Rapport : Estimation non paramétrique – Épreuve PRATIQUE (SOUS R).

NOM ET PRENOM : BERREKSI KHADIDJA

E-MAIL: BERREKSI.KHADIDJA@GMAIL.COM

M2-SPA.

PARTIE 1 : LOI DISCRETE FINIE

1. Générer par simulation un échantillon de taille N=50, 100, 500,1000 de la loi binomiale de paramètres (5, 0.4)

SYNTAXE :

```
p= 0.4 #probabilite de reussir
n=5 #nombre d'experience(essaies)

#Initialiser la taille de chaque echantillon
N1=50    N2=100    N3=500    N4=1000

S1=rbinom(N1,5,0.4)
S2=rbinom(N2,5,0.4)
S3=rbinom(N3,5,0.4)
S4=rbinom(N4,5,0.4)
```

RESULTATS :

```
> S1=rbinom(N1,5,0.4)
> S1
 [1] 2 1 2 1 3 2 1 2 2 3 1 0 2 5 1 1 3 0 1 2 1 1 3 3 1 2 1 1 2 2 2 2 2 2 3 1 2 3
[39] 3 2 3 2 1 1 0 2 2 1 3 4
> S2=rbinom(N2,5,0.4)
> S2
  [1] 0 2 3 0 0 0 2 2 1 3 3 1 3 1 2 1 3 3 1 0 3 1 4 5 1 2 2 2 2 0 3 2 2 1 1 3 0
 [38] 1 3 2 3 1 1 1 2 1 2 2 1 2 2 2 1 2 2 2 3 0 2 1 3 1 0 4 2 1 1 2 4 2 2 5 1 2
 [75] 2 1 2 1 3 3 3 2 2 3 2 3 2 3 2 1 3 1 2 1 2 0 2 3 3 3

> S3=rbinom(N3,5,0.4)
> S3
  [1] 0 3 0 0 1 1 2 5 2 1 1 4 0 1 3 4 1 1 2 2 2 1 1 1 1 2 3 2 2 2 2 3 3 4 3 2 4
 [38] 4 0 0 0 0 2 3 0 1 1 1 3 0 0 1 3 1 2 1 1 4 3 2 2 0 2 4 4 4 2 0 3 2 1 5 2 2
 [75] 2 2 1 4 2 3 3 2 1 2 3 2 2 0 2 2 3 1 1 0 0 3 2 1 3 3 1 2 1 1 2 2 2 3 1 4 4
[112] 1 3 2 1 3 4 2 1 3 1 1 1 2 2 2 3 3 4 1 3 2 1 1 3 1 2 3 4 2 0 3 2 4 3 1 2 3
[149] 1 0 2 2 2 0 2 3 4 3 2 2 2 2 2 1 2 1 0 1 3 1 3 2 2 1 3 1 2 2 1 1 2 3 2 4 1
[186] 2 1 3 1 4 1 1 2 2 3 2 1 2 4 0 2 3 3 1 3 1 3 2 2 1 0 1 0 2 1 0 2 3 2 1 0 3
[223] 2 2 3 1 2 2 3 3 2 2 4 1 2 2 2 2 4 1 1 3 2 2 3 2 2 1 4 2 3 1 3 2 3 3 2 1 1
[260] 2 1 2 3 3 1 2 2 3 0 2 5 3 0 1 2 2 2 2 3 2 3 1 1 3 2 4 1 2 2 1 2 2 2 2 3 2
[297] 1 1 2 2 3 2 0 3 2 3 3 1 3 3 3 4 2 2 1 2 2 2 2 1 3 1 1 2 2 4 1 1 3 2 2 2 1
[334] 2 1 2 3 2 2 3 1 5 1 3 2 3 1 4 0 4 3 3 2 2 3 2 2 2 3 3 1 1 3 1 2 2 1 4 2 4
[371] 2 4 3 4 3 3 1 1 0 3 2 3 2 4 2 2 2 2 4 1 3 2 2 3 1 0 2 2 1 0 2 4 2 2 2 2 2
[408] 3 3 2 4 1 2 3 3 1 4 1 3 2 1 1 3 2 3 1 2 1 1 1 3 0 2 1 3 1 3 2 2 2 2 1 4 3
[445] 1 1 1 3 1 1 2 4 2 2 2 2 0 2 2 2 2 1 0 2 2 1 2 1 1 2 3 2 3 2 1 0 2 3 2 2 3
[482] 4 3 2 2 1 3 2 0 1 1 1 2 3 2 3 0 0 3 1
```

```
> S4=rbinom(N4,5,0.4)
> S4
   [1] 1 2 2 4 3 3 2 1 1 3 1 2 4 4 3 1 3 2 1 5 3 2 2 3 3 1 3 2 1 1 3 2 1 1 2 2 3
  [38] 3 1 1 3 3 1 5 4 2 2 2 1 1 2 3 1 3 2 1 3 3 1 3 2 2 3 2 1 3 2 3 1 1 0 1 3 2
  [75] 2 1 3 2 4 2 2 2 2 3 3 2 5 1 3 1 1 1 4 0 1 1 2 3 2 1 2 3 4 2 4 0 2 0 3 2 3
 [112] 2 1 0 4 2 1 3 1 1 2 1 1 3 2 0 2 3 3 3 2 2 3 3 0 1 1 5 1 3 2 2 4 2 1 2 2 1
 [149] 2 2 3 0 3 1 5 1 1 1 0 2 2 3 1 2 1 3 2 2 2 3 3 1 3 0 4 2 2 1 2 1 1 2 3 0 0
 [186] 3 1 2 4 3 2 1 2 1 2 2 1 2 1 2 1 3 1 1 0 1 1 3 4 1 1 2 1 1 2 2 1 2 4 0 1 1
 [223] 2 3 1 2 2 1 1 2 2 3 3 3 3 0 1 5 0 3 3 2 1 3 0 2 1 0 3 4 2 2 3 3 3 2 2 2 2
 [260] 2 0 1 3 3 2 2 1 1 0 1 1 3 0 1 3 0 1 1 2 2 1 0 2 3 1 1 2 3 3 2 2 2 2 3 2 0
 [297] 1 3 0 3 1 2 3 1 2 2 3 2 3 2 2 5 1 2 1 1 2 2 1 0 4 3 0 2 3 3 2 2 1 2 2 1 3
 [334] 3 0 2 3 3 1 4 4 2 3 2 0 4 2 3 4 1 3 0 4 0 3 1 0 4 1 2 2 1 2 3 3 4 3 2 5 2
 [371] 0 0 3 1 3 2 2 2 1 1 1 1 2 2 3 2 2 2 0 2 1 0 1 1 3 3 3 3 3 3 4 3 2 1 3 3 1
 [408] 1 3 0 0 3 1 3 2 1 2 1 2 2 2 1 0 2 1 3 3 3 1 4 4 3 5 3 1 1 1 1 1 4 2 3 3 4
 [445] 4 4 1 1 4 1 2 1 0 1 2 2 1 0 2 1 5 1 1 1 1 2 2 1 3 3 2 0 2 0 1 2 2 2 3 2 1
 [482] 4 4 3 2 1 2 1 4 5 1 2 3 3 2 2 2 3 2 2 1 1 1 1 2 1 3 1 4 0 2 4 2 4 2 2 0 2
 [519] 1 1 3 1 1 3 3 1 2 1 2 2 3 3 0 1 3 2 3 2 2 2 0 2 2 1 3 1 1 2 3 1 1 2 1 4 2
 [556] 3 2 3 1 1 2 2 1 4 2 0 1 3 2 2 1 3 1 1 4 3 3 2 1 0 2 3 1 0 3 2 1 1 3 4 3 3
 [593] 2 3 2 4 1 4 2 0 4 2 2 4 2 3 1 2 1 4 4 2 2 2 2 3 3 3 1 2 1 3 4 1 2 1 1 2 2
 [630] 0 3 3 3 3 2 1 1 1 1 1 1 2 1 4 1 3 1 2 1 0 3 0 1 1 1 4 3 2 2 1 2 0 0 4 0 2
 [667] 3 2 1 2 2 1 2 1 3 2 2 5 4 1 2 2 1 1 1 1 2 1 3 0 1 1 1 3 1 2 0 0 2 1 4 2 3
 [704] 1 1 1 1 3 2 1 4 2 2 2 3 1 3 2 2 0 1 1 2 2 2 3 2 1 1 1 3 0 3 2 1 2 1 2 0 2
 [741] 2 2 2 1 2 0 2 2 3 2 1 2 2 2 3 4 2 0 4 2 2 2 3 2 1 1 2 1 2 2 2 2 3 2 3 3 3
 [778] 1 0 0 1 2 3 2 0 1 3 2 3 1 1 2 2 1 3 3 3 1 1 1 2 2 1 1 1 1 1 2 3 3 3 2 2 1
 [815] 1 3 1 4 2 3 3 2 3 3 2 1 2 2 1 3 1 0 2 2 2 2 3 2 1 5 4 2 0 2 2 4 0 3 0 2 2
 [852] 4 1 2 1 1 3 1 2 2 2 2 2 1 2 2 1 4 1 1 1 3 0 2 2 2 4 1 1 4 1 2 3 3 2 1 1 2
 [889] 1 3 1 1 3 3 2 2 0 0 3 2 2 2 3 3 5 2 2 2 0 0 2 2 1 2 4 2 1 1 2 0 2 2 2 2 0
 [926] 2 2 3 3 2 2 2 4 3 1 1 0 3 2 1 1 1 4 3 4 3 0 1 1 3 2 3 0 4 2 3 3 3 0 1 2 0
 [963] 4 3 2 1 2 3 3 2 2 0 1 3 1 1 3 2 3 1 4 2 0 2 1 1 3 3 3 1 2 1 2 2 2 4 3 1 1
[1000] 3
```

2. Calcul des fréquences empiriques :

SYNTAXE :

```
#les frequences empiriques
fe1=table(S1)/N1
fe2=table(S2)/N2
fe3=table(S3)/N3
fe4=table(S4)/N4
```

RESULTATS :

```
> #les frequences empiriques
> fe1=table(S1)/N1
> fe1
S1
   0    1    2    3    4    5
0.06 0.32 0.38 0.20 0.02 0.02
> fe2=table(S2)/N2
> fe2
S2
   0    1    2    3    4    5
0.10 0.26 0.36 0.23 0.03 0.02
> fe3=table(S3)/N3
> fe3
S3
    0     1     2     3     4     5
0.078 0.250 0.368 0.216 0.080 0.008
> fe4=table(S4)/N4
> fe4
S4
    0     1     2     3     4     5
0.086 0.283 0.326 0.220 0.071 0.014
>
```

3. Calcul des fonctions de répartition empiriques (en utilisant la libraire stats) :

SYNTAXE :

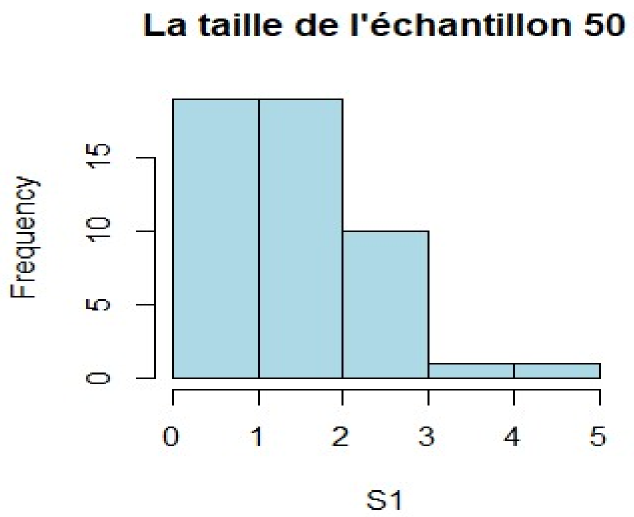
```r
#Fonctions de repartition empiriques
#pour ce faire on utilise le package stats
install.packages("stats")
library(stats)

ecdf_1 <- ecdf(S1)
ecdf_2 <- ecdf(S2)
ecdf_3 <- ecdf(S3)
ecdf_4 <- ecdf(S4)

```

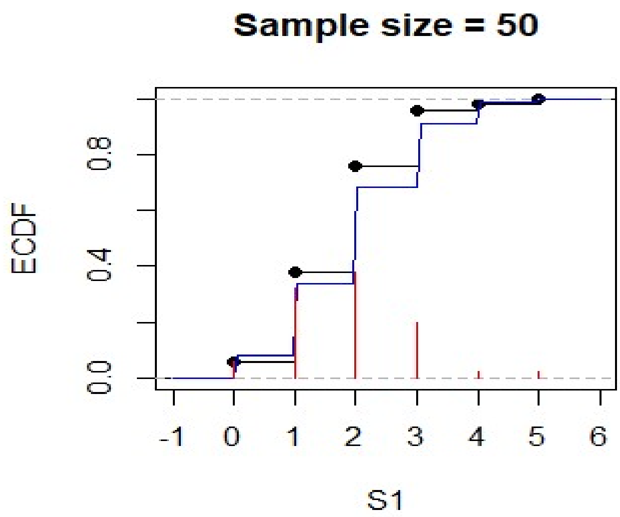
RESULTATS :

```
> ecdf_1
Empirical CDF
Call: ecdf(S1)
 x[1:6] =      0,      1,      2,  ...,      4,      5
> ecdf_2
Empirical CDF
Call: ecdf(S2)
 x[1:6] =      0,      1,      2,  ...,      4,      5
> ecdf_3
Empirical CDF
Call: ecdf(S3)
 x[1:6] =      0,      1,      2,  ...,      4,      5
> ecdf_4
Empirical CDF
Call: ecdf(S4)
 x[1:6] =      0,      1,      2,  ...,      4,      5
>
```

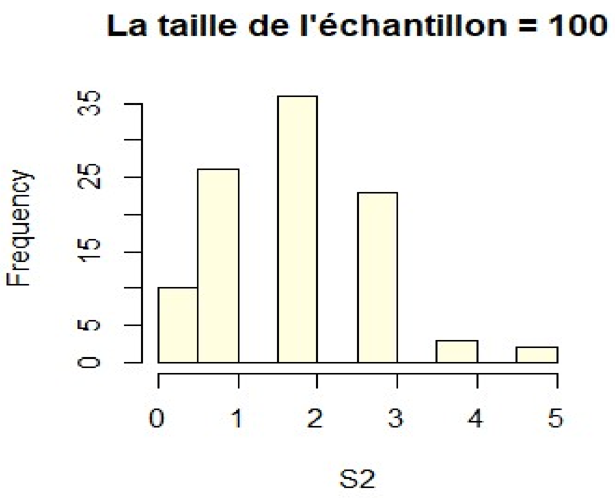
4. Représentations Graphiques :

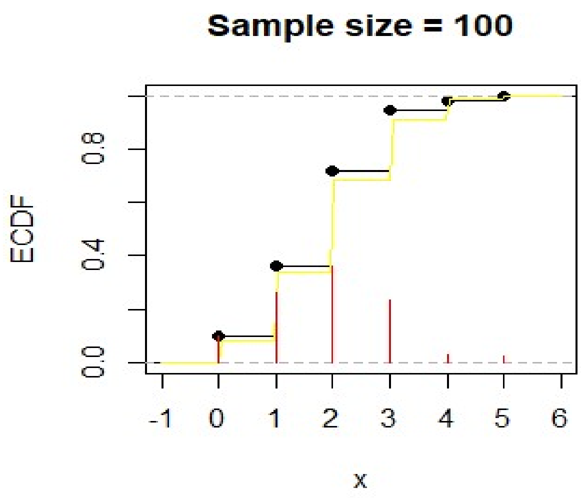
ECHANTILLON DE TAILLE 50 :



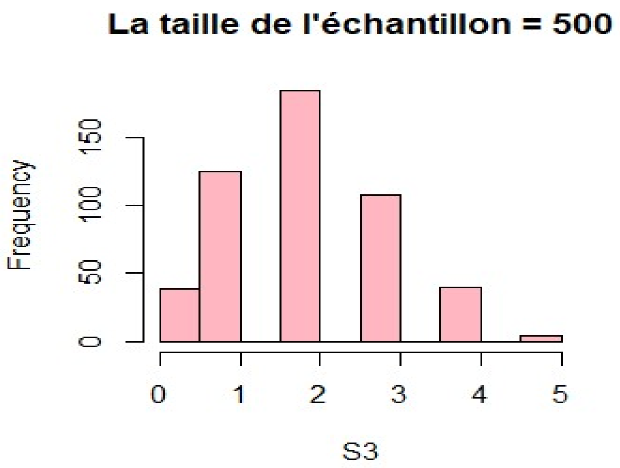


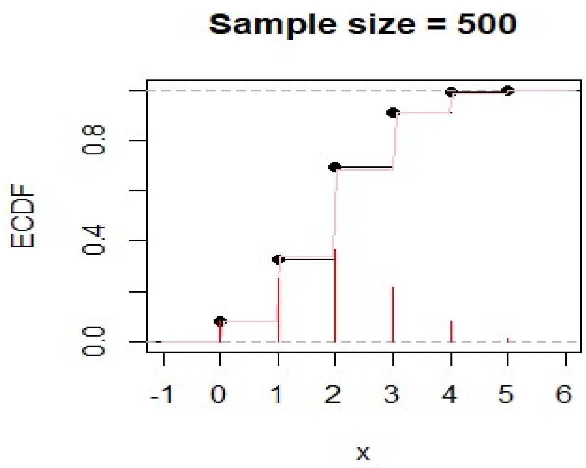
ECHANTILLON DE TAILLE 100 :



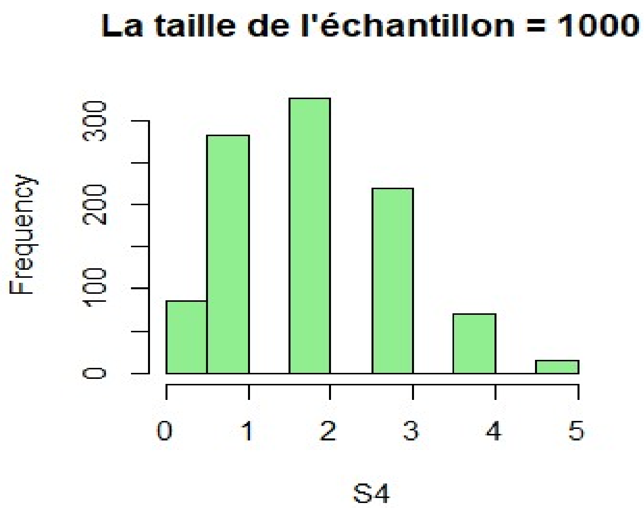


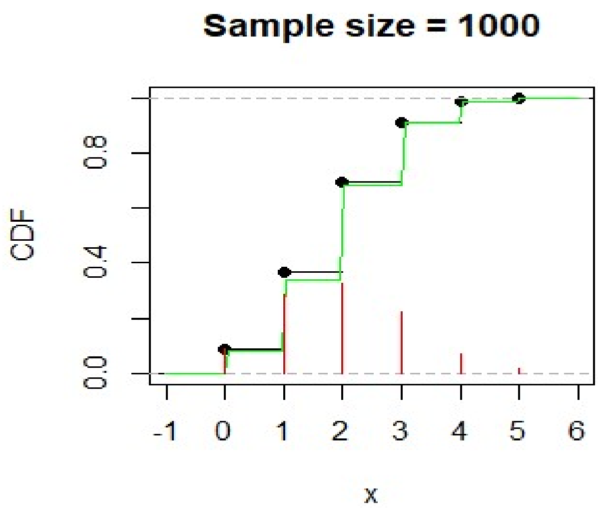
ECHANTILLON DE TAILLE 500 :





ECHANTILLON DE TAILLE 1000 :





COMPARAISON AVEC LES EQUIVALENTS THEORIQUES :

Les graphiques produits nous montreront les fréquences empiriques et la fonction de répartition empirique en noir et bleu, respectivement, comparées aux résultats théoriques en (bleu, jaune, rose et vert). Nous pouvons remarquer que les résultats empiriques s'approchent de plus en plus des résultats théoriques lorsque la taille de l'échantillon augmente.

PARTIE 2 : LOI ABSOLUMENT CONTINUE

LOI EXPONENTIELLE

1. Générer par simulation un échantillon de taille N=50, 100, 500,1000 de la loi Exponentielle

SYNTAXE :

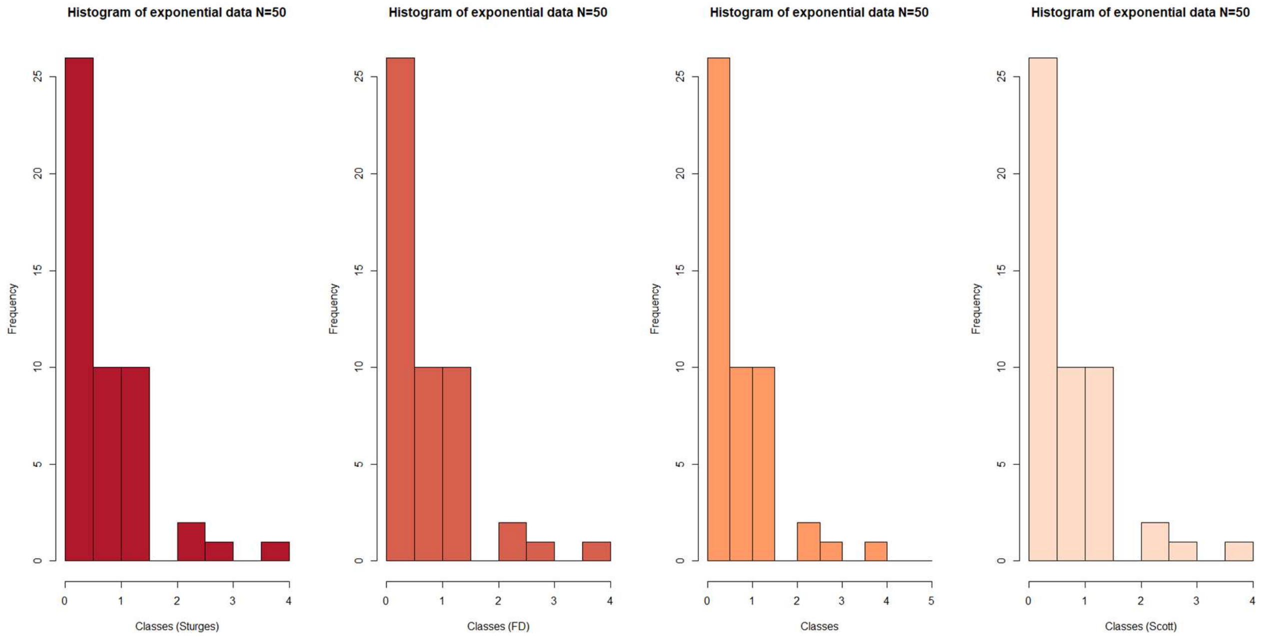
```
#Loi Exponentielle
e1<- rexp(N1)
e2<- rexp(N2)
e3<- rexp(N3)
e4<- rexp(N4)
```

RESULTATS :

```
> e1<- rexp(N1)
> e1
 [1] 1.043512089 1.255112908 0.049383917 0.335513546 0.417209160 0.288332555 0.693220172
 [8] 0.918692809 1.278122254 0.482364119 0.013967506 0.020610885 0.333097997 0.129841488
[15] 1.038235194 1.044212850 0.259770445 1.122780483 0.207115459 0.472699836 0.008123001
[22] 2.773159712 0.908646877 0.002067508 2.274335324 0.623662234 0.539556301 0.462817283
[29] 0.134778550 0.906162170 0.339295636 1.404413577 0.550414765 1.020748994 0.534056354
[36] 1.349646770 0.035845573 3.943345105 0.161797134 0.470516124 0.138600381 0.447673029
[43] 2.439272275 0.150310635 0.071868092 1.332061590 0.072122218 0.004949112 0.656142754
[50] 0.830802131
> e2<- rexp(N2)
> e2
  [1] 1.540015574 1.219293173 0.608322611 2.642699530 0.090942669 1.439239669 0.353047055
  [8] 0.478523049 0.509435654 0.080876675 0.812010527 3.337993133 1.111734637 0.488387968
 [15] 0.025793838 2.288222855 1.979519065 0.101668576 0.424728331 0.766833051 3.268707685
 [22] 0.860683162 0.217129563 1.099622890 1.051754841 3.593663296 0.143002293 1.441196927
 [29] 0.412540485 0.445436367 1.557224212 1.601629227 0.461209798 2.163397660 1.710078750
 [36] 3.129614681 1.280453992 0.457813668 0.640927513 1.877036836 1.636995181 1.010192976
 [43] 1.123196970 1.709853951 1.019320486 0.592219084 0.229739793 0.536385866 0.247481521
 [50] 2.589275656 0.380277138 0.265767566 1.669172713 0.248661484 0.756213632 0.325366379

> e3<- rexp(N3)
> e3
  [1] 0.041805141 0.410859877 0.724295692 0.603870856 0.238353225 0.369521169 1.250587142
  [8] 1.375907146 0.405352217 3.355858244 0.056650138 1.134350515 0.119578188 1.030242036
 [15] 0.747113942 0.175918602 0.206193318 0.377547160 0.558444507 1.318513076 0.813389758
 [22] 0.450942311 1.645542981 0.954730255 0.525579157 1.571907336 0.483463603 0.157788916
 [29] 0.612732205 1.235961598 0.185007470 0.081411559 0.480990421 1.631763372 0.238403675
 [36] 2.717314260 0.619695905 1.263575106 1.672642905 0.873156809 0.524550121 1.574289817
 [43] 0.245970711 2.508236358 1.208473313 0.569788716 0.465024978 0.120935480 0.472701076
 [50] 0.314991599 0.368006499 1.715596460 0.583509804 0.774105961 0.251164758 0.532235870

> e4<- rexp(N4)
> e4
   [1] 0.500661254 0.015546603 0.333604166 0.457382506 1.439508667 2.863401211 0.279968472
   [8] 0.858306949 1.491729341 2.478433328 0.023233615 0.164352422 1.988288742 1.646838976
  [15] 2.412625490 1.111496616 2.303506961 0.294630658 0.405892001 0.463344426 0.843325067
  [22] 3.893046250 0.405372108 1.485172427 1.623768515 0.200403341 1.631284526 0.546404892
  [29] 0.769553979 0.563014726 0.308512779 0.697679874 1.117566361 0.191337076 1.062273237
  [36] 0.554145066 0.768302864 0.919516510 1.614172196 0.443697511 0.775389963 1.003292976
  [43] 2.432220051 0.067001375 1.926266665 1.474508628 1.222979222 0.208344123 1.333074581
  [50] 0.180415994 0.047395749 1.715713242 0.696482963 0.118801949 1.478886930 0.087775380
```

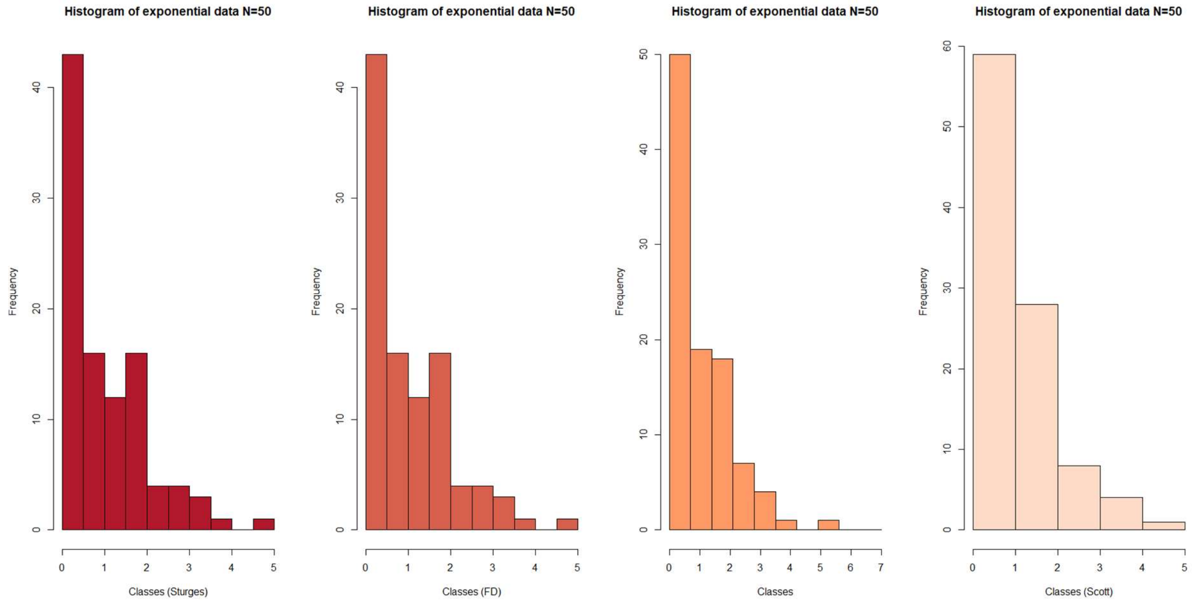
2. Representation des histogrammes en faisant varier l’amplitude des classes

POUR N=50 :



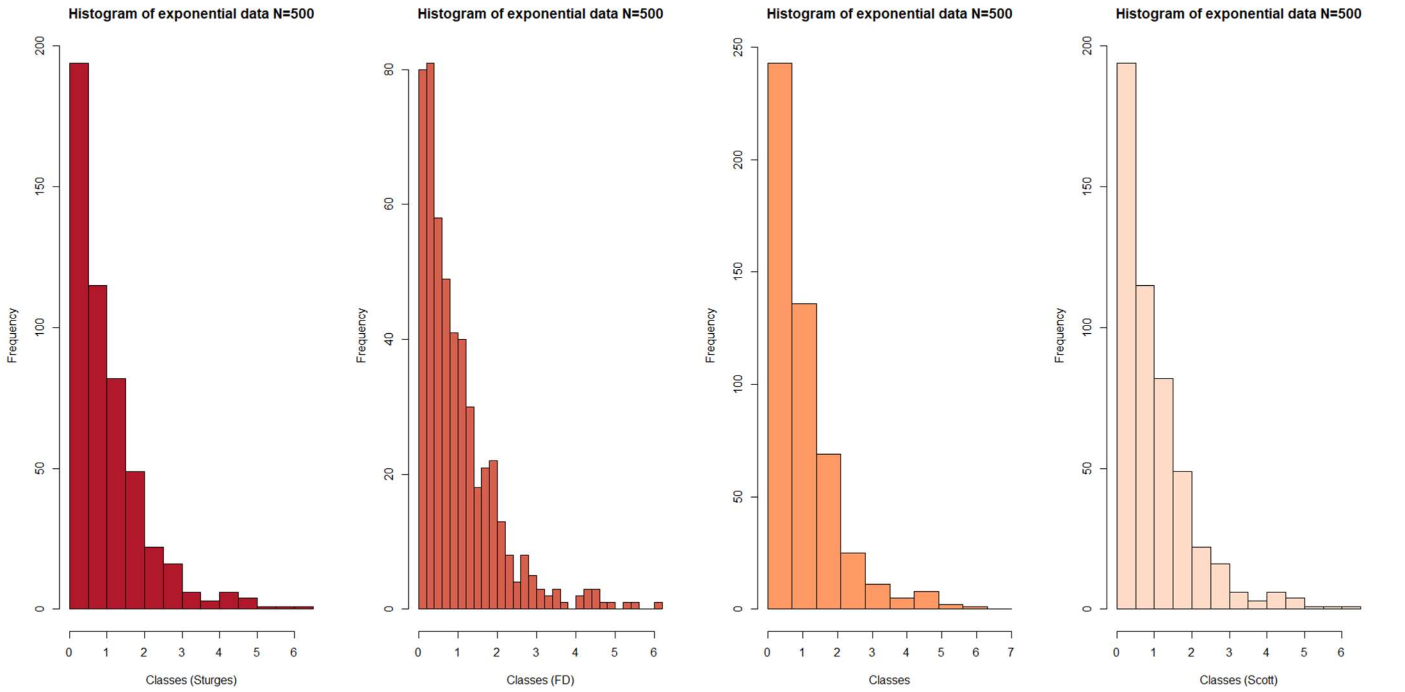
- Les histogrammes ci-dessus sont de jeux de données exponentielles de taille d'échantillon 50.
- La variation de l'amplitude de classes est faite avec quatre méthodes telles que : STURGES, FD, SCOTT et pour l'autre méthode j'ai spécifié la plage des valeurs (DE 0 à 5 avec 11 valeurs espacées de manière égale entre elles)
- Après avoir testé ces methodes il en résulte qu'on ne peut pas choisir une méthode appropriée pour varier l'amplitude des classes (car elles sont toutes identiques et l'adéquation entre la forme de l'histogramme et la forme de la distribution exponentielle et guère remarquable)

POUR N=100 :



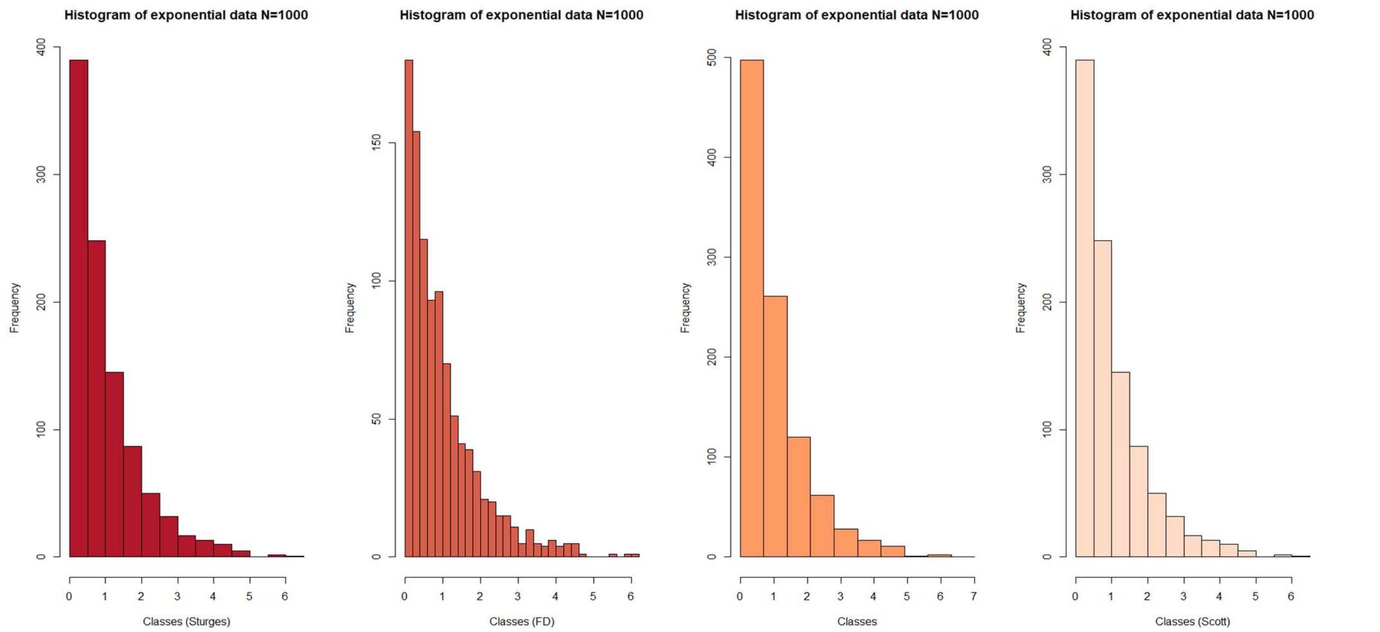
- Les histogrammes ci-dessus sont de jeux de données exponentielles de taille d'échantillon 100.
- La variation de l'amplitude de classes est faite avec quatre méthodes telles que : STURGES, FD, SCOTT et pour l'autre méthode j'ai spécifié la plage des valeurs (DE 0 à 7 avec 11 valeurs espacées de manière égale entre elles)
- Après avoir testé ces méthodes il en résulte que la méthode de Scott et plus appropriée par rapport aux autres méthodes car sa forme est la plus identique à la forme de la distribution exponentielle.

POUR N=500 :



- Les histogrammes ci-dessus sont de jeux de données exponentielles de taille d'échantillon 500.
- La variation de l'amplitude de classes est faite avec quatre méthodes telles que : STURGES, FD, SCOTT et pour l'autre méthode j'ai spécifié la plage des valeurs (De 0 à 7 avec 11 valeurs espacées de manière égale entre elles) , après avoir testé ces méthodes il en résulte que les histogrammes de Sturges, Scott est celui dans lequel on a spécifié les classes sont identiques, seul l'histogramme de FD qui est différent (il parait qu'il est divisé en plusieurs classes par rapport aux autres méthodes).

POUR N=1000 :

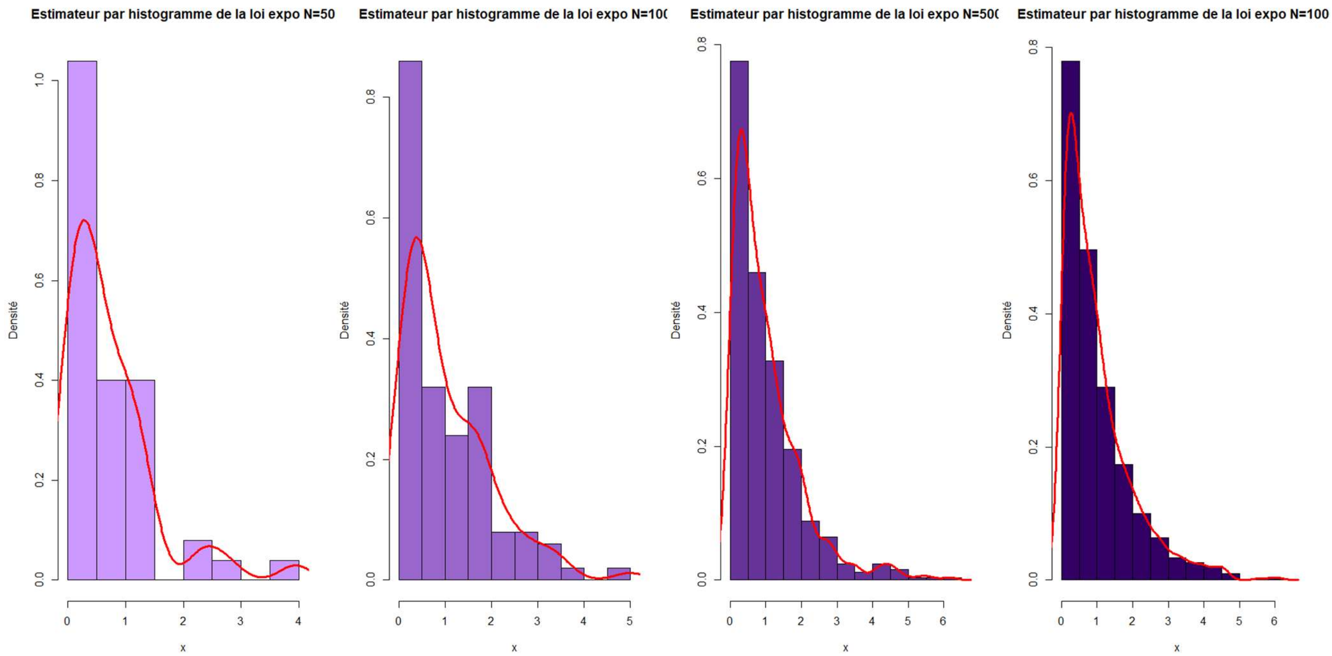


- Les histogrammes ci-dessus sont de jeux de données exponentielles de taille d'échantillon 1000.
- La variation de l'amplitude de classes est faite avec quatre méthodes telles que : STURGES, FD, SCOTT et pour l'autre méthode j'ai spécifié la plage des valeurs (De 0 à 7 avec 11 valeurs espacées de manière égale entre elles)
- Après avoir testé ces méthodes il en résulte que la méthode de Scott et Sturges sont plus appropriée par rapport aux autres méthodes car leurs forme est plus identique à la forme de la distribution exponentielle.

- On remarque qu'avec l'augmentation de la taille de l'échantillon les histogrammes sont plus représentatifs de la loi exponentielle.

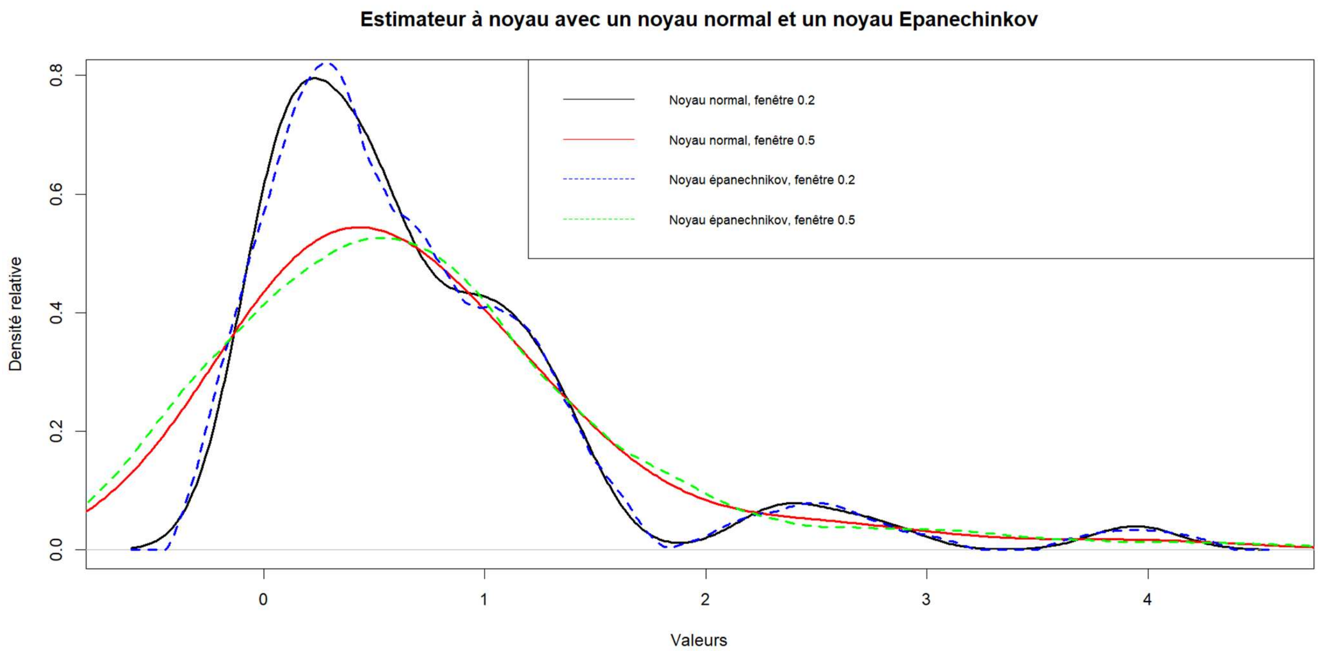
LA METHODE "STURGES" OU "SCOTT" PEUVENT ETRE CONSIDEREES COMME DES METHODES APPROPRIEES POUR DETERMINER L'AMPLITUDE DES CLASSES DU HISTOGRAMME DANS LE CAS DE LA DISTRIBUTION EXPONENTIELLE.

3. Les estimateurs par histogramme de la densité

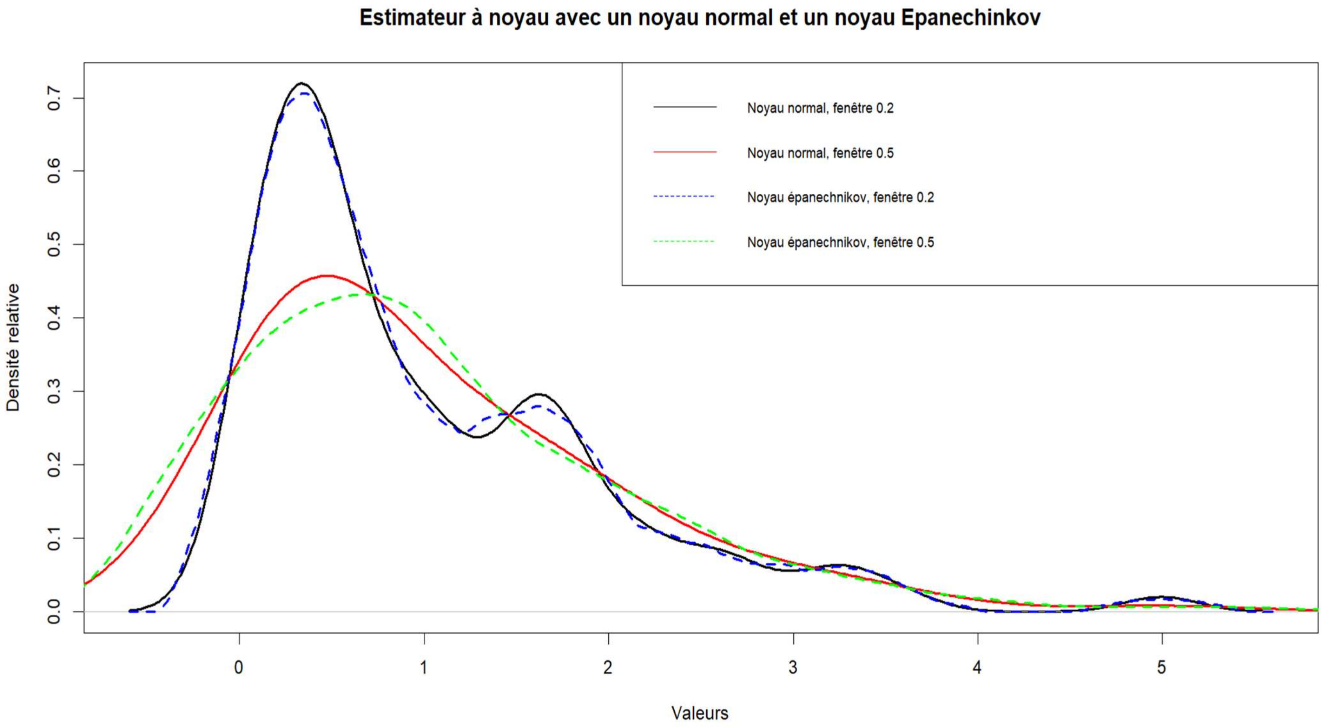


4. Les estimateurs à noyau de la densité, avec la variation des fenêtres (2 noyaux différents)

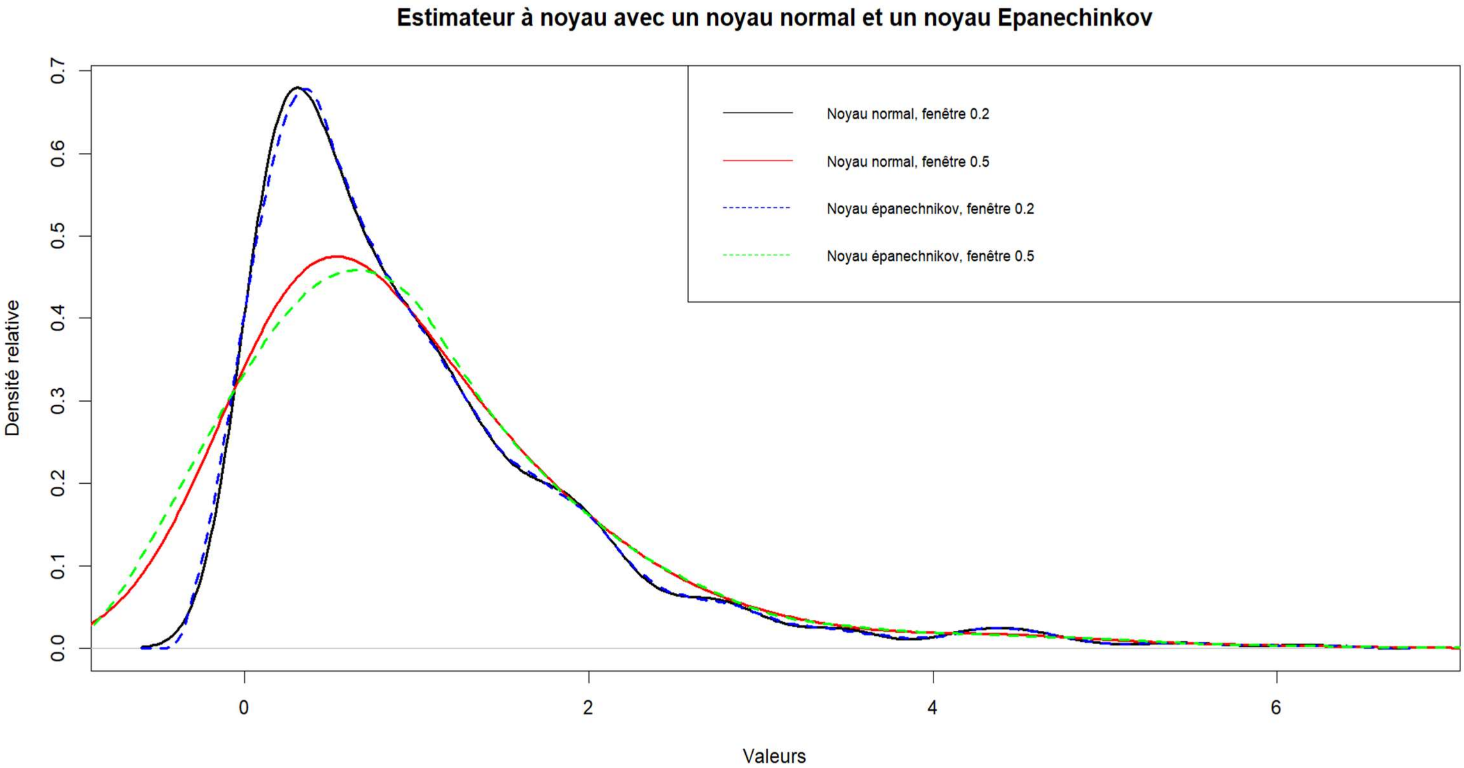
POUR N=50 :



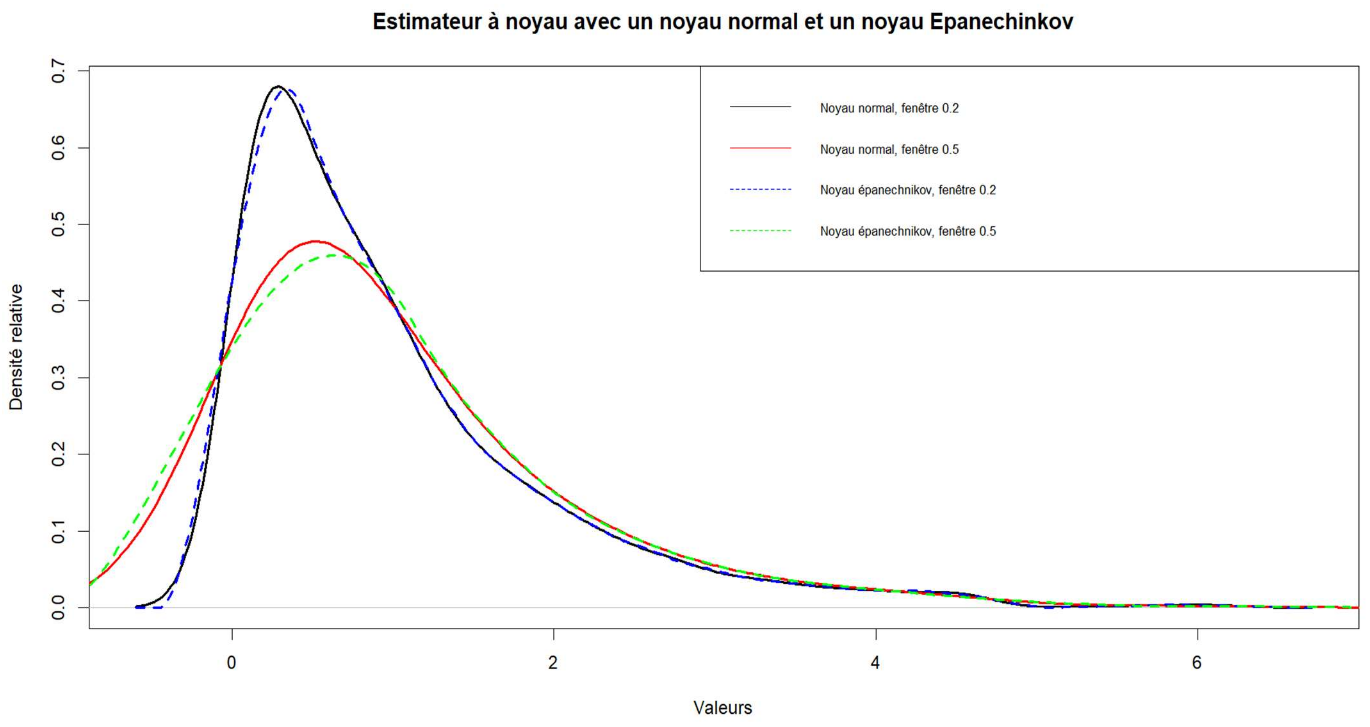
POUR N=100 :



POUR N=500 :



POUR N=1000 :



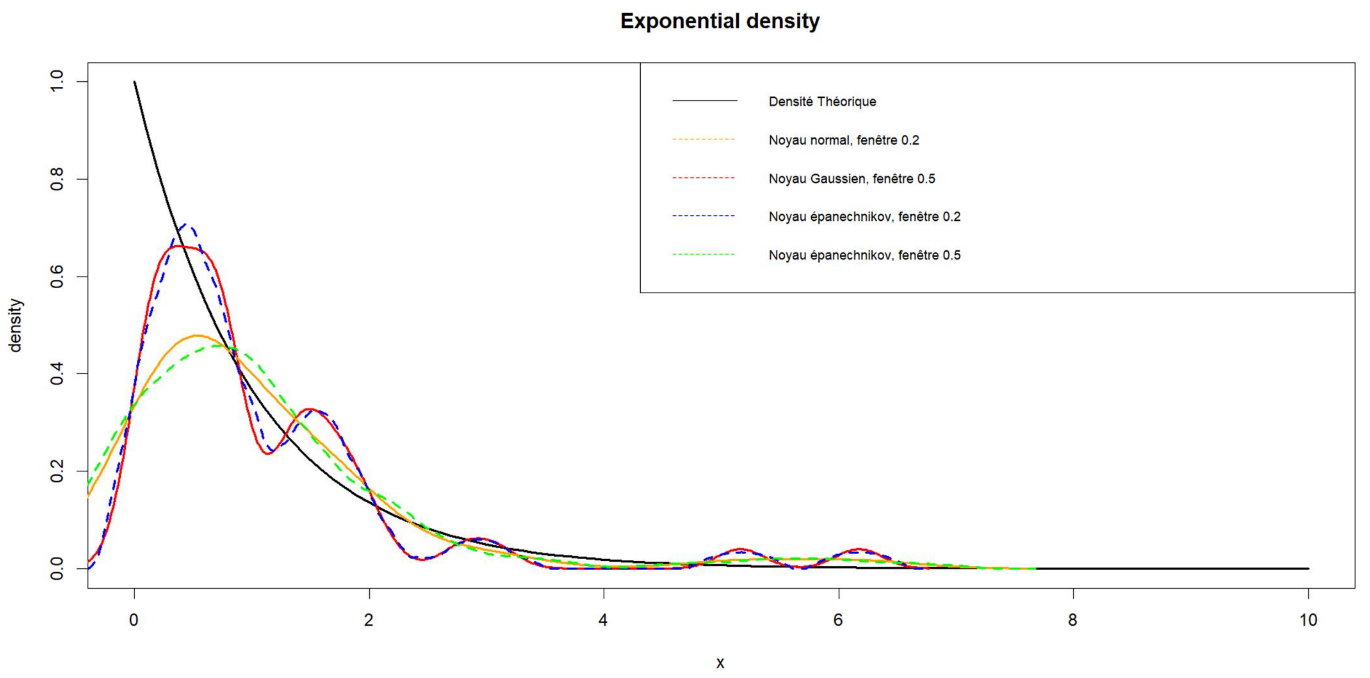
-LE NOYAU GAUSSIEN EST GENERALEMENT UTILISE CAR IL EST ASYMETRIQUE ET POSSEDE DES PROPRIETES STATISTIQUES SOUHAITABLES TELLES QUE LA CONVERGENCE UNIFORME VERS LA FONCTION DE DENSITE.

-CEPENDANT, D'AUTRES NOYAUX, TELS QUE LE NOYAU EPANECHNIKOV, POUR SA CAPACITE A GERER LES DONNEES DE BRUIT.

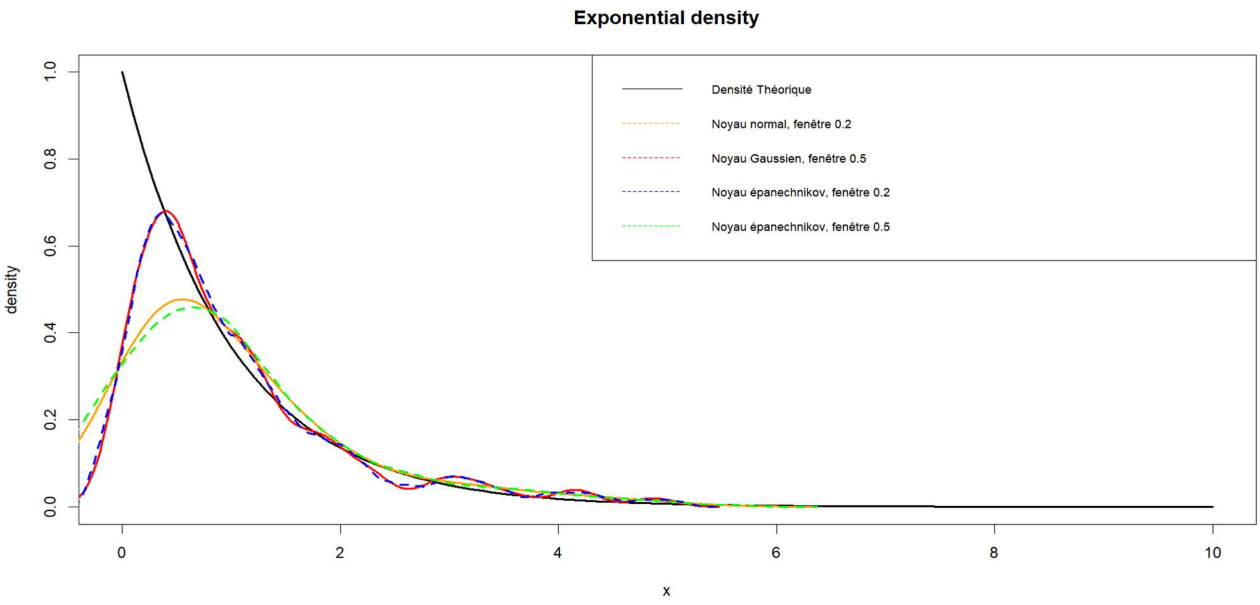
5. Les graphes des estimateurs et des densités théoriques sur le même graphe

L'intervalle choisi est [0,10]

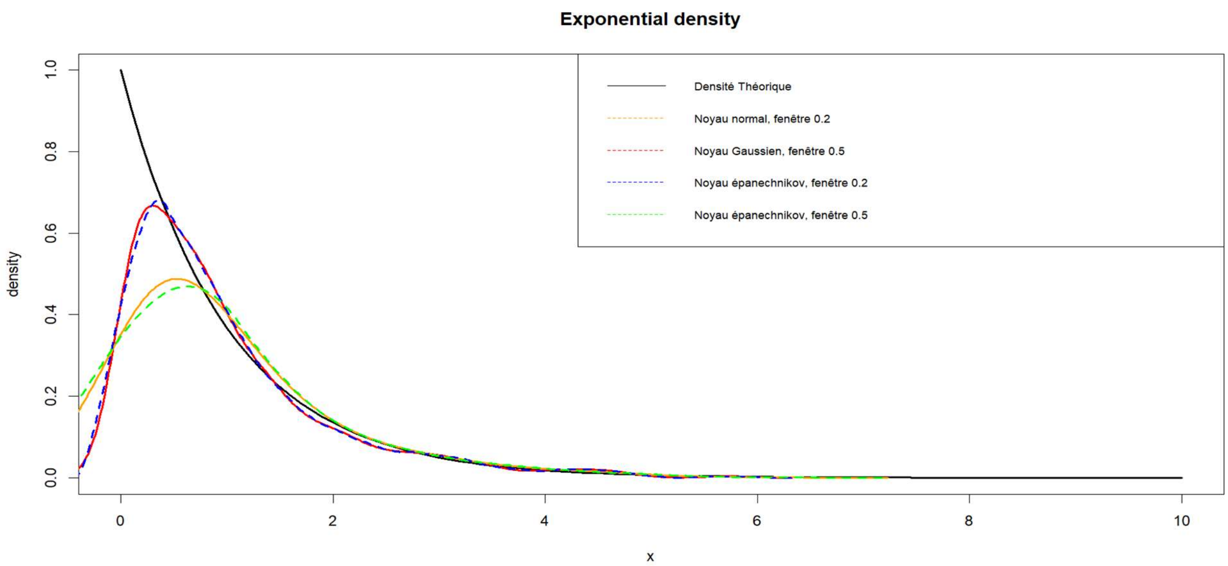
POUR N=50 :



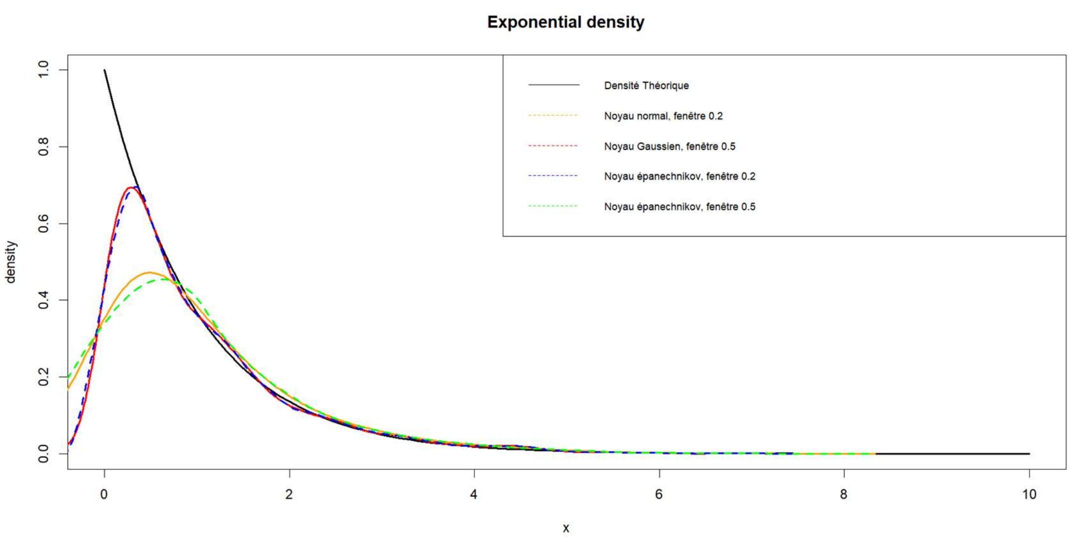
POUR N=100 :



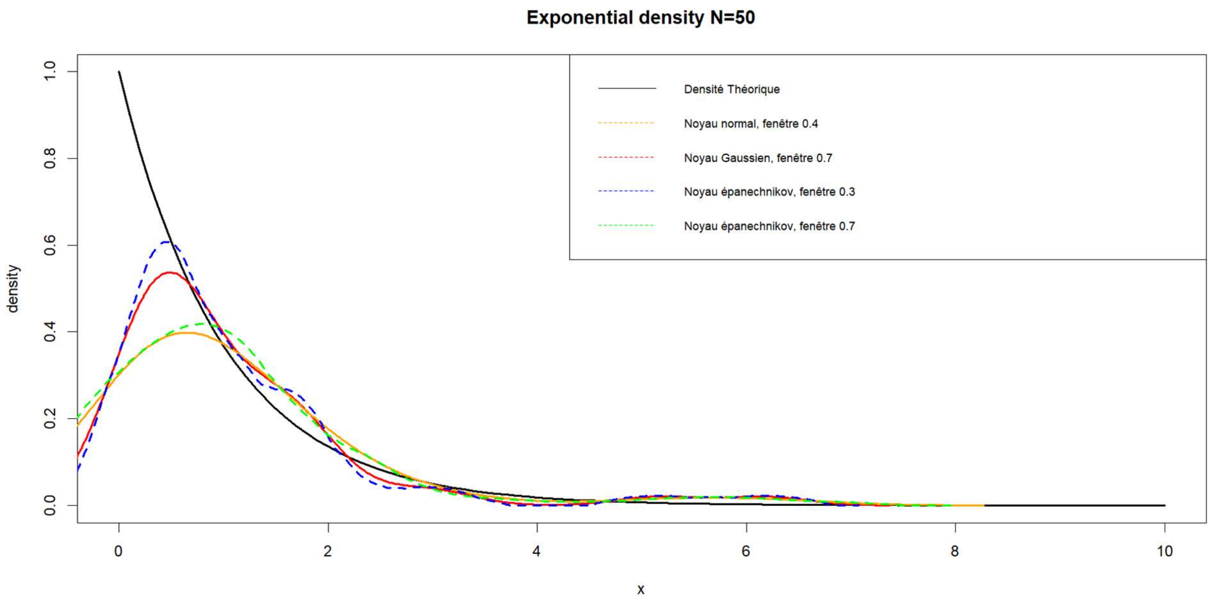
POUR N=500 :

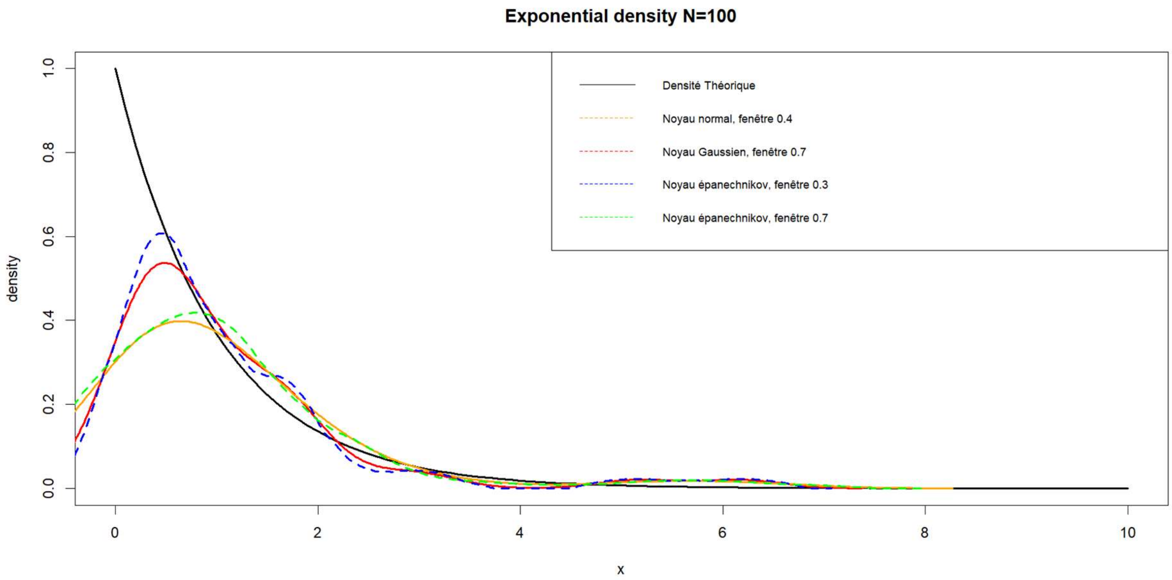


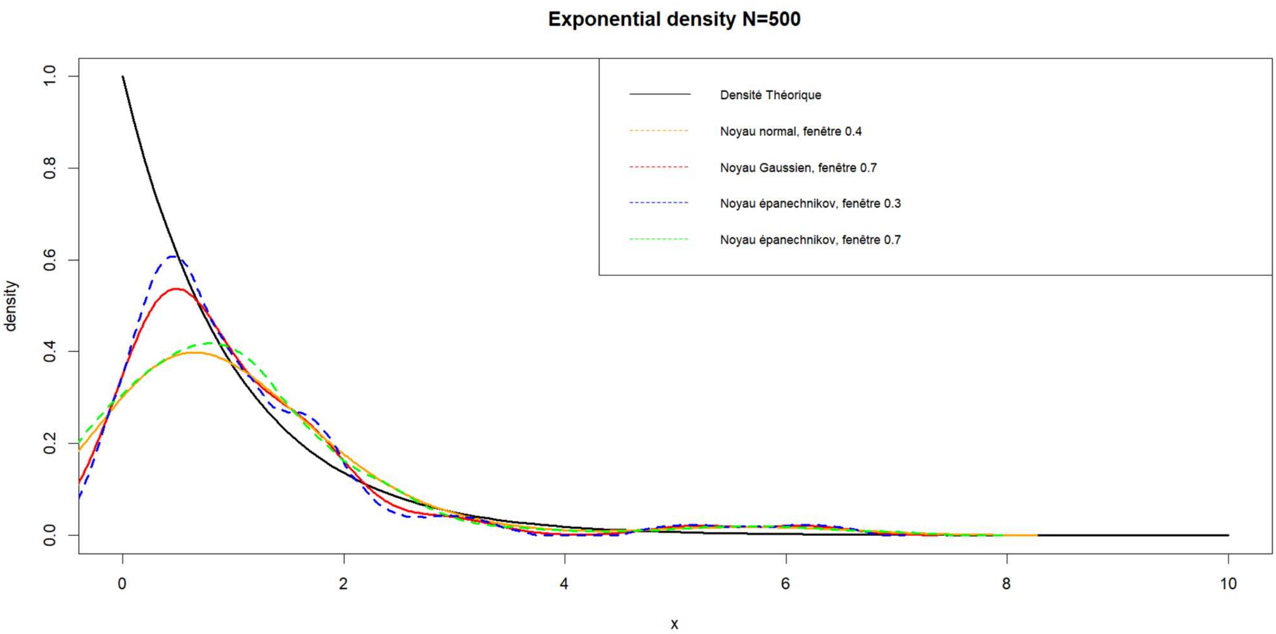
POUR N=1000 :

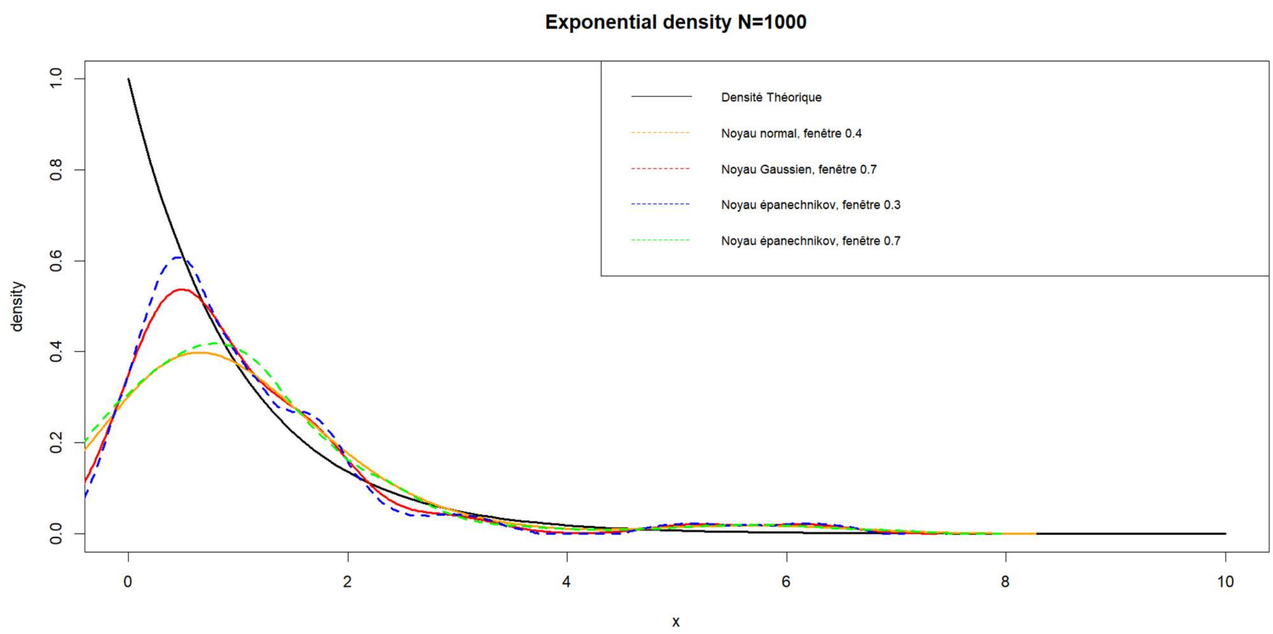


Avec une autre variation des fenêtres









INTERPRETATION :

Lorsque nous augmentons ou diminuons la fenêtre du noyau d'une distribution exponentielle, cela a pour conséquence une distribution plus lisse lorsque la fenêtre est augmentée, et une distribution plus variable lorsque la fenêtre est réduite. Dans les deux cas, la forme globale de la distribution exponentielle reste inchangée.

LOI GAUSSIENNE

1. Générer par simulation un échantillon de taille N=50, 100, 500,1000 de la loi gaussienne

SYNTAXE :

```
#Loi Normale
g1<- rnorm(N1)
g2<- rnorm(N2)
g3<- rnorm(N3)
g4<- rnorm(N4)

```

RESULTATS :

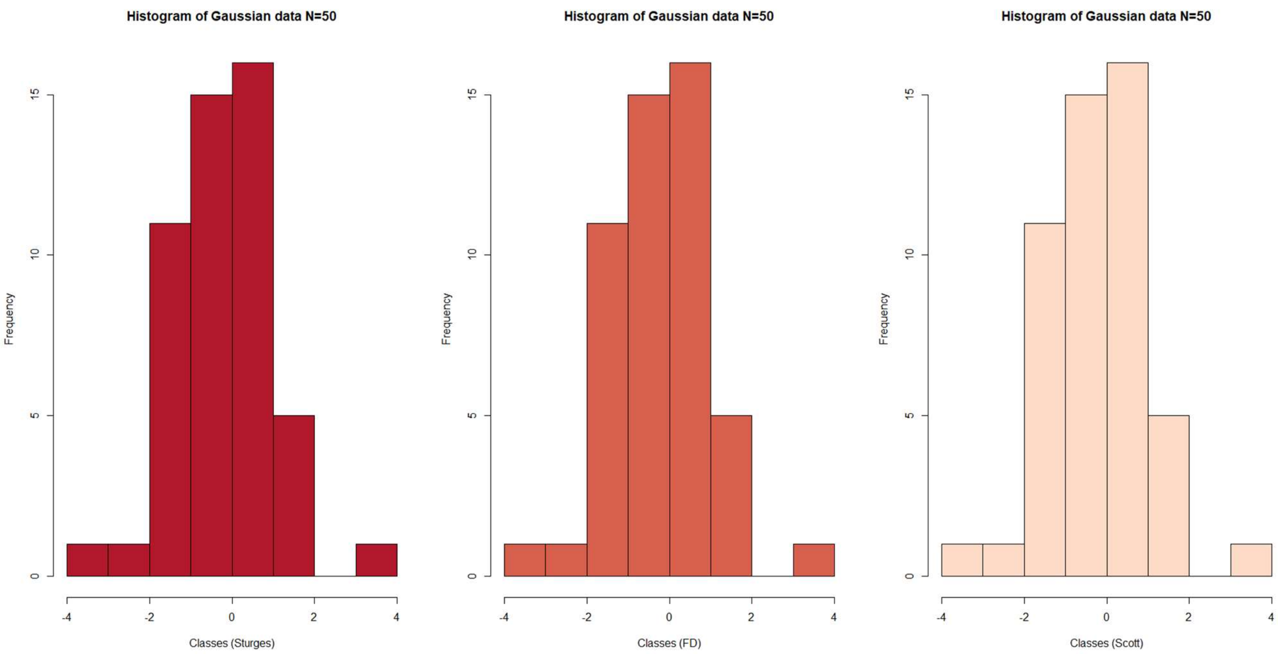
```
> g1
 [1] -0.29246368 -0.71216559  1.13146948 -0.23462004 -0.71108494  0.92932887 -1.71177675
 [8] -1.81023651 -0.19011909 -3.91864861  0.44117115  0.76283803  0.63411886 -1.38310531
[15]  0.16668708 -0.53258692 -0.68020760  0.66693347 -1.07815416  3.38497221  0.52092704
[22]  0.23916640 -0.50609236  0.99672876 -0.04073686  0.64377778 -0.69153956 -0.65518170
[29]  0.43702384 -1.91041728 -1.14653007 -1.15079309 -1.03901718  0.01421675  1.30628932
[36] -0.36822830 -2.22709556 -0.12307736 -0.83212514  0.64035953  1.78725922  0.86910028
[43] -1.00743144 -1.74940041 -0.45045704  1.03819590 -1.39093441  0.66139145  0.09801795
[50]  1.23542298
> g2
  [1]  0.56008301 -1.00699123  1.25727445  0.56199503 -1.15186314  0.10155234 -1.00090165
  [8]  1.00732024 -1.90778745 -1.75574339 -1.13933176  2.66319974 -0.11344089  0.20720943
 [15]  1.51009550 -1.33318365 -1.78821988  0.63963155  0.68054271 -0.02923428  0.35878460
 [22] -0.27708595  0.79482364  1.13974891 -1.04493612  1.33611020 -0.88401109  1.50284771
 [29] -0.66220400 -0.88634889  0.37765923  0.83202452  1.07540453  0.91226736  0.97723569
 [36]  1.55991219  0.21265840  0.11029515 -0.98402974 -0.66525063 -0.06462275  0.56925839
 [43]  0.71273197  0.47259159  0.66197088 -0.08847392 -0.76706709  0.92011742  0.45715501
 [50] -0.92895930 -0.83755535 -1.95984759 -0.94924275 -0.21735704 -1.17374020 -0.40265442
 [57]  0.85066994  1.02000712  0.01041580  0.35481123  1.37408560 -0.37717085  1.46695610
```

```
> g3
 [1]  0.457045101  0.116776326  0.426383387 -0.235241830  0.730549390  1.424280925
 [7]  0.850491731  0.678369831  0.612362726 -0.112482004  1.794431546  0.616922503
[13]  0.554796496  0.329884626 -0.005521245  0.384895357 -0.870005587 -0.602675021
[19] -0.793839347  1.133315612  0.025425812  1.689536288  0.392618510 -0.265153143
[25]  1.107380281 -1.222188464 -0.570192393 -0.790151001  0.372652375  1.566903800
[31] -1.326778344  0.989941820  1.687120825 -0.737589613 -1.747633782 -0.544868660
[37]  0.364733975 -0.200949942 -0.598031918  0.652284312 -0.029821464  1.281801685
[43] -0.244575655  0.585771715 -0.400676237  2.121373272 -0.386602429  0.873229306
[49]  1.691009896 -1.182679702  0.470891264 -0.628981171 -2.835390914 -1.663349057
[55]  0.346247492 -0.365487427 -1.155926519 -0.899487301  0.241050192 -0.501858933
```

```
> g4
  [1] -0.3799665008  0.7598092460  0.3827413296  2.5187819872 -1.7089790542
  [6] -0.6687323534  1.1371361631  0.3674128880  1.0566553458  0.3939267057
 [11]  0.6993374158  0.4864317890 -0.7356993644 -1.9387139702 -0.9773327581
 [16] -0.1170491607 -0.0541716747  1.4117127745  1.1716035989 -1.0594147297
 [21] -1.8685640250 -1.9194157379  1.0870388312 -0.3358729010  0.0414561071
 [26] -0.2784759014 -1.6812494469 -1.0021653712 -1.4516161369 -0.2543441500
 [31]  1.4719641387  1.3819883702 -0.0869334966 -0.9769733695  1.5744496746
 [36] -1.5470532175 -0.9793971076  0.6329799329  0.2553828653 -0.4595986849
 [41]  0.9184512763 -0.8221698289  1.1986479113  2.3655123503  1.2921579077
 [46]  0.6141467153  1.6339338574 -0.4777778735 -1.3906754399  1.4792149222
 [51]  1.8152144413  0.4268317411 -0.6724750988  0.0647017321  1.2529081024
 [56] -0.4330553048  1.0551107549 -0.4888095907 -0.9448505226  0.0535974619
```

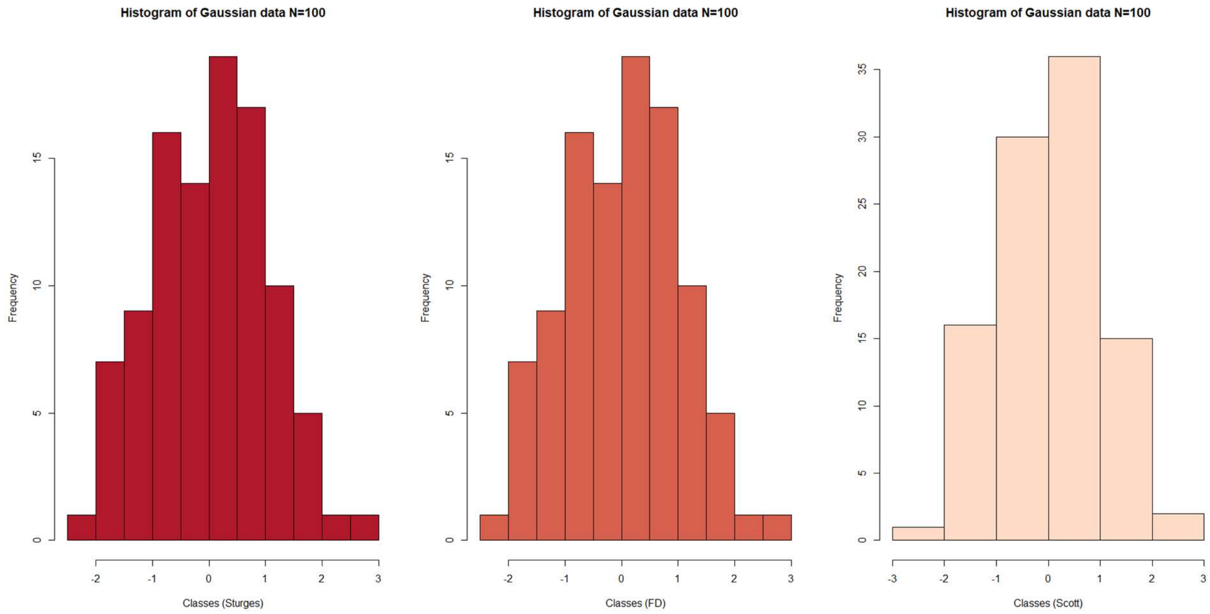
2. Représentation des histogrammes en faisant varier l'amplitude des classes

POUR N=50 :



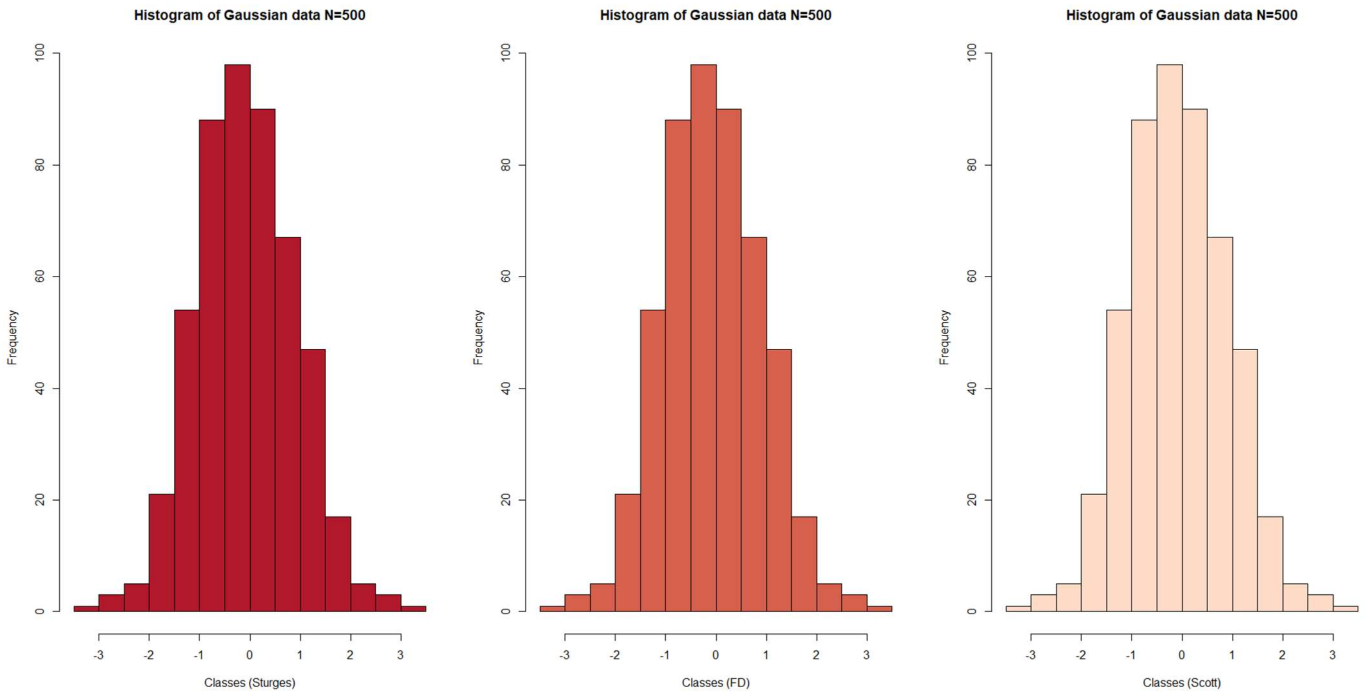
- Les histogrammes ci-dessus sont de jeux de données gaussiennes de taille d'échantillon 50.
- La variation de l'amplitude de classes est faite avec trois méthodes telles que : STURGES, FD et SCOTT
- Après avoir testé ces méthodes il en résulte qu'on ne peut pas choisir une méthode appropriée pour varier l'amplitude des classes (car elles sont toutes identiques et l'adéquation entre la forme de l'histogramme et la forme de la distribution exponentielle et guère remarquable)

POUR N=100 :



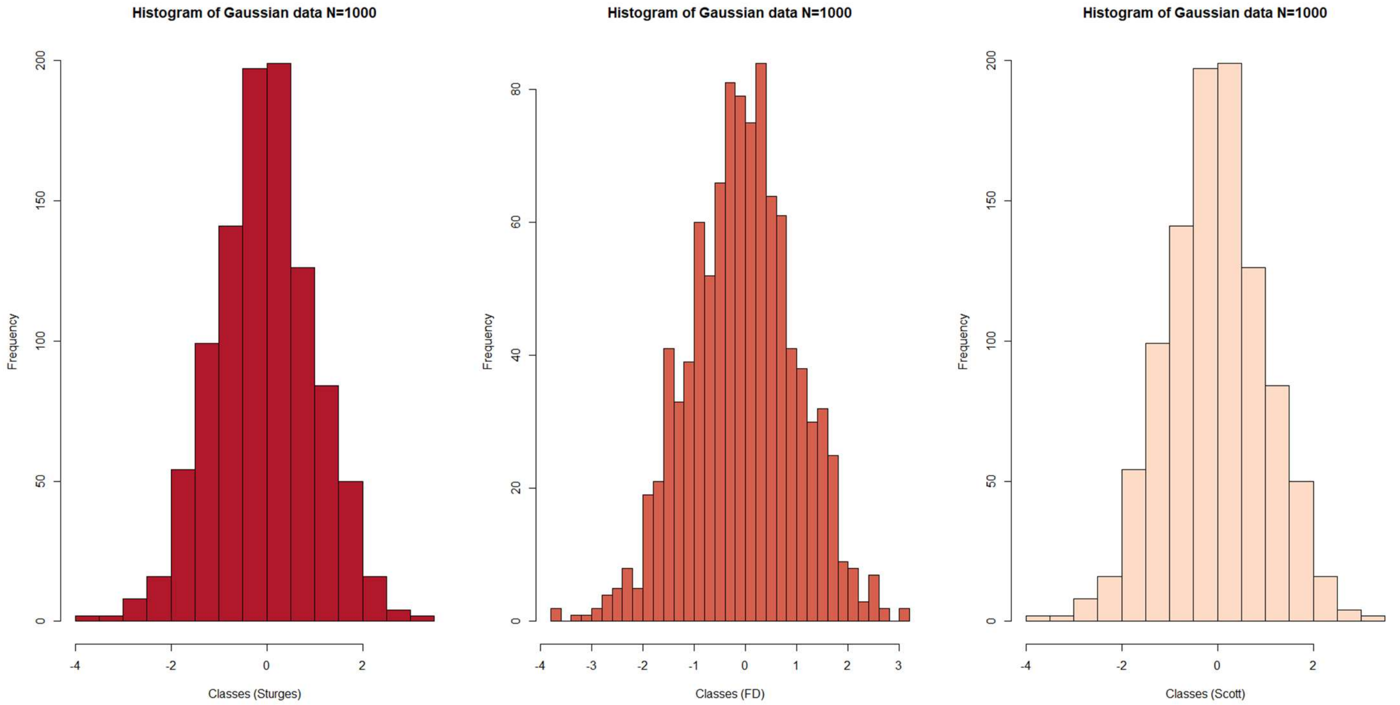
- Les histogrammes ci-dessus sont de jeux de données gaussiennes de taille d’échantillon 100.
- La variation de l’amplitude de classes est faite avec trois méthodes telles que : STURGES, FD et SCOTT
- Après avoir testé ces méthodes il en résulte que la méthode de Scott et plus appropriée par rapport aux autres méthodes car sa forme est la plus identique à la forme de la distribution exponentielle.

POUR N=500 :



- Les histogrammes ci-dessus sont de jeux de données gaussiennes de taille d’échantillon 500.
- La variation de l’amplitude de classes est faite avec trois méthodes telles que : STURGES, FD et SCOTT
- Après avoir testé ces méthodes il en résulte que les histogrammes sont tous identiques et on constate que leur forme ressemble à celle de la loi normale.

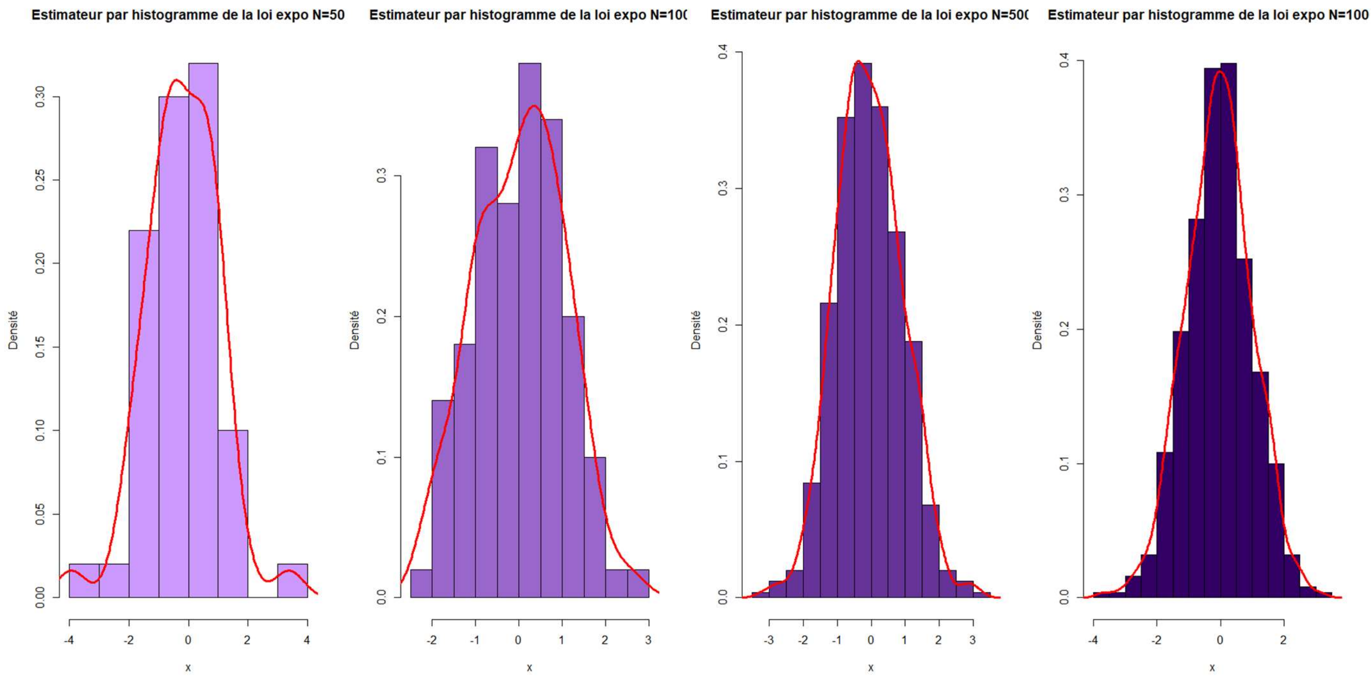
POUR N=1000 :



- Les histogrammes ci-dessus sont de jeux de données exponentielles de taille d’échantillon 1000.
- La variation de l’amplitude de classes est faite avec trois méthodes telles que : STURGES, FD et SCOTT
- Après avoir testé ces méthodes il en résulte que la méthode de Scott et Sturges sont plus appropriée par rapport à la méthode FD car leurs forme est plus identique à la forme de la distribution exponentielle (car elles ont le moins nombre de classes).
- On remarque qu’avec l’augmentation de la taille de l’échantillon les histogrammes sont plus représentatifs de la loi gaussienne.

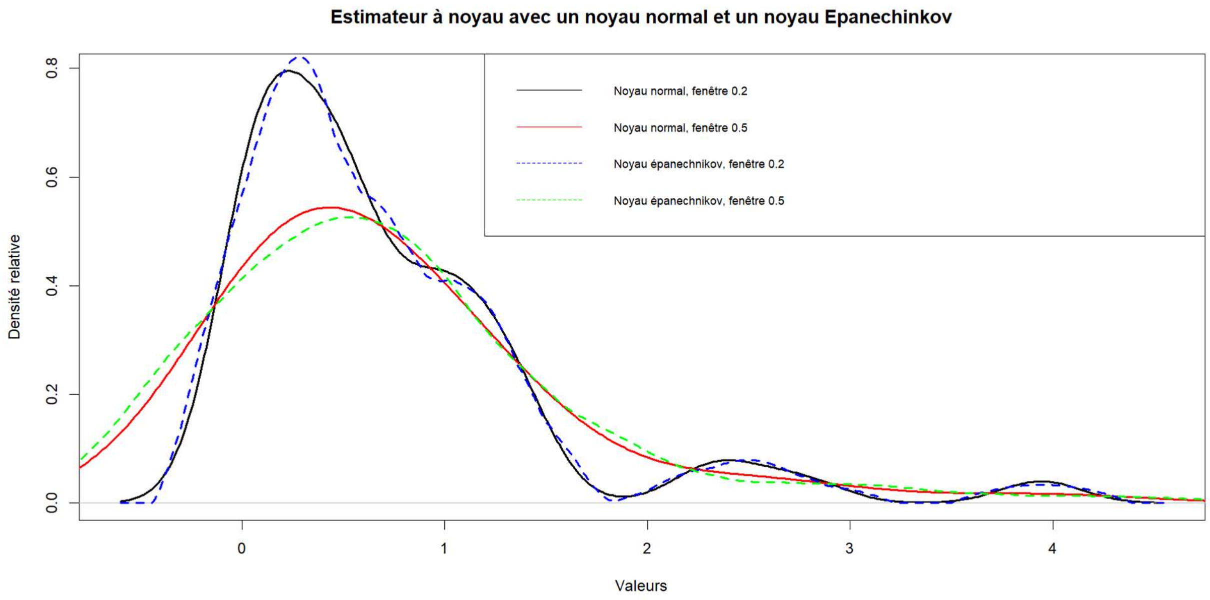
LA METHODE "STURGES" OU "SCOTT" PEUVENT ETRE CONSIDEREES COMME DES METHODES APPROPRIEES POUR DETERMINER L’AMPLITUDE DES CLASSES DU HISTOGRAMME DANS LE CAS DE LA DISTRIBUTION GAUSSIENNE.

1. Les estimateurs par histogramme de la densité

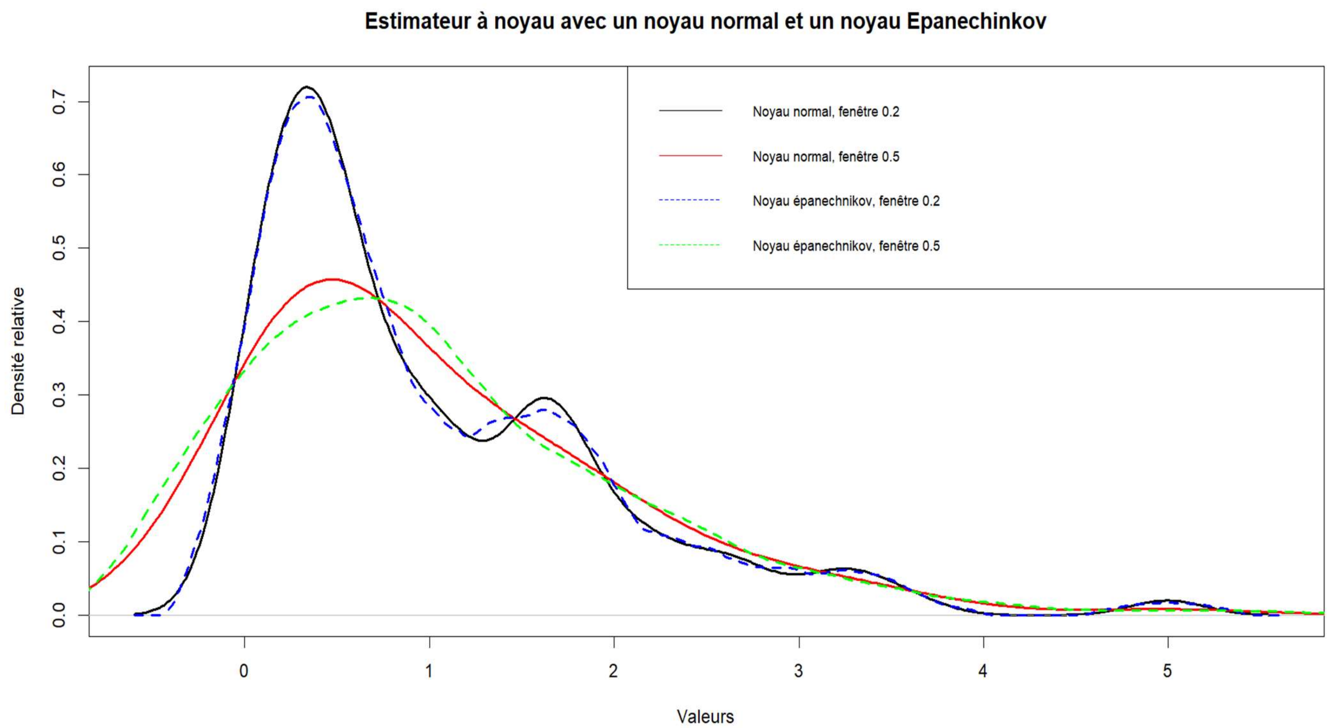


2. Les estimateurs à noyau de la densité, avec la variation des fenêtres (2 noyaux différents)

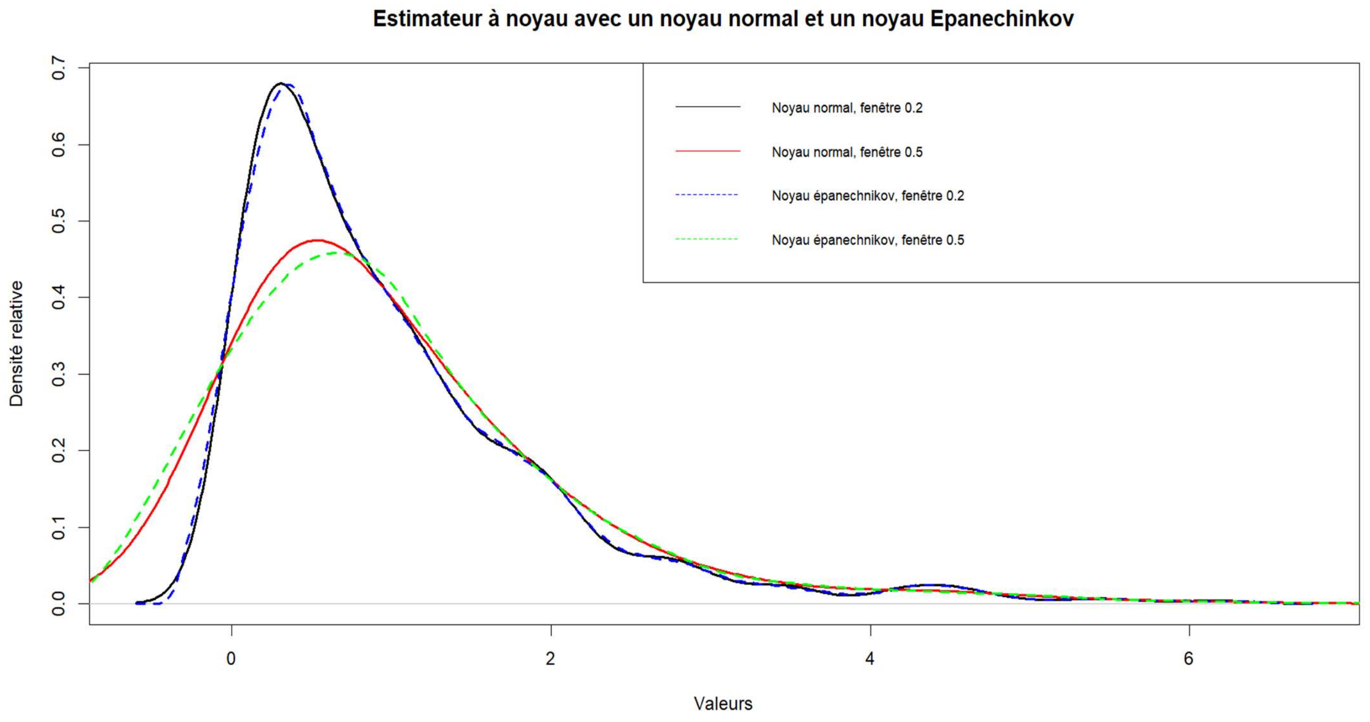
POUR N=50 :



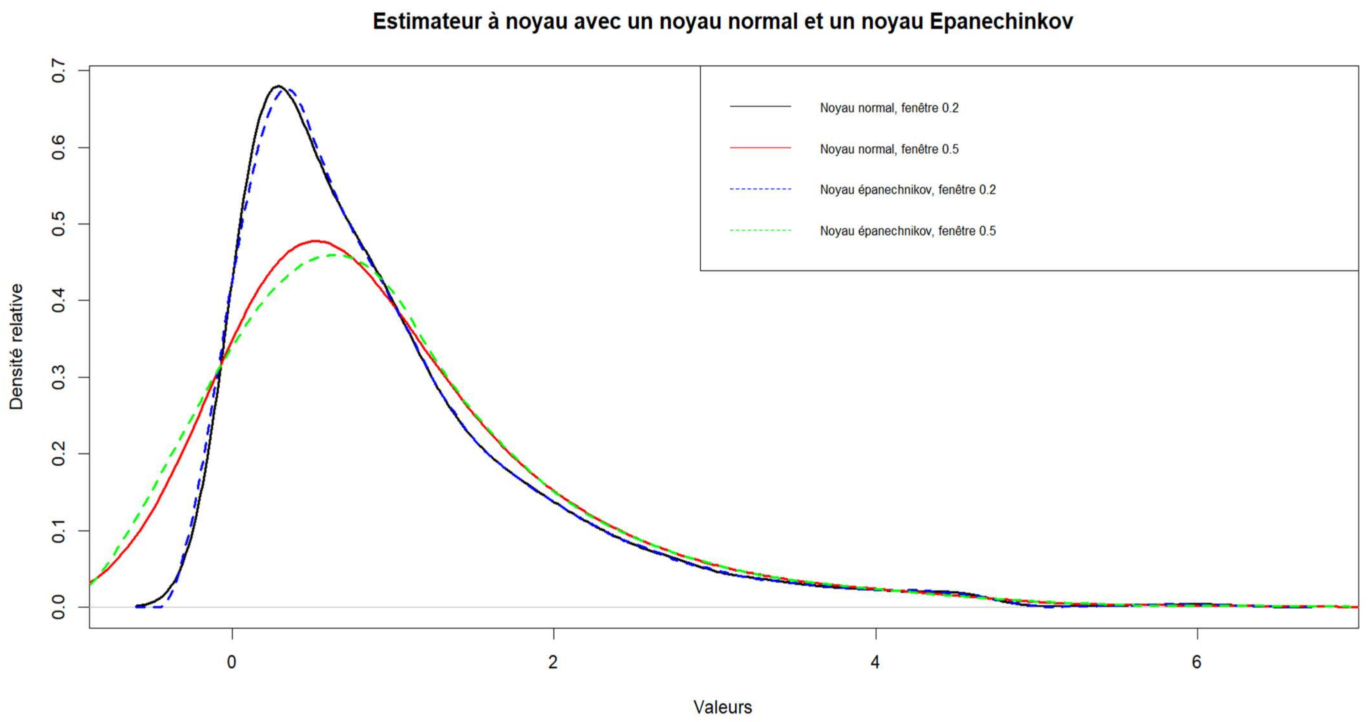
POUR N=100 :



POUR N=500 :


Estimateur à noyau avec un noyau normal et un noyau Epanechinkov
Noyau normal, fenêtre 0.2
Noyau normal, fenêtre 0.5
Noyau épanechnikov, fenêtre 0.2
Noyau épanechnikov, fenêtre 0.5
Densité relative
Valeurs

POUR N=1000 :


Estimateur à noyau avec un noyau normal et un noyau Epanechinkov
Noyau normal, fenêtre 0.2
Noyau normal, fenêtre 0.5
Noyau épanechnikov, fenêtre 0.2
Noyau épanechnikov, fenêtre 0.5
Densité relative
Valeurs

-LE NOYAU GAUSSIEN EST GENERALEMENT UTILISE CAR IL EST ASYMETRIQUE ET POSSEDE DES PROPRIETES STATISTIQUES SOUHAITABLES TELLES QUE LA CONVERGENCE UNIFORME VERS LA FONCTION DE DENSITE.

-CEPENDANT, D'AUTRES NOYAUX, TELS QUE LE NOYAU EPANECHNIKOV, POUR SA CAPACITE A GERER LES DONNEES DE BRUIT.

INTERPRERATION

Lorsque nous augmentons la fenêtre (ou la largeur) du noyau d'une distribution gaussienne, cela se traduit par une distribution plus lisse, avec moins de variation. Inversement, une réduction de la taille de la fenêtre entraîne une distribution plus aiguë et donc plus variable. Dans les deux cas, la forme générale de la distribution gaussienne restera inchangée.