Marcel Délèze
Edition 2017

Thème : Statistiques I, § 4 Distributions théoriques à une variable continue
Lien vers les énoncés des exercices:
https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_1/4-stat_I.pdf

Packages de l’auteur

- On peut consulter le mode d’emploi du package **Statistique**:
  https://www.deleze.name/marcel/sec2/applmaths/packages/aide/Statistique.pdf
- Avant d’utiliser le package, il faut le charger en donnant son adresse web:

```
Needs["Statistique`",
 "https://www.deleze.name/marcel/sec2/applmaths/packages/Statistique.m"]
```
nécessite

Voici la liste des instructions disponibles :

```
Names["Statistique`*"]
```
noms

```
{amplitudes, densiteContinue, densites, diagrammeBatons,
 diagrammeCumulatif, distributionContinue, distributionLissee, fctDensite,
 fctFrequenceCumulee, frequenceCumuleeContinue, frequenceCumuleeLissee,
 histogramme, InterpolatedQuantile, noeudsPolygonaux, polygoneDeDensite,
 quantileC, quantileLisse, sommesCumulees, StandardDeviationMLE, VarianceMLE}
```

- Le package **Tableaux** contient des commandes qui facilitent la présentation des données et résultats sous la forme de tableaux:

```
Needs["Tableaux`",
 "https://www.deleze.name/marcel/sec2/applmaths/packages/Tableaux.m"]
```
nécessite

```
Names["Tableaux`*"]
```
noms

```
{afficheTableau, afficheTableauTitre, arrondis, fusionneColonnes,
 fusionneLignes, fusionneTableaux, prodCart, prodCartTrans, tableauGraph}
```

- On peut consulter le mode d’emploi du package **Tableaux**:
  https://www.deleze.name/marcel/sec2/applmaths/packages/aide/Tableaux.pdf

Pour ne pas oublier d'exécuter ces instructions au début de chaque session de travail, il est conseillé de déclarer les instructions **Needs** comme étant des cellules d'initialisation. Pour ce faire, sélectionnez les cellules voulues puis passez par le menu

Cell / Cell properties / Initialization cell

Corrigé de l'exercice 4.1 - 1

```
distr = UniformDistribution[{0, 1}];
```
distribution uniforme

Paramètres empiriques

```
n = 1000;
x = RandomReal[distr, n];
    nombre réel aléatoire
```

```
m = Mean[x]
    valeur moy
```

```
0.50002
```

```
v = VarianceMLE[x]
```

```
0.0822342
```

Comparons avec les valeurs théoriques

```
μ = N[Mean[distr]]
    ·· valeur moyenne
```

```
0.5
```

```
V = N[Variance[distr]]
    ·· variance
```

```
0.0833333
```

Calcul théorique (facultatif)

Partageons l'intervalle en *n* classes de même amplitude $\frac{1}{n}$. Les centres des classes sont

```
Clear[c, j, n, v]
efface
```

$$c[j\_] = \left(j - \frac{1}{2}\right)\frac{1}{n};$$

La variance calculée au moyen de ces classes dépend du nombre de classes :

$$v[n\_] = \sum_{j=1}^{n}\left(c[j] - \frac{1}{2}\right)^2 \frac{1}{n}$$

$$\frac{-1 + n^2}{12\, n^2}$$

Pour un nombre de classes qui tend vers l'infini, on obtient comme limite la variance de la distribution continue :

```
Limit[v[n], n → ∞]
limite
```

$$\frac{1}{12}$$

Corrigé de l'exercice 4.2 - 1

Pour les mathématiques, la variable centrée réduite nous donne

$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{5.2 - 4.8}{0.5} = 0.8$$

Pour la physique, la variable centrée réduite nous donne

$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{5.5 - 5.1}{0.8} = 0.5$$

Signification

$Z > 0$ la note de l'étudiant est au-dessus de la moyenne : $X > \mu_X$

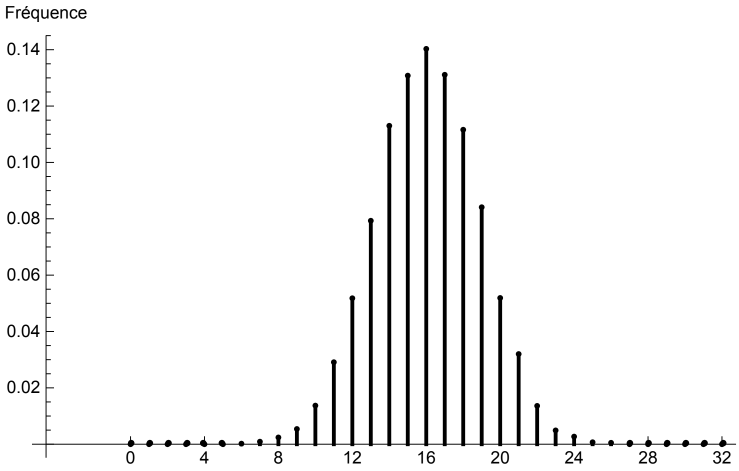
$0 < Z < 1$ la note X de l'étudiant est située dans $]\mu_X, \mu_X + \sigma_X[$

L'étudiant est relativement meilleur en mathématiques.

Corrigé de l'exercice 4.2 - 2

Diagramme à bâtons d'une simulation

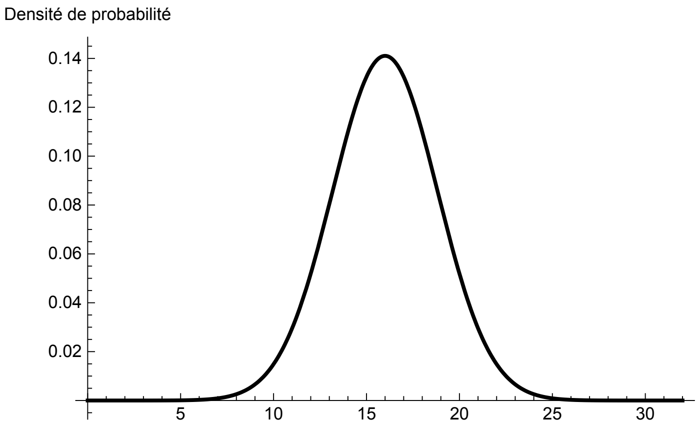
```
k = 32;
p = 1/2;
n = 10000;
distr = BinomialDistribution[k, p];
        distribution binomiale
x = RandomInteger[distr, n];
    entier aléatoire
eff = Map[Count[x, #] &, Range[0, k]];
      app· compte         plage
freq = eff/n;
g1 = diagrammeBatons[Range[0, k], freq,
                     plage
  AxesLabel → {None, "Fréquence"}, Ticks → {Range[0, k, 4], Automatic}]
  titre d'axe  aucun                 graduati···  plage          automatique
```



Graphique de la fonction "densité de probabilité" :

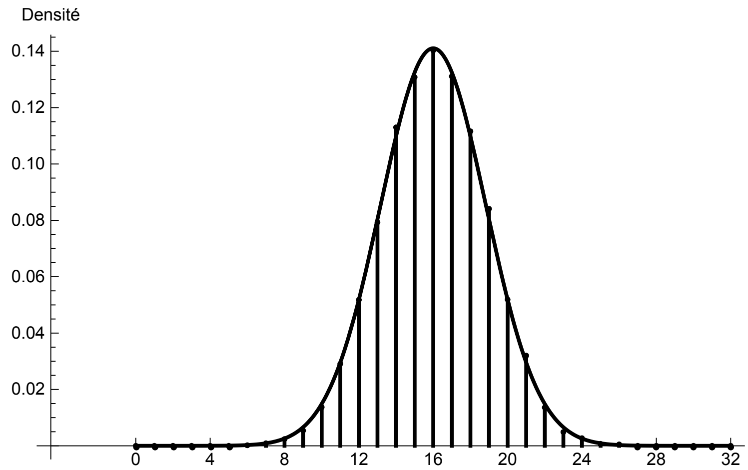
```
μ = k p; σ = √(k p (1 - p));
normal = NormalDistribution[μ, σ];
         distribution normale
```

```
Clear[t];
efface
g2 = Plot[PDF[normal, t], {t, 0, k},
     trac··· fonction de densité de probabilité
  PlotStyle → {Black, Thick}, AxesLabel → {None, "Densité de probabilité"}]
              noir    épais   titre d'axe    aucun
```



Superposons les deux graphiques:

```
Show[g1, g2, Ticks → {Range[0, k, 4], Automatic}, AxesLabel → {None, "Densité"}]
montre       graduati··· plage       automatique   titre d'axe    aucun
```

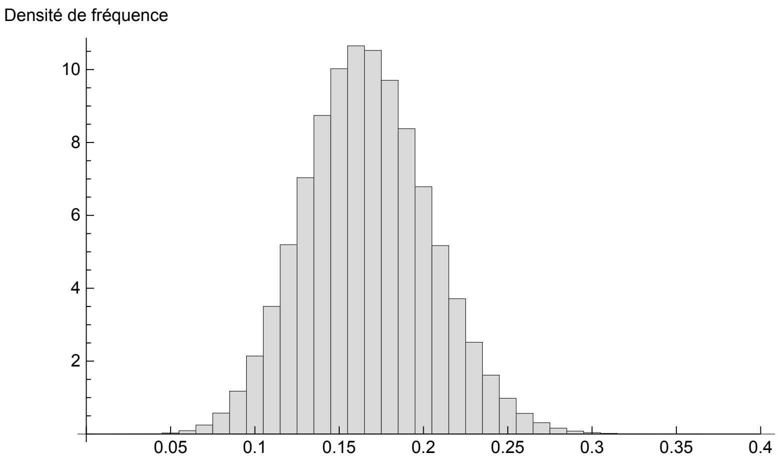


Dans le graphique précédent, l'axe vertical désigne non pas la fréquence mais la densité de fréquence et la densité de probabilité. Les écarts entre les densités empiriques et les densités théoriques tendent vers zéro (en probabilité) lorsque la taille n de l'échantillon tend vers l'infini.

Corrigé de l'exercice 4.2 - 3 a)

Comparaison d'une distribution théorique binomiale et de la distribution normale.

```
k = 100;
p = 1/6;
us = Range[-1/2, k + 1/2];                     (* us = bornes des classes de S *)
     plage
u = us/k;                                      (* u = bornes des classes de M  *)
distr = BinomialDistribution[k, p];            (*  distr = distribution de S  *)
        distribution binomiale
freq = Table[PDF[distr, t], {t, 0, k}];
       table fonction de densité de probabilité
g1 = histogramme[u, freq,
  PlotRange → {{0, 2/5}, All}, Ticks → {Range[0., k, 5.]/k, Automatic},
  zone de tracé         tout    graduations                automatique
  AxesLabel → {None, "Densité de fréquence"}, AxesOrigin → {0, 0}]
  titre d'axe  aucun                           origine des axes
```

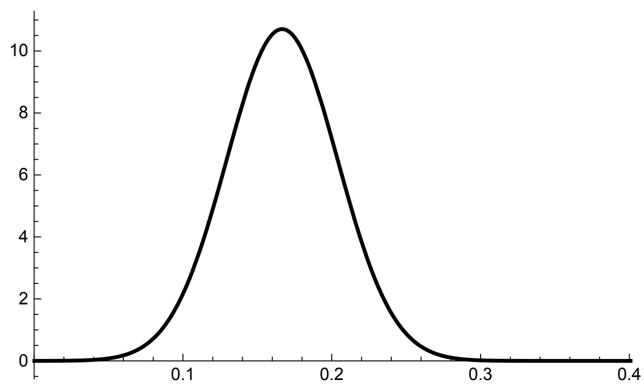


Graphique de la fonction "densité de probabilité" :

```
μs = k p; σs = √(k p (1 - p)); μm = μs/k; σm = σs/k;
normal = NormalDistribution[μm, σm];
         distribution normale
```

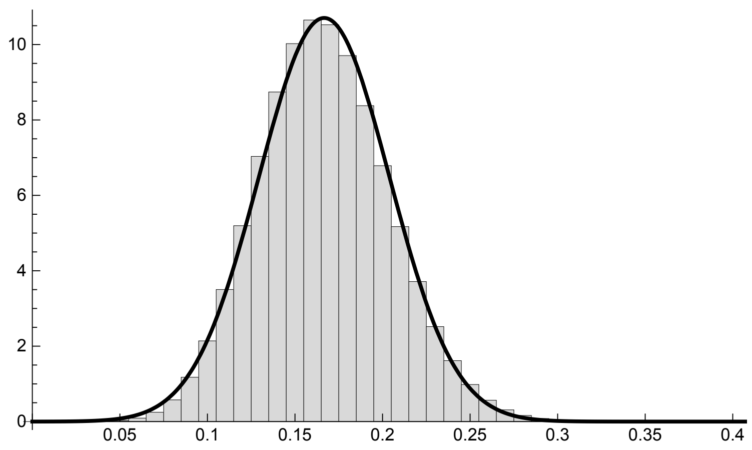
```
Clear[t];
```
efface

```
g2 = Plot[PDF[normal, t], {t, 0, 1}, PlotRange → {{0, 2/5}, All},
  PlotStyle → {Black, Thick}, AxesLabel → {None, "Densité de probabilité"}]
```
trac··· fonction de densité de probabilité zone de tracé tout style de tracé noir épais titre d'axe aucun



Superposons les deux graphiques:

```
Show[g1, g2, PlotRange → {{0, 2/5}, All},
  Ticks → {Range[0., k, 5.]/k, Automatic}, AxesLabel → {None, "Densité"}]
```
montre zone de tracé tout graduations automatique titre d'axe aucun

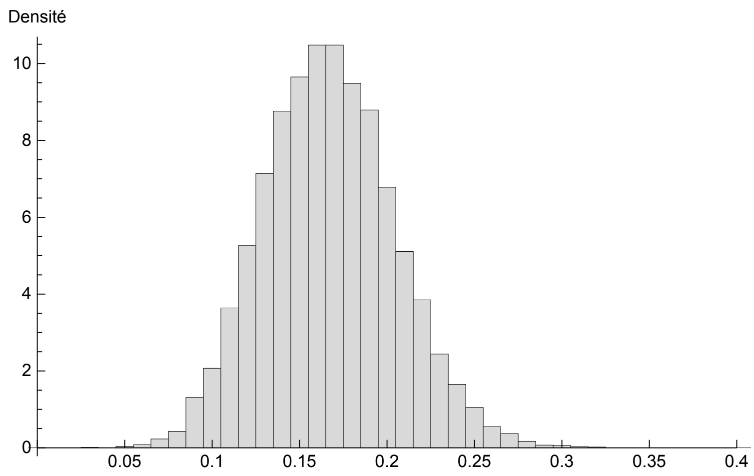


Dans le graphique précédent, l'axe vertical désigne la densité de fréquence et la densité de probabilité. Les écarts entre les densités binomiales et les densités normales tendent vers zéro lorsque le nombre de lancers du dé *k* tend vers l'infini.

Corrigé de l'exercice 4.2 - 3 b)

Comparaison d'une distribution empirique binomiale et de la distribution normale

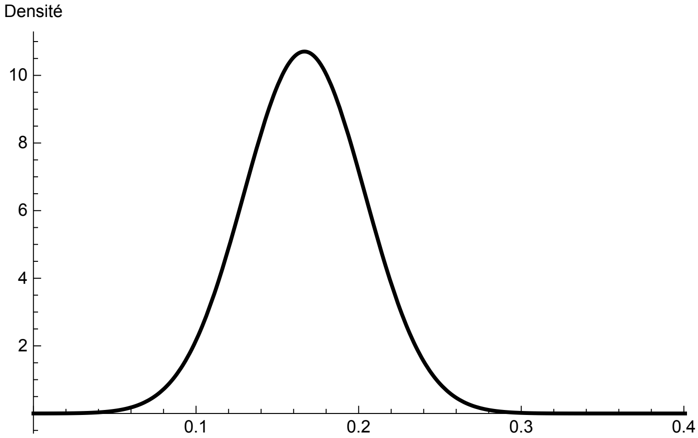
```
k = 100;
p = 1/6;
us = Range[-1/2, k + 1/2];                    (* us = bornes des classes de S *)
     plage
u = us/k;
(* u = bornes des classes de M  *)
c = Range[0, k]/k;
(* c = modalités de M           *)
n = 10 000;
distr = BinomialDistribution[k, p];           (*  distr = distribution de S  *)
        distribution binomiale
x = RandomInteger[distr, n]/k;                 (*   x = échantillon de M      *)
eff = Map[Count[x, #] &, c];
      app· compte
freq = eff/n;
g1 = histogramme[u, freq,
  PlotRange → {{0, 2/5}, All}, Ticks → {Range[0., k, 5.]/k, Automatic},
                          tout    graduations                     automatique
  AxesLabel → {None, "Densité"}, AxesOrigin → {0, 0}]
               aucun              origine des axes
```



Graphique de la fonction "densité de probabilité" :

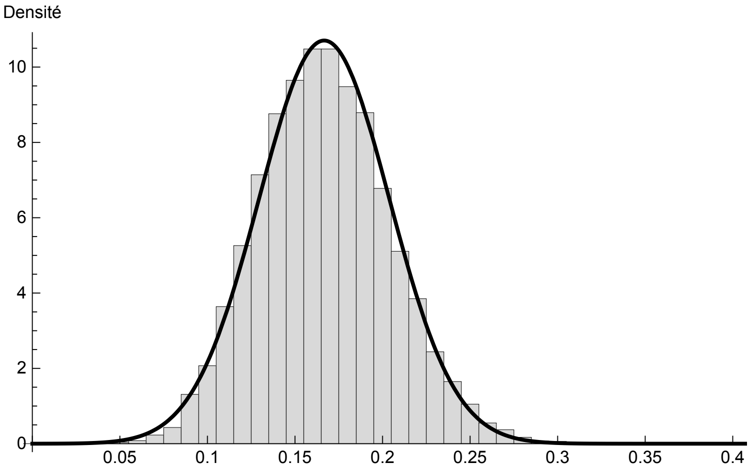
```
μs = k p; σs = Sqrt[k p (1 - p)]; μm = μs/k; σm = σs/k;
normal = NormalDistribution[μm, σm];
                  distribution normale
Clear[t];
effacе
```

```
g2 = Plot[PDF[normal, t], {t, 0, 1}, PlotRange → {{0, 2/5}, All},
     trac…  fonction de densité de probabilité   zone de tracé   tout
  PlotStyle → {Black, Thick}, AxesLabel → {None, "Densité"}]
  style de tracé   noir   épais   titre d'axe   aucun
```



Superposons les deux graphiques:

```
Show[g1, g2, PlotRange → {{0, 2/5}, All},
montre       zone de tracé        tout
 Ticks → {Range[0., k, 5.]/k, Automatic}, AxesLabel → {None, "Densité"}]
 graduations                  automatique   titre d'axe     aucun
```

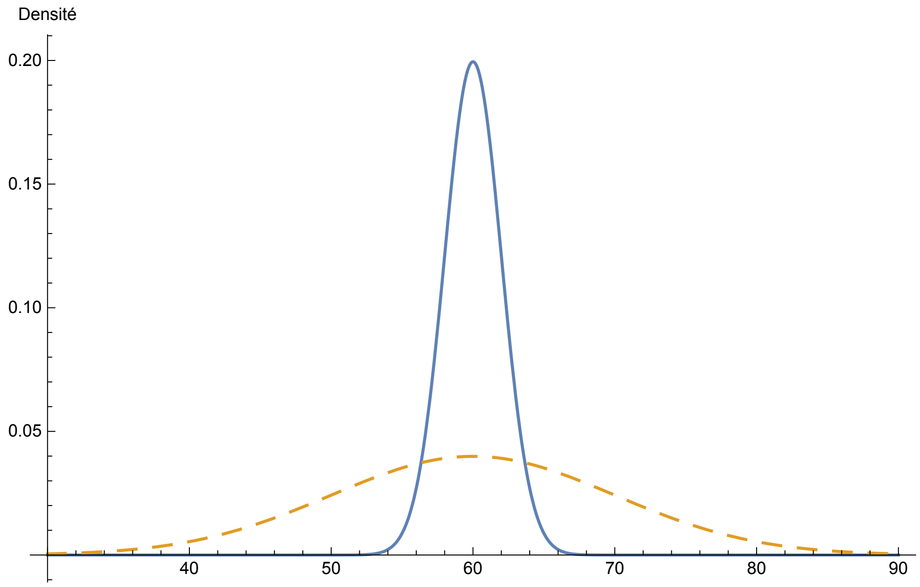


Dans le graphique précédent, l'axe vertical désigne la densité de fréquence et la densité de probabilité. Les écarts entre les densités empiriques et les densités théoriques tendent vers zéro (en probabilité) lorsque la taille n de l'échantillon tend vers l'infini.

Corrigé de l'exercice 4.2 - 4

```
μ = 60; σ1 = 2; σ2 = 10;
distr1 = NormalDistribution[μ, σ1];
          distribution normale
distr2 = NormalDistribution[μ, σ2];
          distribution normale

Plot[{PDF[distr1, t], PDF[distr2, t]},
tracé··· fonction de densité··· fonction de densité de probabilité
 {t, 30, 90}, PlotStyle → {Dashing[{}], Dashing[{0.02`}]},
                style de tracé   style de rayures   style de rayures
 PlotRange → All, AxesLabel → {None, "Densité"}]
 zone de tracé  tout  titre d'axe    aucun
```




```
n = 100;
x1 = RandomReal[distr1, n]
     nombre réel aléatoire
```

{62.029, 63.204, 60.6183, 58.415, 61.8478, 59.339, 60.6492, 61.7088, 59.591, 55.5997, 60.3586, 60.1878, 58.1836, 59.9267, 63.0188, 59.9171, 57.3904, 63.2574, 60.0117, 60.0662, 56.0691, 59.323, 64.2438, 59.0243, 59.789, 62.1895, 58.6399, 58.6061, 58.67, 58.8001, 60.6891, 60.8893, 61.249, 61.0036, 57.4908, 62.0615, 59.0829, 54.5665, 61.1438, 59.5611, 60.3907, 59.4079, 61.5677, 61.2234, 58.9928, 62.092, 59.0767, 61.4668, 60.0199, 61.6375, 62.5364, 64.6468, 60.5742, 59.4523, 61.7667, 58.7825, 60.81, 61.0267, 60.1251, 61.5612, 58.8535, 62.0831, 59.0243, 59.2105, 64.1663, 58.5769, 58.9906, 57.4805, 59.1891, 60.2996, 59.8822, 59.7654, 60.5913, 60.253, 60.7402, 55.8595, 62.4466, 58.422, 60.9149, 61.8058, 64.2004, 61.7809, 57.6198, 59.5071, 60.7722, 59.3351, 60.6235, 59.1509, 60.5228, 58.13, 57.4052, 65.1651, 58.7958, 63.418, 56.3258, 61.4804, 61.0552, 59.1014, 59.1683, 59.5885}

```
x2 = RandomReal[distr2, n]
     nombre réel aléatoire
```

```
{54.5884, 53.2811, 77.7288, 45.5374, 79.1412, 58.8197, 47.0615, 62.8008, 66.2187, 42.1801,
 79.5246, 61.926, 73.1107, 55.8617, 63.794, 46.193, 77.6665, 54.6951, 66.6284,
 59.4605, 71.7291, 69.2173, 67.1606, 57.1426, 46.7822, 64.9038, 63.6124, 69.4355,
 54.2188, 86.9881, 75.1811, 58.2277, 67.7707, 81.8326, 69.2, 57.3312, 55.5378,
 58.786, 52.1056, 51.991, 28.6643, 57.593, 65.0521, 66.3749, 65.9589, 69.9556,
 56.3824, 43.6207, 62.9942, 48.2175, 55.1926, 76.7097, 55.9079, 53.9549, 56.4276,
 58.805, 67.5318, 75.3043, 63.4608, 52.2214, 60.9184, 56.472, 47.6467, 47.9687,
 49.3316, 78.9128, 63.1542, 66.8014, 60.6143, 57.004, 68.9721, 64.6786, 55.423,
 48.9427, 73.0253, 57.6143, 46.5318, 63.202, 71.4492, 57.3808, 66.9415, 63.9203,
 50.0593, 55.7861, 55.0926, 68.4947, 46.6756, 56.4655, 52.5504, 87.3302, 71.9393,
 61.5622, 61.8274, 83.2164, 66.3864, 45.2837, 64.6548, 61.0153, 51.237, 62.8565}
```

Corrigé de l'exercice 4.2 - 5 a)

La méthode graphique consiste à superposer l'histogramme de l'échantillon et la densité normale.

u = Range$\left[0, 4, \frac{1}{2}\right]$
plage

$\left\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \frac{7}{2}, 4\right\}$

```
eff = {113, 277, 278, 191, 97, 34, 10, 0};
```

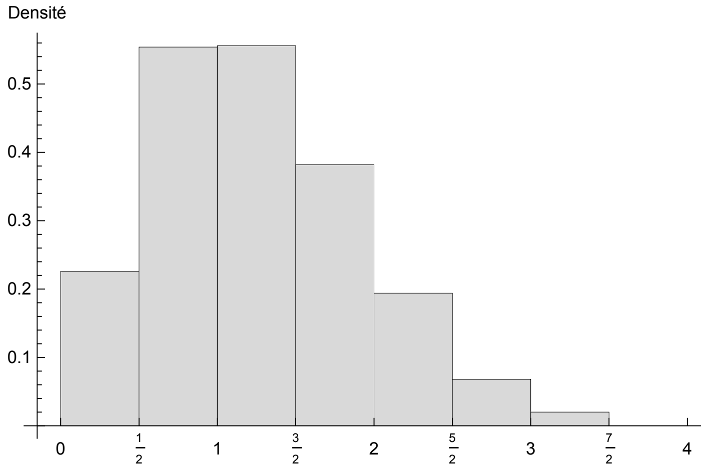
```
n = Apply[Plus, eff]
    remp… plus
```

1000

freq = $\frac{\text{eff}}{\text{n}}$

$\left\{\frac{113}{1000}, \frac{277}{1000}, \frac{139}{500}, \frac{191}{1000}, \frac{97}{1000}, \frac{17}{500}, \frac{1}{100}, 0\right\}$

```
g1 = histogramme[u, freq, AxesLabel → {None, "Densité"}]
                          titre d'axe     aucun
```



```
c = (Drop[u, 1] + Drop[u, -1]) / 2
```

$\left\{\frac{1}{4}, \frac{3}{4}, \frac{5}{4}, \frac{7}{4}, \frac{9}{4}, \frac{11}{4}, \frac{13}{4}, \frac{15}{4}\right\}$

```
m = c.freq; N[m]
```
valeur

1.262

```
me = quantileC[u, freq, 1/2]; N[me]
```
valeur n

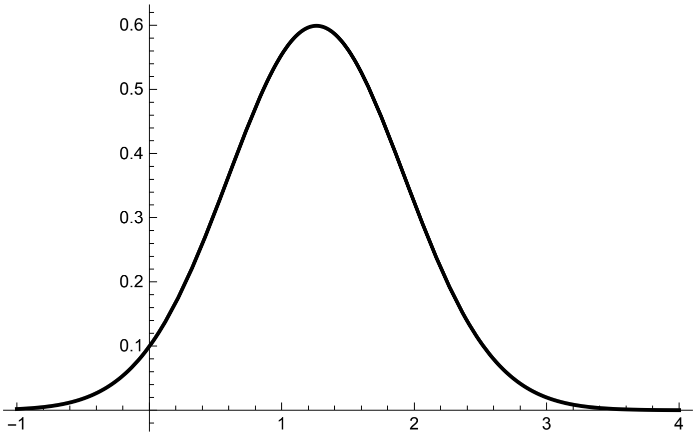
1.19784

```
s = √((c - m)^2.freq); N[s]
```
valeur

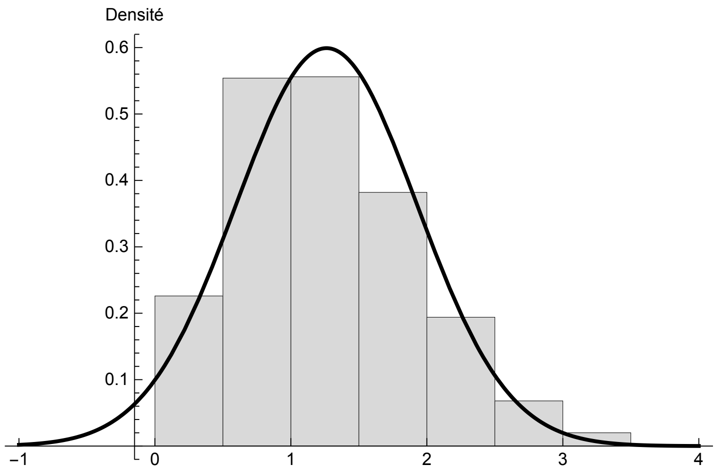
0.66585

```
distr = NormalDistribution[m, s];
```
distribution normale

```
g2 = Plot[PDF[distr, t], {t, -1, 4}, PlotStyle → {Black, Thick}]
```
trac··· fonction de densité de probabilité style de tracé noir épais




```
Show[g1, g2, AxesLabel → {None, "Densité"}, Ticks → {Range[-1, 4], Automatic}]
```
montre titre d'axe aucun graduati··· plage automatique

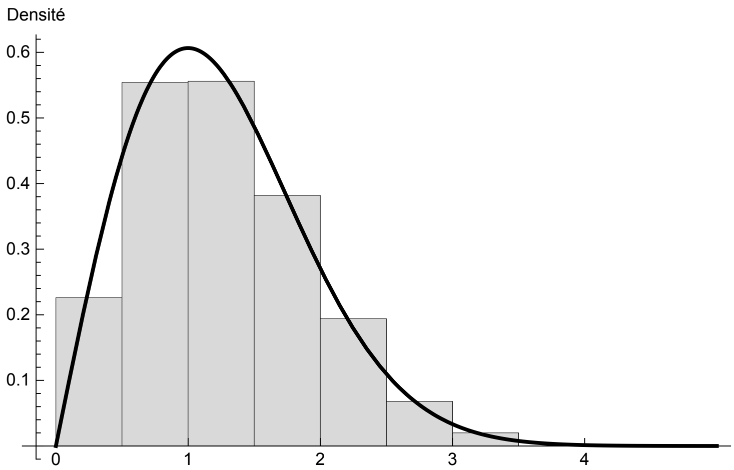


Commentaires : les écarts entre les distributions empirique et théorique sont importants sur les intervalles

$[-1;\ 0[$ et $\left[\frac{1}{2};\ 1\right[$;

l'histogramme n'est pas symétrique par rapport à la moyenne;

la distribution théorique a plutôt la forme suivante:



Réponse : non, la distribution théorique de l'échantillon *a* n'est vraisemblablement pas normale.

Corrigé de l'exercice 4.2 - 5 b)

La méthode graphique consiste à superposer l'histogramme de l'échantillon et la densité normale.

```
u = Range[0, 64, 8]
    plage
```

```
{0, 8, 16, 24, 32, 40, 48, 56, 64}
```

```
eff = {8, 48, 172, 362, 272, 115, 21, 1};
```

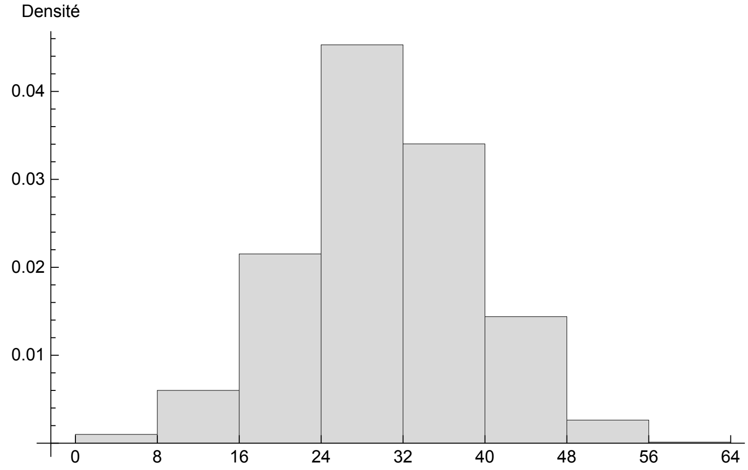
```
n = Apply[Plus, eff]
    remp… plus
```

```
999
```

$$\text{freq} = \frac{\text{eff}}{\text{n}}$$

$$\left\{\frac{8}{999}, \frac{16}{333}, \frac{172}{999}, \frac{362}{999}, \frac{272}{999}, \frac{115}{999}, \frac{7}{333}, \frac{1}{999}\right\}$$

```
g1 = histogramme[u, freq, AxesLabel → {None, "Densité"}]
                          titre d'axe    aucun
```



$$c = \frac{\text{Drop}[u, 1] + \text{Drop}[u, -1]}{2}$$

```
{4, 12, 20, 28, 36, 44, 52, 60}
```

```
m = c.freq; N[m]
            valeur
```

```
30.2182
```

$$me = \text{quantileC}\left[u, \text{freq}, \frac{1}{2}\right]$$

```
30
```

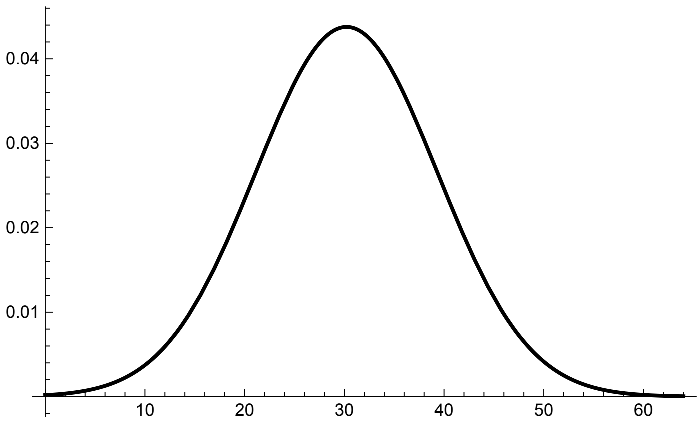
$$s = \sqrt{(c - m)^2.\text{freq}}\ ; N[s]$$

valeur

```
9.1126
```

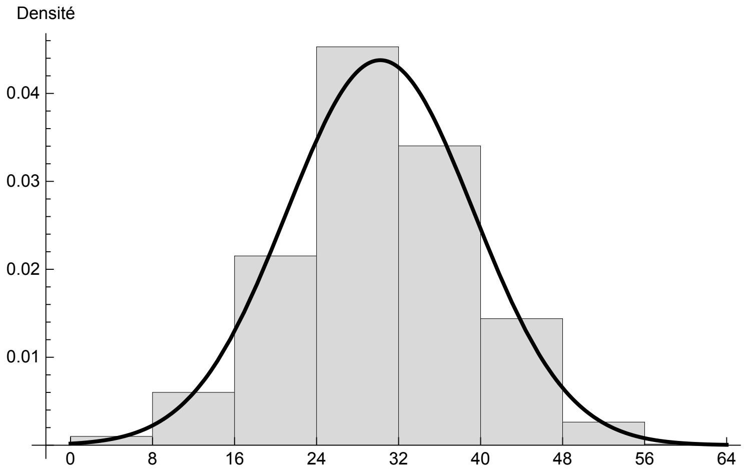
```
distr = NormalDistribution[m, s];
        distribution normale
```

```
g2 = Plot[PDF[distr, t], {t, 0, 64}, PlotStyle → {Black, Thick}]
     trac… fonction de densité de probabilité  style de tracé  noir  épais
```



```
Show[g1, g2]
```

montre



Commentaires : l'histogramme est une approximation convenable de la densité normale.

Réponse : oui, la distribution normale est un modèle acceptable pour l'échantillon *b*.

Corrigé de l'exercice 4.2 - 6

```
μ = 1.275; σ = 0.013;
```

```
distr = NormalDistribution[μ, σ];
```

distribution normale

Fonction de distribution normale

```
ϕ[t_] := CDF[distr, t];
```

fonction de distributi

Pourcentage de rondelles dans les bornes de tolérance

$$p = P([1.26;\ 1.29])$$

```
p = ϕ[1.29] - ϕ[1.26]
```

```
0.751437
```

Pourcentage de rondelles défectueuses

```
1 - p
```

```
0.248563
```

Corrigé de l'exercice 4.2 - 7

$p = \frac{1}{2}$; $k = 500$;

$\mu = k\,p$; $\sigma = \sqrt{k\,p\,(1-p)}$;

```
distr = NormalDistribution[μ, σ];
```

distribution normale

Fonction de distribution normale

```
ϕ[t_] := CDF[distr, t];
```

fonction de distributi

Probabilité pour que le nombre de faces soit situé entre 240 et 260, approximée par la distribution normale

p ≈ P ([239.5; 260.5])

```
p = ϕ[260.5] - ϕ[239.5]
```

0.652346

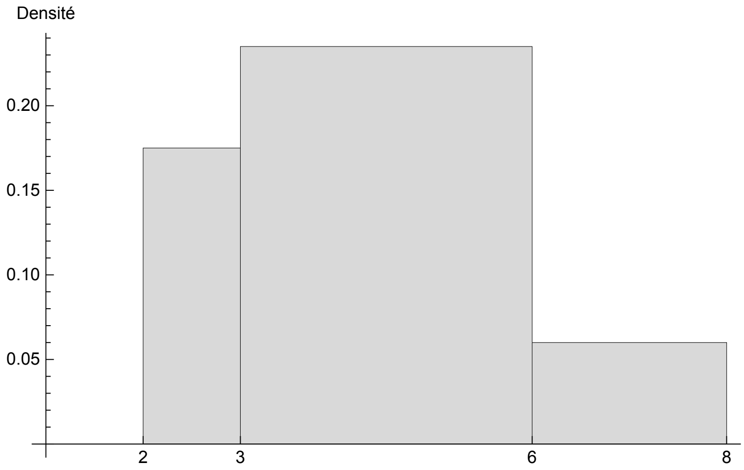
La valeur exacte a été calculée avec la distribution binomiale dans l'exercice 3.2 - 3

0.652336

Corrigé de l' exercice 4.3 - 1 a)

$$u = \{2, 3, 6, 8\};\ p = \left\{\frac{7}{40}, \frac{141}{200}, \frac{3}{25}\right\};$$

```
histogramme[u, p, AxesOrigin → {1, 0}, AxesLabel → {None, "Densité"}]
                  origine des axes      titre d'axe    aucun
```



Milieux des classes

```
c = {2.5, 4.5, 7};
```

Espérance mathématique (= moyenne théorique)

```
μ = c.p
```

4.45

Corrigé de l'exercice 4.3 - 1 b)

```
va[] := Module[{z},
        module
  z = Random[UniformDistribution[{0, 1}]];
      aléatoire distribution uniforme
  Which[z < 0.705, Random[UniformDistribution[{3, 6}]],
  quel             aléatoire distribution uniforme
   z < 0.88, Random[UniformDistribution[{2, 3}]],
             aléatoire distribution uniforme
   True, Random[UniformDistribution[{6, 8}]]]]
   vrai  aléatoire distribution uniforme
```

```
n = 100000;
x = Table[va[], {n}];
```

```
m = Mean[x]; N[m]
```

4.45298

```
s = StandardDeviationMLE[x]; N[s]
```

1.4237

Corrigé de l'exercice 4.4 - 1

La fonction est un polynôme du deuxième degré en *t* dont nous déterminons les coefficients:

$$f(t) = \frac{1}{n}\sum_{i=1}^{n}(x_i - t)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i t + t^2\right) = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) - \left(2\frac{1}{n}\sum_{i=1}^{n}x_i\right)t + \left(\frac{1}{n}\sum_{i=1}^{n}t^2\right) =$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) - (2m)\,t + \left(\frac{1}{n}\left(n\,t^2\right)\right) = t^2 - 2mt + \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) = a\,t^2 + b\,t + c$$

$$\text{où} \qquad a = 1 > 0; \qquad b = -2m; \qquad c = \frac{1}{n}\sum_{i=1}^{n}x_i^2$$

La parabole est ouverte vers le haut. La fonction admet un minimum en son sommet dont l'abscisse est

$$x_S = -\frac{b}{2a} = -\frac{(-2m)}{2} = m$$

Corrigé de l'exercice 4.4 - 2

Partie a)

Moyenne empirique

$$m = \frac{95 + 84 + 76 + 82 + 87}{5} = \frac{424}{5} \simeq 84.8$$

Ecart-type empirique

$$s = \sqrt{\left(\frac{1}{5}\left(\left(95 - \frac{424}{5}\right)^2 + \left(84 - \frac{424}{5}\right)^2 + \left(76 - \frac{424}{5}\right)^2 + \left(82 - \frac{424}{5}\right)^2 + \left(87 - \frac{424}{5}\right)^2\right)\right)} =$$

$$\frac{\sqrt{974}}{5} \simeq 6.24179$$

Partie b)

```
x = {95, 84, 76, 82, 87};
```

Moyenne empirique

```
m = Mean[x]
```

$$\frac{424}{5}$$

```
N[m]
```
valeur numérique

84.8

Ecart-type empirique

```
s = StandardDeviationMLE[x]
```

$$\frac{\sqrt{974}}{5}$$

```
N[s]
```
valeur numérique

6.24179

Partie c)

Estimation de l'espérance mathématique

$$\hat{\mu} = m = \frac{424}{5} \simeq 84.8$$

Estimation de l'écart-type théorique

$$\hat{\sigma} = \sqrt{\left(\frac{1}{4}\left(\left(95-\frac{424}{5}\right)^2+\left(84-\frac{424}{5}\right)^2+\left(76-\frac{424}{5}\right)^2+\left(82-\frac{424}{5}\right)^2+\left(87-\frac{424}{5}\right)^2\right)\right)} =$$

$$\sqrt{\frac{487}{10}} \simeq 6.97854$$

Partie d)

```
x = {95, 84, 76, 82, 87};
```

Estimation de l'espérance mathématique

```
μ = Mean[x]
```
valeur moy

$$\frac{424}{5}$$

```
N[μ]
```
valeur numérique

84.8

Estimation de l'écart-type théorique

```
σ = StandardDeviation[x]
```
écart-type

$$\sqrt{\frac{487}{10}}$$

```
N[σ]
```
valeur numérique

6.97854