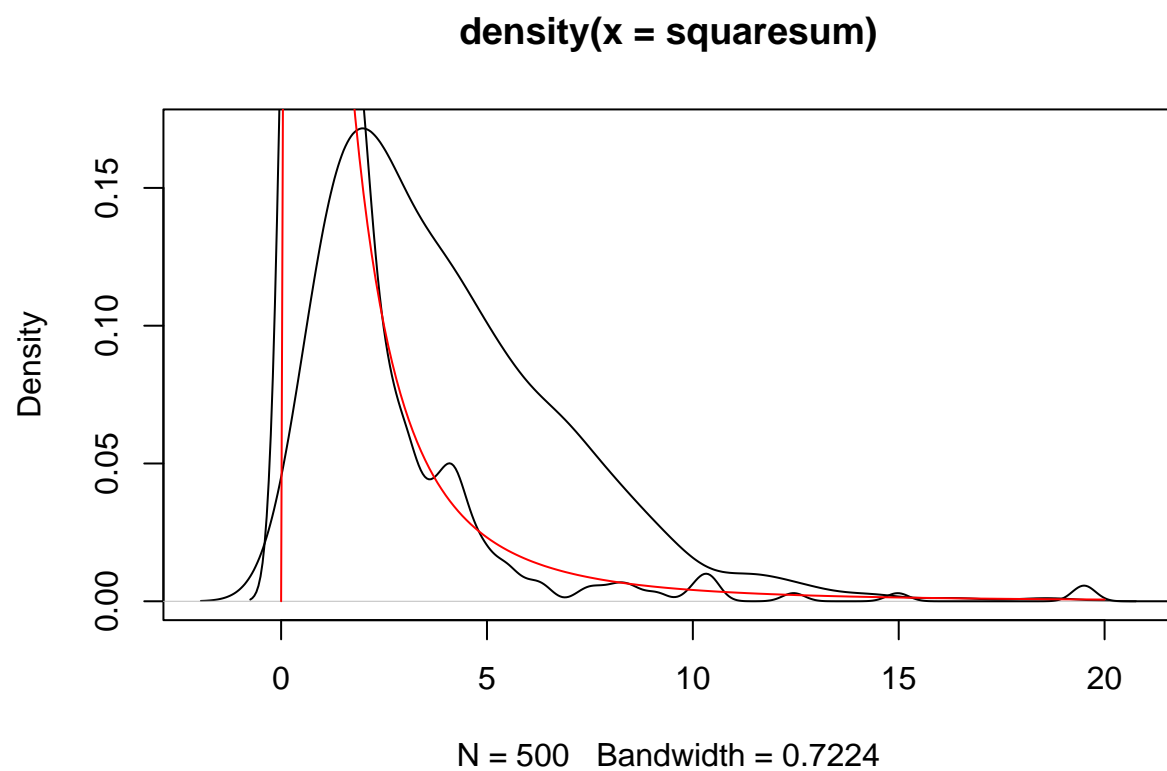
density(x = (squaresum/4)/(chi/3))



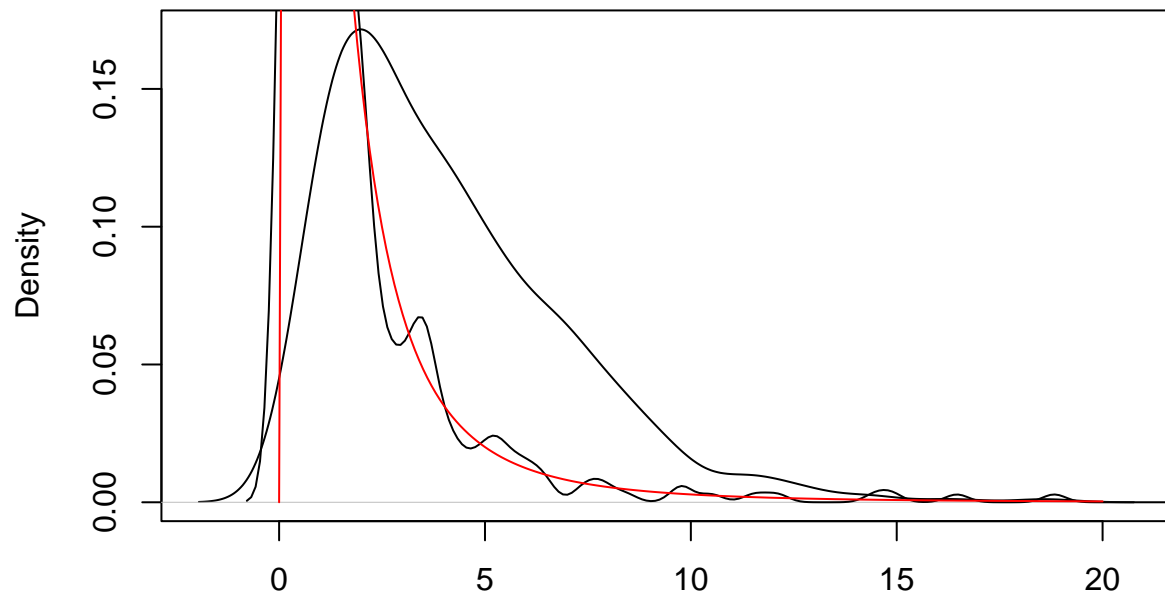
N = 500 Bandwidth = 0.3736

La commande plot ne peut s'exécuter qu'une fois sinon, on écrase le graphe précédent. On rajoutera des éléments avec la fonction `lines()`. Prêt pour une boucle for ?

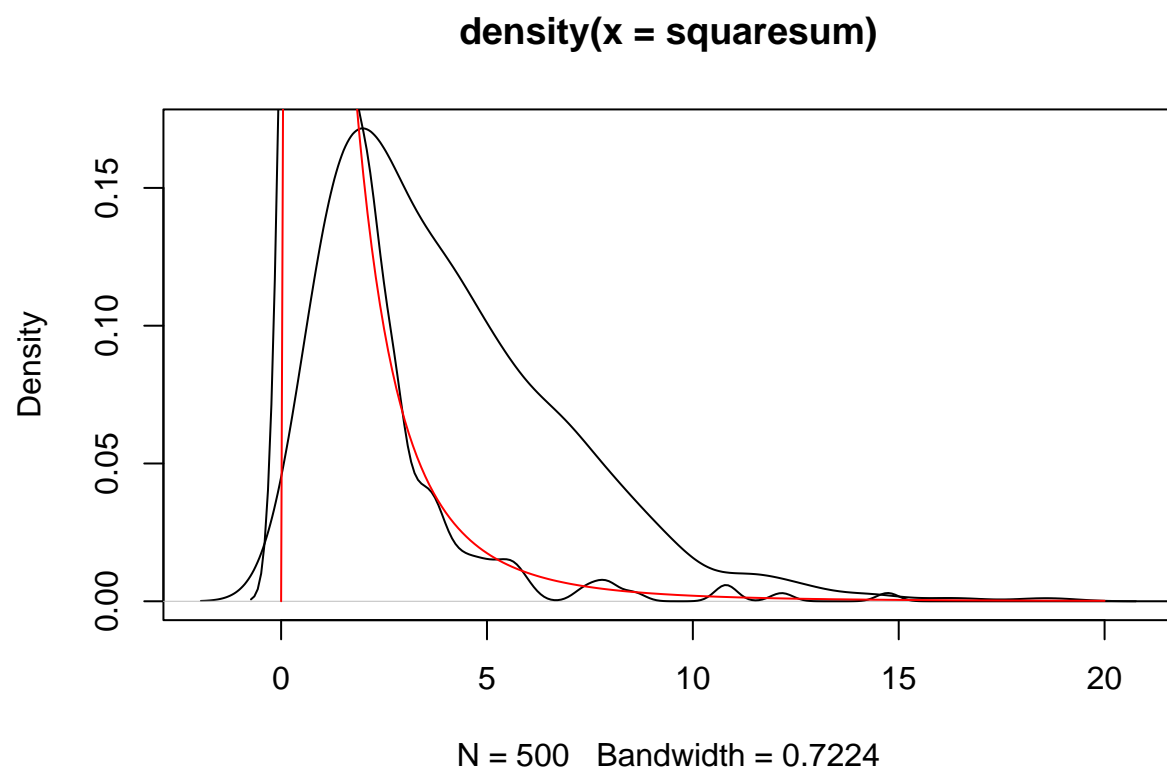
```
for (i in 4:10){  
  chi=rchisq (500,i)  
  plot(density(squaresum))  
  lines (density ((squaresum/4)/(chi/i)))  
  lines (discr,df (discr,4,i), col="red")  
}
```

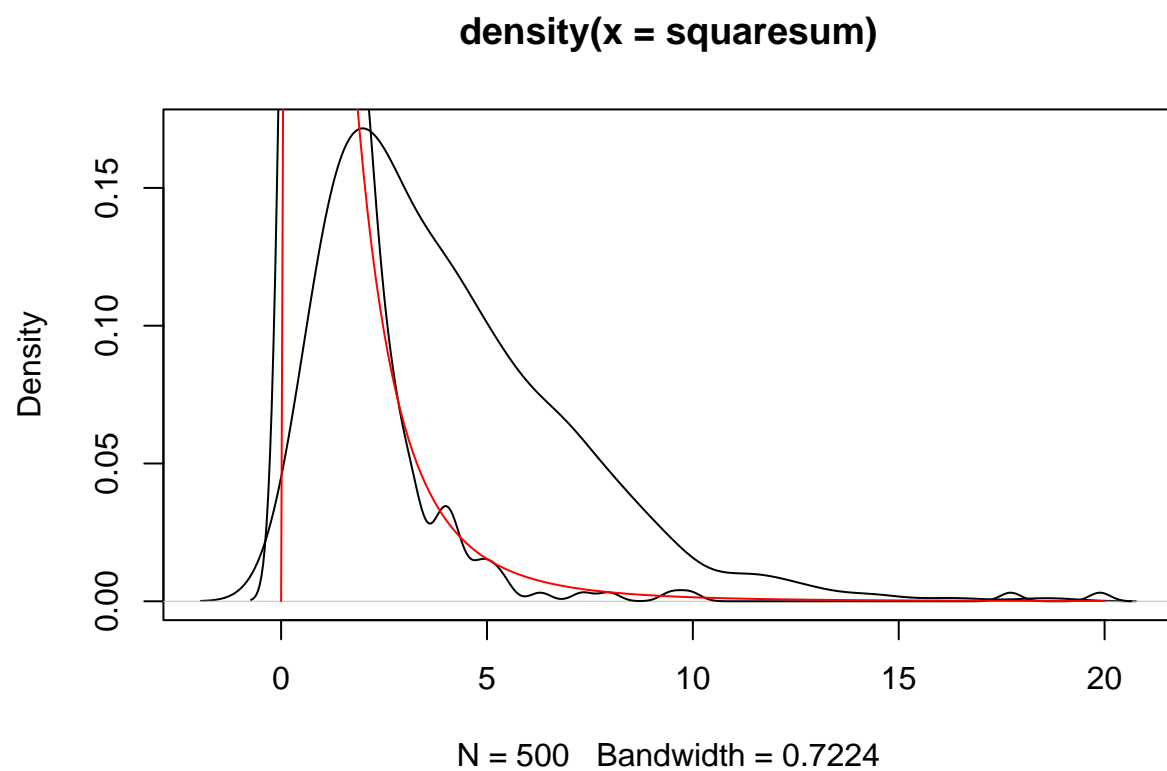


density(x = squaresum)

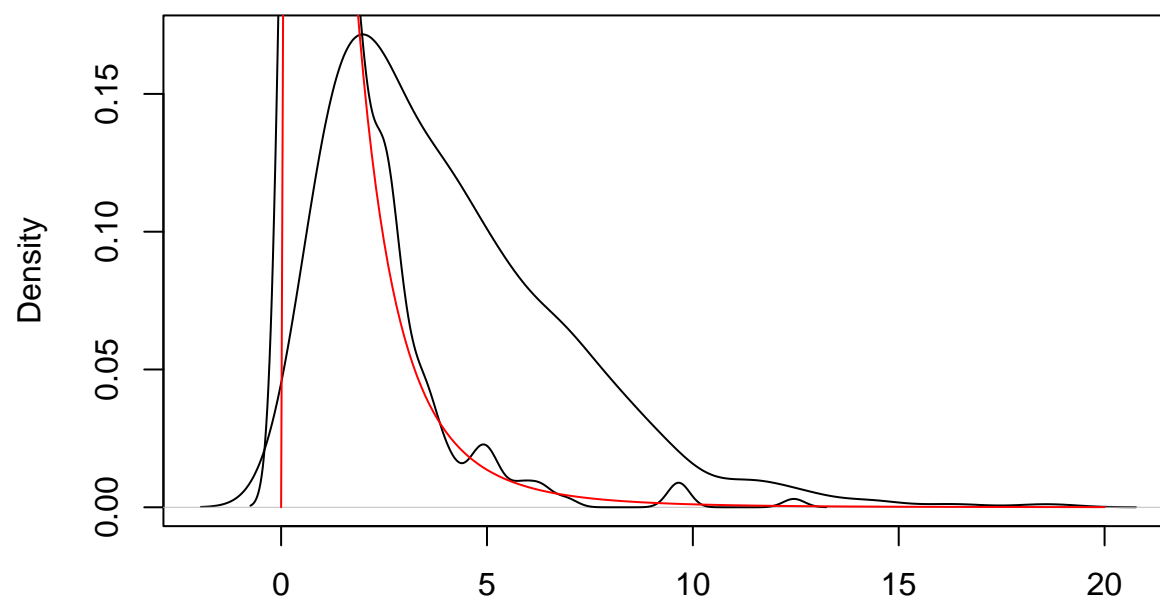


N = 500 Bandwidth = 0.7224

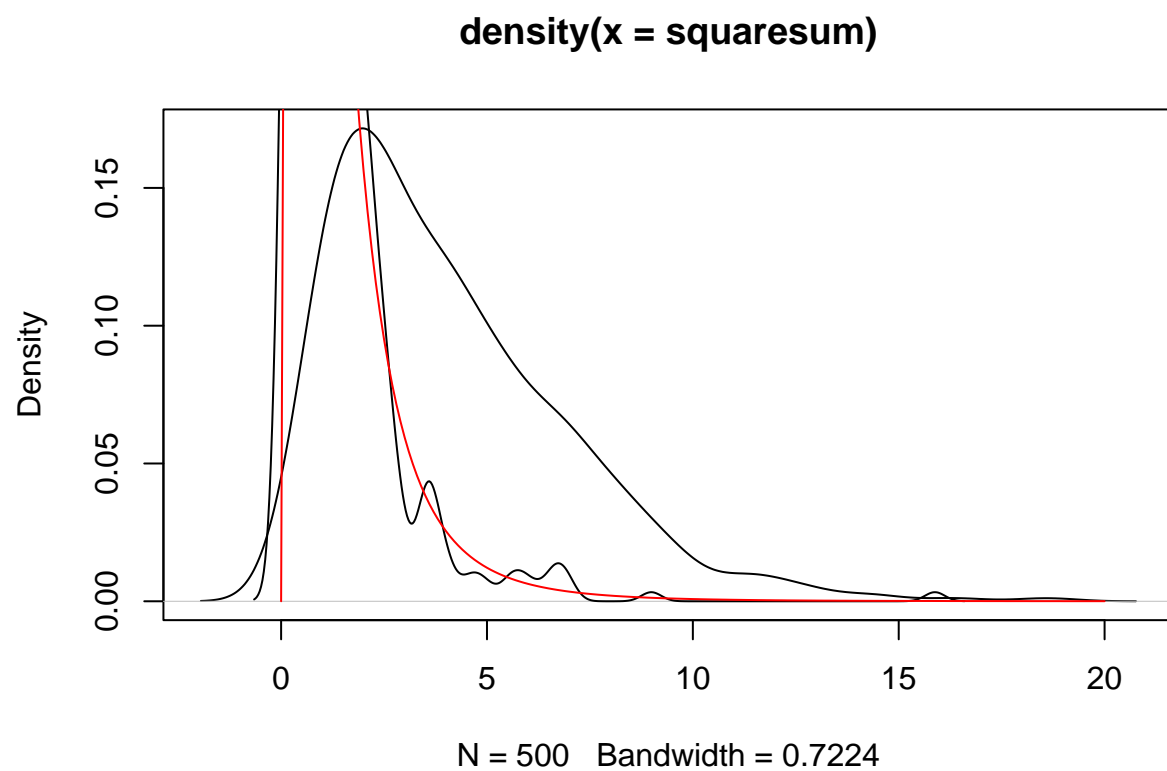


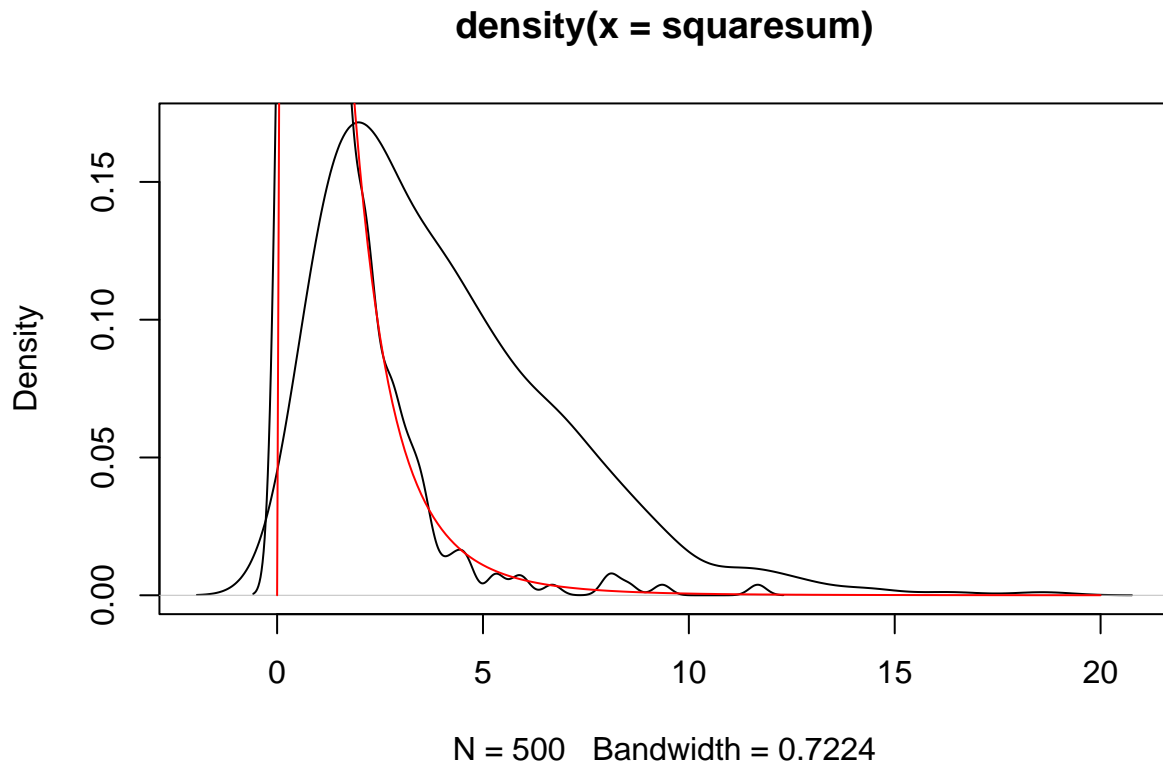


density(x = squaresum)



N = 500 Bandwidth = 0.7224





On voit que pour des grandes valeurs d_1 et d_2 , la loi de Fisher a une variance qui tend vers 0. Est-ce clair sur votre graphique ?

2 Intervalles de confiance

Dans cette dernière partie du TP, on souhaite illustrer par la simulation la construction de l'intervalle de confiance pour une proportion en suivant les lignes du cours.

On considère un stock de rouleaux de câbles de cuivre destiné à la livraison à une entreprise d'installation de lignes téléphoniques. On souhaite estimer la fréquence inconnue p des rouleaux de ce stock ayant une section inférieure à 0,4 mm de diamètre.

On prélève un échantillon aléatoire de 100 rouleaux dans ce stock. Ce stock est assez important pour que l'on puisse assimiler ce prélèvement à un tirage avec remise de 100 rouleaux.

On constate que seuls 4 rouleaux de cet échantillon ont une section inférieure à 0,4 mm. Ce sont les numéros 30,38,72,81.

On va principalement répondre aux questions suivantes.

- Donner une estimation ponctuelle de la fréquence inconnue f des rouleaux de ce stock ayant une section inférieure à 0,4 mm.
- Soit F la variable aléatoire qui, à tout échantillon de 100 rouleaux ainsi prélevé dans ce stock, associe la fréquence des rouleaux de cet échantillon ayant une section inférieure à 0,4 mm.

On suppose que F suit la loi normale de moyenne inconnue p et d'écart type $\sqrt{\frac{p(1-p)}{100}}$

* Déterminer un intervalle de confiance de la fréquence p avec le coefficient de confiance 95%.

* On considère l'affirmation suivante: "la fréquence p est obligatoirement dans l'intervalle de confiance". Cette affirmation est-elle vraie ?

2.1 Construction de l'intervalle de confiance

2.1.1 Construction de l'échantillon

On peut construire le vecteur d'observations des rouleaux défectueux. On fait un vecteur de 100 zéros.

```
x=rep(0,100)
```

On remplace les valeurs 30, 38, 72 et 81 par 1.

```
x[c (30,38,72,81)]=1
```

2.1.2 Estimation ponctuelle

L'estimation ponctuelle de la fréquence observée est la moyenne du vecteur x :

```
f=mean(x)
```

2.1.3 Loi X qui compte le nombre de rouleaux défectueux dans le lot de n pièces

On considère un échantillon de n pièces, on observe donc X_1, \dots, X_n des variables indépendantes et identiques à 2 issues la réussite (le rouleau est défectueux) avec probabilité p et l'échec. Pour compter, on attribuera donc à X_i la valeur 1 en cas de réussite et 0 sinon.

La variable X est la somme des X_i ,

$$X = \sum_{i=1}^n X_i$$

La variable X suit une loi binomiale $\mathcal{B}(n, p)$.

2.1.4 On pose $f_n = X/n$. Que représente f_n ?

La variable f_n est l'estimateur de la fréquence de pièces défectueuses.

2.1.5 Approximation de f_n

On suppose que f_n peut être approchée par une loi $N(m, \sigma)$ en vertu du théorème central limit. (pour n assez grand). On prendra

$$m_n = E(f_n) = E[X/n] = \frac{E(X)}{n} = \frac{np}{n} = p$$

et

$$\sigma_n = \sqrt{\text{Var}(f_n)} = \sqrt{\text{Var}(X/n)} = \sqrt{\frac{1}{n^2} \text{Var}(X)} = \sqrt{\frac{np(1-p)}{n^2}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{f(1-f)}{n}}$$

A supposer que $p = f = 0.04$, on calcule σ_n

```
sigma=sqrt((f*(1-f))/100)
```

2.1.6 Risque 5%

- On cherche a tel que

$$P(m_n - a\sigma_n \leq f_n \leq m_n + a\sigma_n) = 0.95$$

On prendra pour valeur de **a**

```
a=qnorm(0.975)
```

soit à peu près 1,96.

- Il y a donc un risque de se tromper de probabilité 0,05 si on affirme que

$$p - 1,96\sigma_n \leq f_n \leq p + 1,96\sigma_n.$$

- En faisant l'opération $-f_n - p$, puis en multipliant par -1, on déduit qu'au risque 0,05

$$f - 1,96\sigma_n \leq p \leq f + 1,96\sigma_n.$$

Intervalle de p avec un risque de 5% On a donc, au risque 5%,

$$f - 1,96\sigma \leq p \leq + 1,96\sigma.$$

Calculons les bornes

```
binf=f-a*sigma  
bsup=f+a*sigma
```

2.2 Simulations

On fait l'hypothèse suivante (invérifiable sauf à débiller tous les rouleaux de cuivre). **On suppose que la fréquence estimée f est effectivement correcte.**

Simulons 1000 valeurs de la variable X associées à 1000 échantillons.

```
X=rbinom (1000, 100,0.04)
```

Calculons les valeurs de f_n estimées pour chaque réalisation de X :

```
fn=X/100
```

2.2.1 Approximation de f_n par une loi normale

On a approché F_n par une loi normale de paramètres

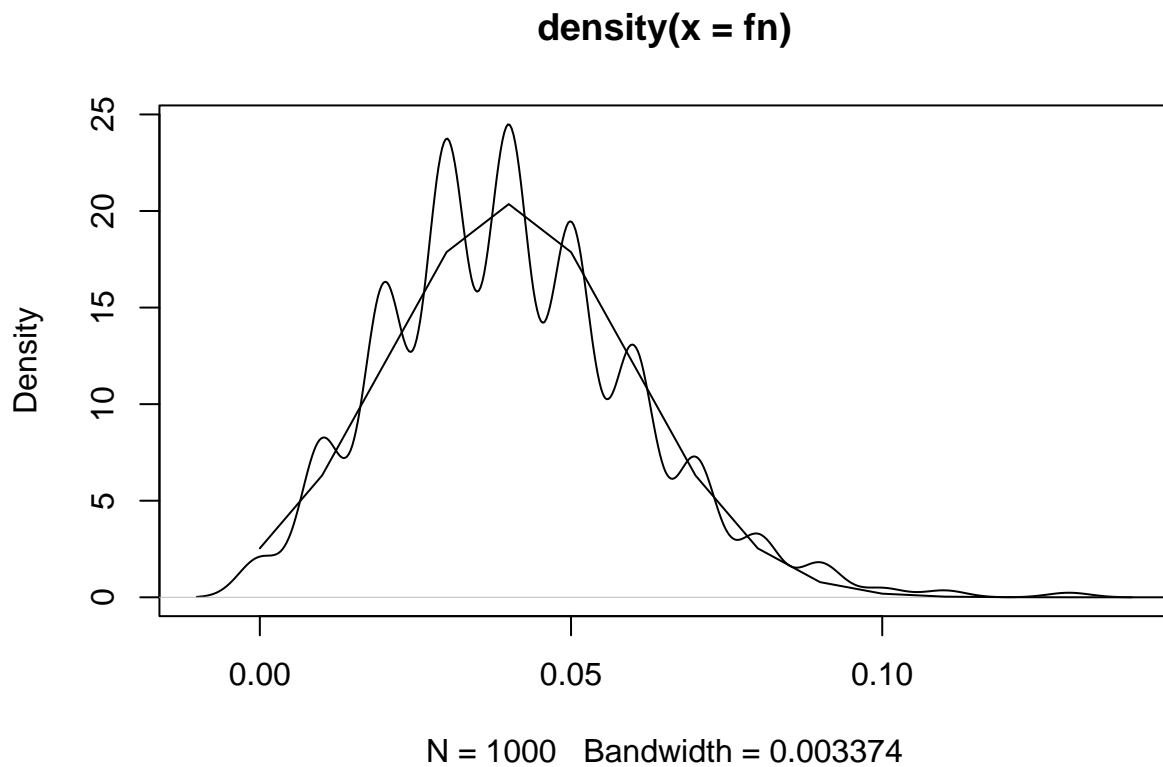
$$m_n = E(f_n) = p,$$

et

$$\sigma_n = \sqrt{Var(f_n)} = \sqrt{\frac{f(1-f)}{n}}$$

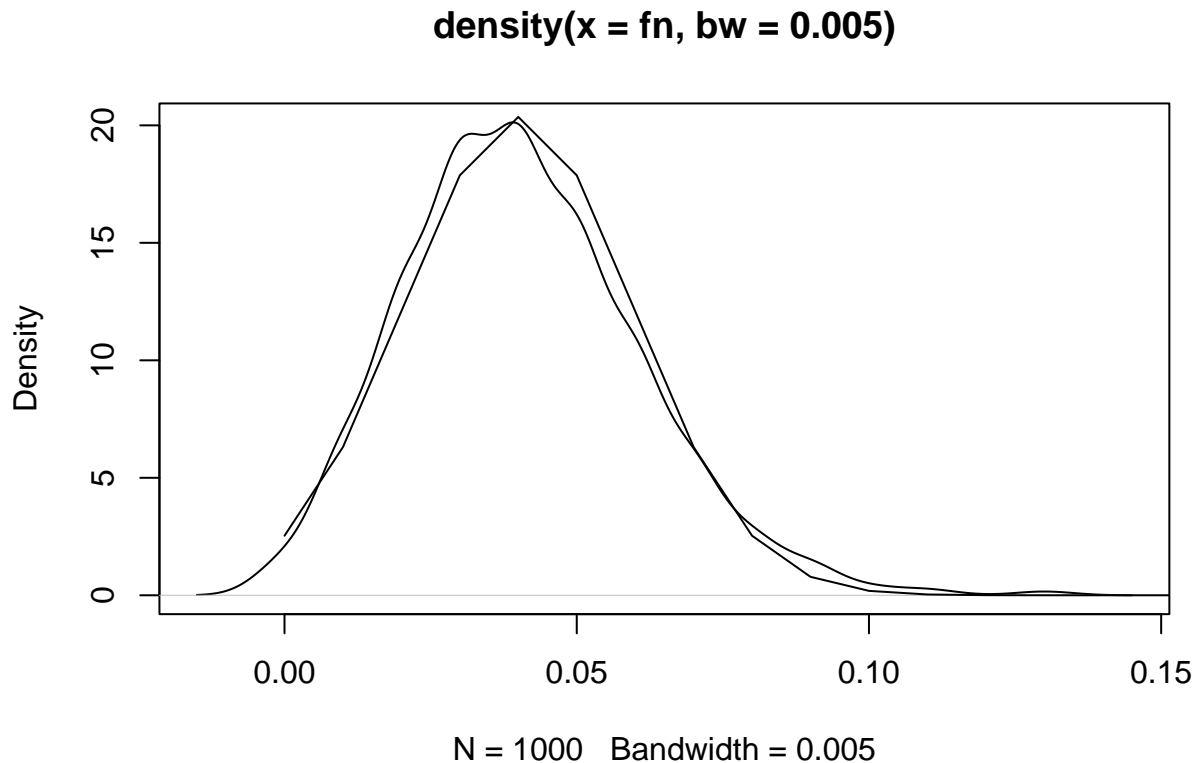
On compare la densité de f_n avec une loi normale $\mathcal{N}(f, \sigma_n)$.

```
plot (density(fn))  
discr=seq(0,1,0.01)  
lines (discr, dnorm (discr,f,sigma))
```



Ce n'est pas très joli, on peut élargie le lissage de la fonction `density()` pour éviter les trous entre groupement de valeurs

```
plot (density (fn, bw=0.005))  
discr=seq(0,1,0.01)  
lines (discr, dnorm(discr,f,sigma))
```



2.2.2 Intervalle de confiance

Cherchons le nombre de valeurs de `fn` qui sortent de l'intervalle de confiance 2.1.7.

```
sortie = fn [fn<binf | fn>bsup]
```

On devrait être aux alentours de 50 valeurs. On devrait même vraiment avoir exactement 50 valeurs. On remarque que ce n'est pas le cas (le plus souvent). Pourquoi ?!

2.2.3 Indépendance ?

Reprenons les lignes de codes pour créer une fonction qui, pour une proportion estimée p et un nombre de simulations N de n tirages nous donnera les simulations en dehors de l'intervalle de confiance.

```
confiance=function(p,n,N) {
  X=rbinom (N,n,p)
  fn=X/n
  sigma=sqrt((p* (1-p))/n)
  binf=p-1.96*sigma
  bsup=p+1.96*sigma
  sortie=fn [fn<binf | fn>bsup]
  return(sortie)
}
```

On peut jouer avec notre fonction, tester des valeurs de p, n, N . On s'aperçoit que quelque-soit la valeur de p , même centrale, quelque-soit la valeur de N et de n , on ne récupère pas toujours 5% de N valeurs.

C'est assez troublant puisqu'on a construit un intervalle de confiance à 5% qui devrait être assez fiable.

En réalité (pour avoir fait enchaîner des lancers de pièces en amphi quand on pouvait encore le faire), je n'ai jamais été mis en défaut. On peut avancer une explication: la machine a du mal à produire des tirages totalement indépendants. Il faut bien avoir conscience que l'aléa et l'indépendance sont des choses compliquées dans l'électronique.

Il y a fort à parier que si nous mettons en commun chacun une série de 10 simulations, 5% seront en dehors de l'intervalle. En multipliant les machines, nous devrions retrouver l'indépendance nécessaire.