

§ 4 Distributions théoriques à une variable continue

■ Packages locaux

A partir de la commande `aide[]`, charger

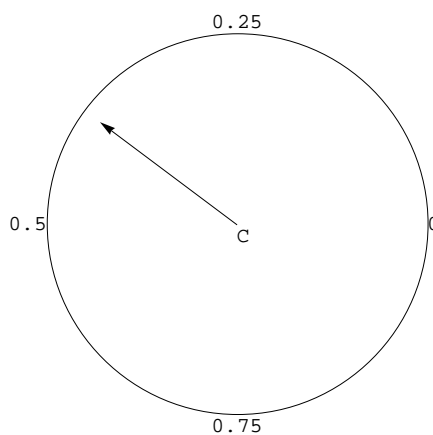
```
<< Statistique`
```

```
<< Tableaux`
```

§ 4.1 Distribution uniforme

■ Expérience aléatoire

Sur un cadran gradué en tours, on lance une aiguille pivotant autour du centre C qui, après quelques tours, s'arrête sur une position aléatoire décrite par un nombre réel de l'intervalle $[0; 1[$.



■ Simulation et nombres pseudo-aléatoires

Remplaçons l'expérience par une simulation. Générons un échantillon de taille n :

```
n = 40;
```

```
? UniformDistribution
```

UniformDistribution[{*min*, *max*}] represents a continuous uniform statistical distribution giving values between *min* and *max*. »

? RandomReal

RandomReal[] gives a pseudorandom real number in the range 0 to 1.
 RandomReal[{ x_{\min} , x_{\max} }] gives a pseudorandom real number in the range x_{\min} to x_{\max} .
 RandomReal[x_{\max}] gives a pseudorandom real number in the range 0 to x_{\max} .
 RandomReal[range, n] gives a list of n pseudorandom reals.
 RandomReal[range, { n_1 , n_2 , ...}] gives an $n_1 \times n_2 \times \dots$ array of pseudorandom reals.
 RandomReal[dist, ...] samples from the symbolic continuous distribution dist. »

```
unif = UniformDistribution[{0, 1}];
```

```
x = RandomReal[unif, n]
```

```
{0.439718, 0.398533, 0.630744, 0.979984, 0.146217, 0.0863413, 0.624052, 0.226531,
0.223894, 0.404439, 0.808872, 0.529708, 0.0968376, 0.980971, 0.266152, 0.813695,
0.546484, 0.760167, 0.607438, 0.698479, 0.0455955, 0.342857, 0.556776, 0.109555,
0.311776, 0.738897, 0.0212202, 0.665896, 0.914848, 0.58573, 0.415591, 0.589492,
0.0775477, 0.00502102, 0.475898, 0.950306, 0.706696, 0.716376, 0.210086, 0.640736}
```

Comme dans une expérience aléatoire, une nouvelle exécution donne habituellement d'autres résultats

```
x = RandomReal[unif, n]
```

```
{0.967761, 0.581698, 0.583742, 0.568875, 0.249594, 0.947597, 0.223746, 0.241367,
0.761044, 0.509062, 0.048778, 0.859043, 0.479469, 0.983595, 0.809885, 0.400238,
0.134955, 0.0801806, 0.52072, 0.212442, 0.729779, 0.0636717, 0.760921, 0.207484,
0.342948, 0.353326, 0.447068, 0.750724, 0.0655731, 0.411838, 0.250797, 0.335637,
0.0288569, 0.375971, 0.271862, 0.419907, 0.180578, 0.348748, 0.500565, 0.862747}
```

■ Probabilité

Notons X la variable aléatoire uniformément distribuée sur l'intervalle $[0; 1[$ et exprimons la probabilité que X appartienne à un intervalle donné

$$P[0.2 \leq X < 0.3] = \frac{0.3 - 0.2}{1} = \frac{1}{10}$$

$$P[0.2 \leq X < 0.25] = \frac{0.25 - 0.2}{1} = \frac{1}{20}$$

$$P[0.2 \leq X < 0.201] = \frac{0.201 - 0.2}{1} = \frac{1}{1000}$$

La probabilité de tirer exactement un nombre donné est nulle

$$P[X = 0.201] = 0$$

Lorsque la variable X est continue, on peut parler de *la probabilité d'un intervalle*, peu importe s'il est ouvert ou fermé

$$P[0.2 \leq X < 0.3] = P[0.2 < X < 0.3] = P[0.2 \leq X \leq 0.3]$$

■ Modèle théorique

La distribution uniforme sur l'intervalle $[a, b[$ est caractérisée par son histogramme qui se réduit à un rectangle. L'aire du rectangle étant 1 et sa largeur $b - a$, sa hauteur est

$$h_1 = \frac{1}{b - a}$$

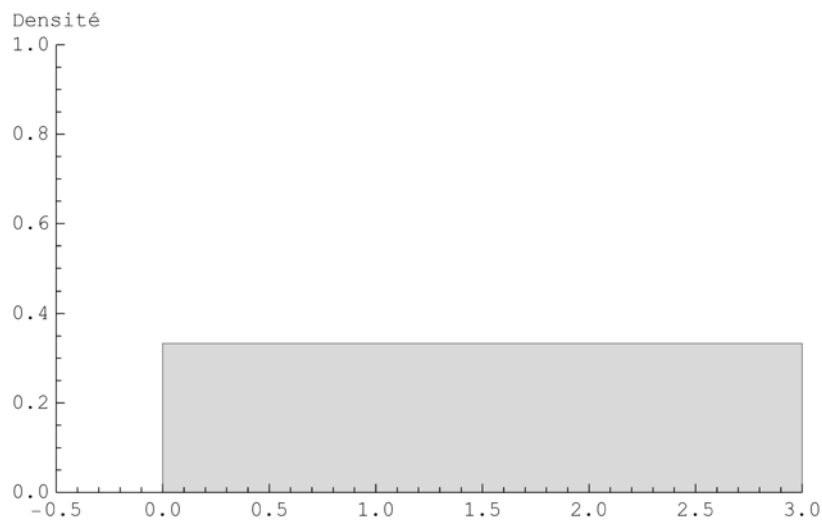
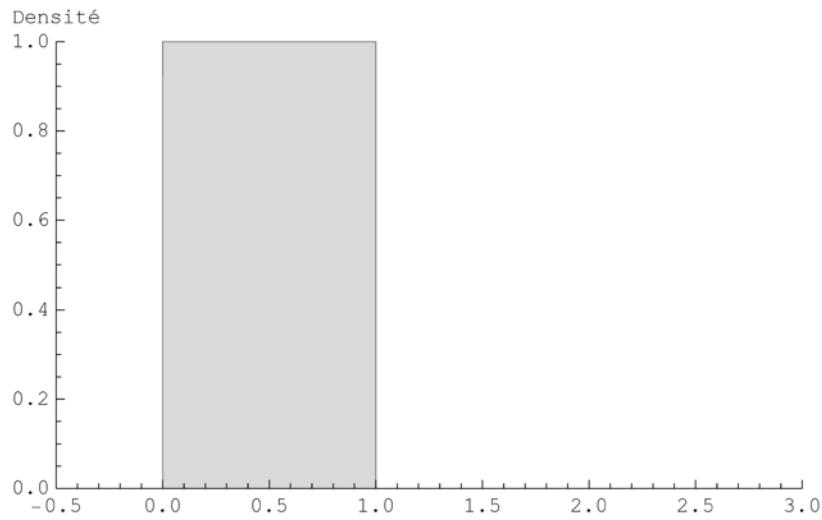
La fonction densité est donc

$$h(x) = \begin{cases} h_1 & \text{pour } a \leq x < b \\ 0 & \text{sinon} \end{cases}$$

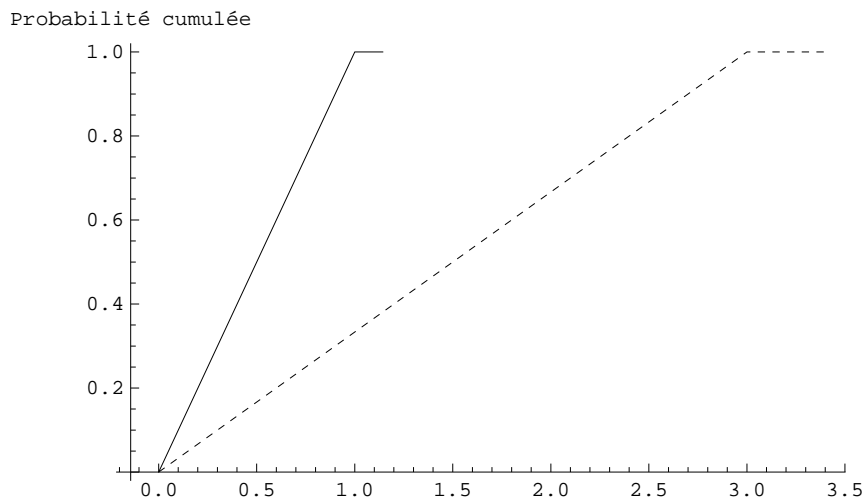
Voici par exemple deux histogrammes successifs

le premier représente la distribution uniforme sur $[0; 1[$,

le second la distribution uniforme sur $[0; 3[$.



Remarquez que les aires sont toutes deux égales à 1. La fonction qui exprime les probabilités cumulées est appelé distribution. La distribution uniforme sur $[0; 1[$ est représentée ci-dessous en trait continu; la seconde, sur $[0; 3[$, en traitillé.



Remarquez que les distributions uniformes sont continues.

■ Espérance et écart-type de la distribution continue uniforme

```
Clear[a, b]
```

```
distr = UniformDistribution[{a, b}];
```

La moyenne de la distribution uniforme est la valeur centrale:

```
 $\mu = \text{Mean}[\text{distr}]$ 
```

$$\frac{a + b}{2}$$

Le calcul de l'écart-type des distributions continues dépasse les ambitions de ce cours. Mais le résultat nous sera utile:

```
 $\sigma = \text{Variance}[\text{distr}]$ 
```

$$\frac{1}{12} (b - a)^2$$

■ Exercice 4.1 - 1

Dans le cas particulier de l'intervalle $[a, b] = [0, 1]$, déterminez la moyenne empirique m et la variance empirique s^2 d'un échantillon obtenu par simulation.

Comparez les résultats obtenus avec l'espérance mathématique $\mu = E(X) = \frac{1}{2}$ et la variance théorique $\sigma^2 = V(X) = \frac{1}{12}$.

Le but est de se convaincre que les résultats théoriques sont plausibles.

Facultativement : écrivez la limite à calculer pour obtenir la variance théorique.

§ 4.2 Distribution normale

■ Introduction

D'une manière générale, la fonction densité $h(x)$ se caractérise par sa moyenne, son écart-type et sa forme. La forme de h , c'est-à-dire son allure, peut être quelconque; en particulier, elle peut être symétrique ou asymétrique. Certaines formes particulières apparaissent souvent dans les applications ou dans les modèles mathématiques. La plus importante est la distribution normale dont la fonction de densité est aussi appelée "courbe en cloche de Gauss". Nous allons la construire en partant de la distribution binomiale.

Distribution binomiale

Partons du problème suivant

on lance k fois un dé et on compte le nombre de six obtenus.

Les résultats possibles sont $0, 1, 2, \dots, k$. La distribution est donc discrète.

Chaque issue a une probabilité déterminée

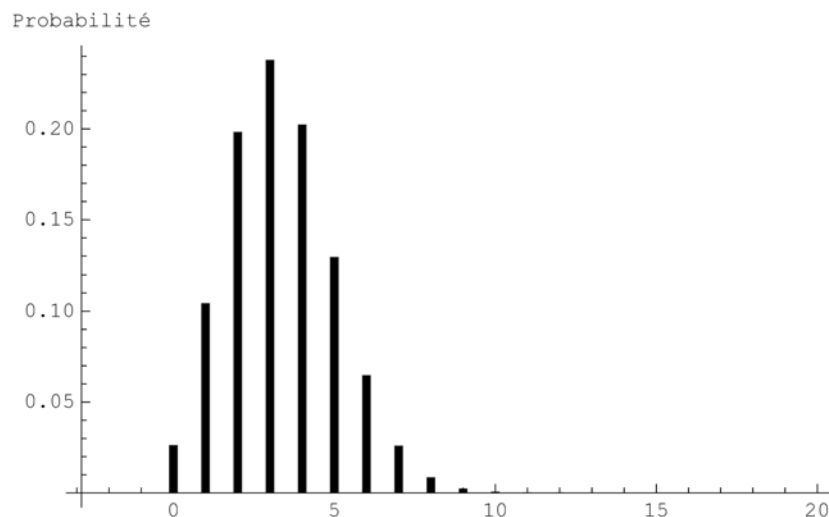
$$p_0 = P\{X = 0\}, \dots, p_k = P\{X = k\}$$

Nous avons déjà résolu le problème dans le cas particulier $k = 3$ (voir § 3.2). Comme il n'est pas dans l'objectif de ce cours de savoir calculer les probabilités par les mathématiques, nous nous contenterons ici de demander les résultats à *Mathematica*.

Distribution discrète

```
k = 20; p = 1/6;
distr = BinomialDistribution[k, p];
c = Range[0, k];
freq = Table[PDF[distr, x], {x, 0, k}];

diagrammeBats[c, freq, AxesLabel -> {None, "Probabilité"}]
```



L'espérance mathématique et l'écart-type théorique de la distribution binomiale sont donnés par les formules suivantes (qui seront démontrées dans *Statistique II*, § 2) :

$$\mu = k p; \text{ N}[\mu]$$

$$3.33333$$

$$\sigma = \sqrt{k p (1 - p)}; \text{ N}[\sigma]$$

$$1.66667$$

■ Transformation de la distribution binomiale en une distribution continue ou distribution binomiale interpolée

Notre but étant de construire une distribution continue, nous allons commencer par rendre continue la distribution précédente. Utilisons la méthode qui consiste à "étaler" le bâton en j sur l'intervalle $[j - \frac{1}{2}, j + \frac{1}{2}[$ avec la même fréquence f_j . La densité correspondante sera donc $h_j = \frac{f_j}{1}$. La distribution obtenue s'appelle "*distribution binomiale interpolée*".

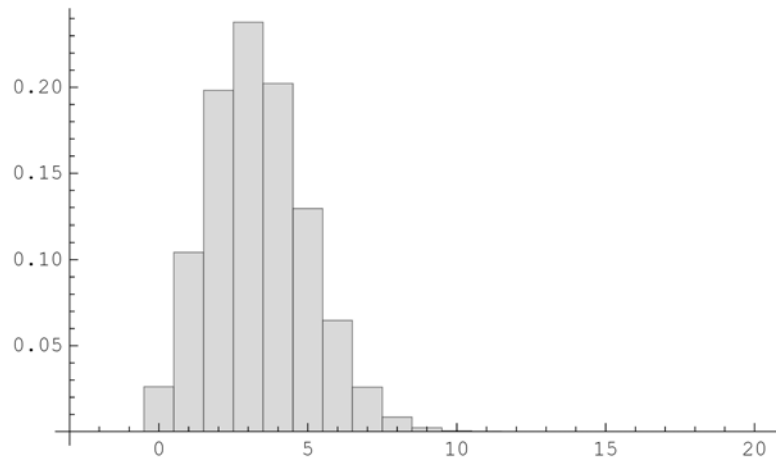
```

k = 20; p =  $\frac{1}{6}$ ;
distr = BinomialDistribution[k, p];
b = Range[- $\frac{1}{2}$ , k +  $\frac{1}{2}$ ];
freq = Table[PDF[distr, j], {j, 0, k}];

histogramme[b, freq,
  AxesLabel → {None, "Densité de probabilité"}, AxesOrigin → {-3, 0}]

```

Densité de probabilité



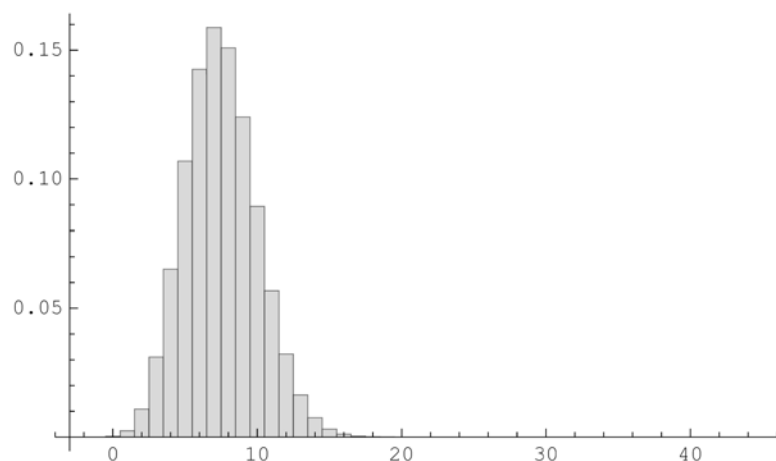
```

k = 45; p =  $\frac{1}{6}$ ;
distr = BinomialDistribution[k, p];
b = Range[- $\frac{1}{2}$ , k +  $\frac{1}{2}$ ];
freq = Table[PDF[distr, j], {j, 0, k}];

histogramme[b, freq,
  AxesLabel → {None, "Densité de probabilité"}, AxesOrigin → {-3, 0}]

```

Densité de probabilité



On peut observer que les densités ne sont pas symétriques par rapport à la moyenne, particulièrement lorsque k est petit.

■ Variables centrées réduites

Comment comparer les formes (ou allures) des deux distributions précédentes alors que leurs moyennes et leurs écarts-types sont différents?

L'idée est de leur donner

le même centre en leur soustrayant l'espérance mathématique et
la même dispersion en divisant par l'écart-type :

$$Z = \frac{X - \mu}{\sigma} \quad \text{où} \quad \mu = E(X) \quad \text{et} \quad \sigma^2 = V(X)$$

On obtient alors une nouvelle distribution appelée *centrée réduite* dont l'espérance mathématique est nulle (voir les propriétés de E dans le § 3.4)

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = E\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0$$

et dont la variance est 1 (voir les propriétés de V dans § 3.4)

$$V(Z) = V\left(\frac{X - \mu}{\sigma}\right) = V\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$

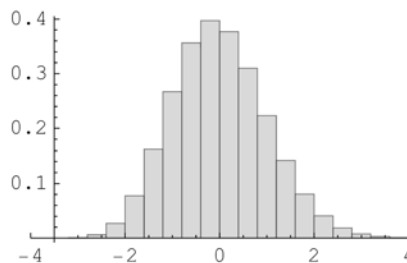
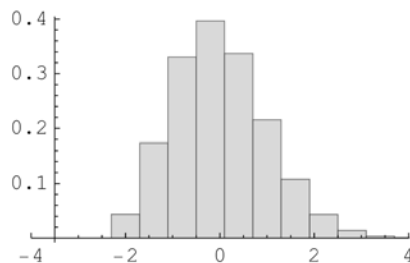
Comparons les deux histogrammes:

```

k = 20; p = 1/6;
distr = BinomialDistribution[k, p];
b = Range[-1/2, k + 1/2];
freq1 = Table[PDF[distr, j], {j, 0, k}];
μ = k p;
σ = Sqrt[k p (1 - p)];
z1 = (b - μ)/σ;

k = 45; p = 1/6;
distr = BinomialDistribution[k, p];
b = Range[-1/2, k + 1/2];
freq2 = Table[PDF[distr, j], {j, 0, k}];
μ = k p;
σ = Sqrt[k p (1 - p)];
z2 = (b - μ)/σ;

GraphicsRow[{histogramme[z1, freq1, PlotRange -> {{-4, 4}, All},
  AxesOrigin -> {-3.5, 0}, ImageSize -> {200, 120}], histogramme[z2, freq2,
  PlotRange -> {{-4, 4}, All}, AxesOrigin -> {-3.5, 0}, ImageSize -> {200, 120}]}]
```



■ Exercice 4.2 - 1 [Sans ordinateur]

Un étudiant a obtenu 5.2 à l'examen final de mathématiques pour lequel la moyenne était de 4.8 et l'écart-type de 0.5.

A l'examen final de physique, dont la moyenne était de 5.1 et l'écart-type de 0.8, il a obtenu 5.5.

Dans quelle discipline était-il relativement meilleur ?

Indication : comparez les variables centrées réduites.

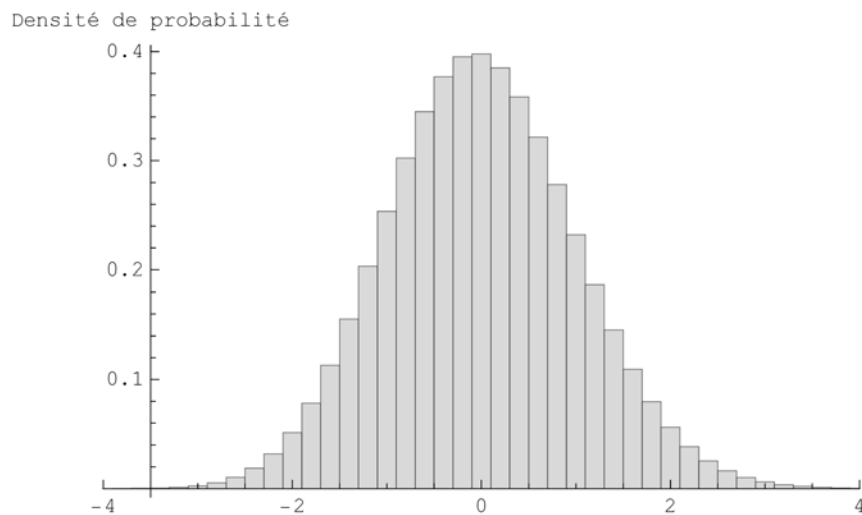
■ Distribution normale (0, 1)

Pour une valeur plus grande de k , la fonction *densité de la distribution binomiale interpolée* tend vers une courbe qui est pratiquement symétrique par rapport à la moyenne

```

k = 180; p = 1/6;
distr = BinomialDistribution[k, p];
b = Range[-1/2, k + 1/2];
freq = Table[PDF[distr, j], {j, 0, k}];
μ = k p;
σ = Sqrt[k p (1 - p)];
z = (b - μ)/σ;
histogramme[z, freq, PlotRange -> {{-4, 4}, All},
  AxesLabel -> {None, "Densité de probabilité"}, AxesOrigin -> {-3.5, 0}]

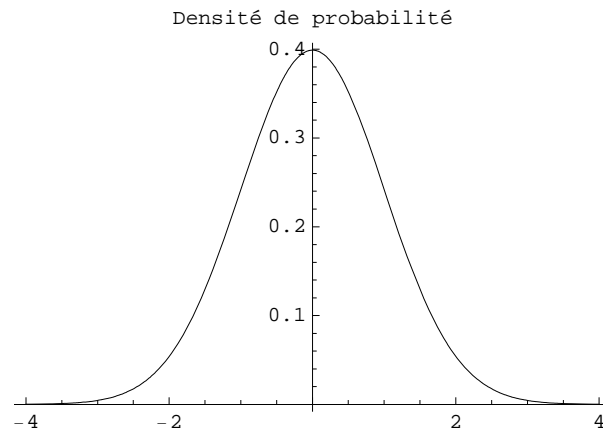
```



Les mathématiciens démontrent que, quel que soit p , la densité de la distribution binomiale tend vers une courbe bien définie appelée *courbe en cloche* ou *courbe de Gauss* dont l'expression mathématique est

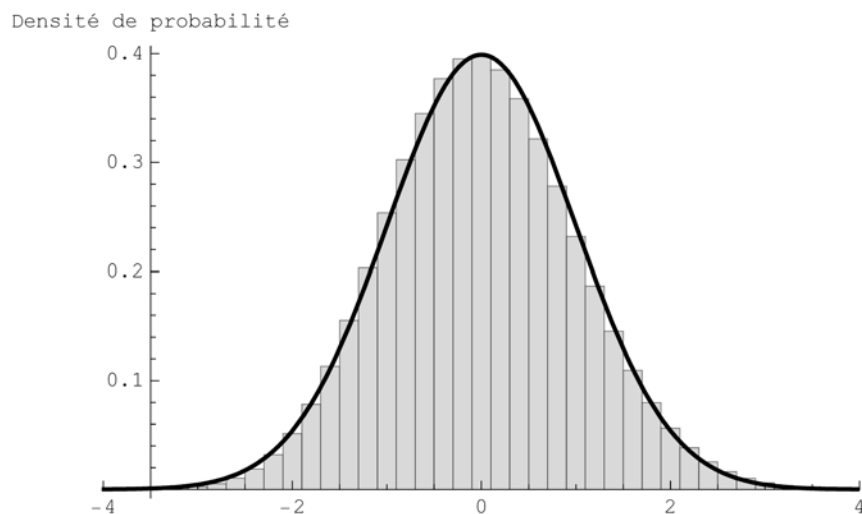
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$


```
distr = NormalDistribution[0, 1];
Plot[PDF[distr, x], {x, -4, 4}, PlotStyle -> Black,
  ImageSize -> {400, 200}, AxesLabel -> {None, "Densité de probabilité"}]
```



Superposons les deux derniers graphiques

```
Show[histogramme[z, freq],
  Plot[PDF[distr, t], {t, -4, 4}, PlotStyle -> {Black, Thick}],
  PlotRange -> {{-4, 4}, All}, AxesOrigin -> {-3.5, 0},
  AxesLabel -> {None, "Densité de probabilité"}]
```



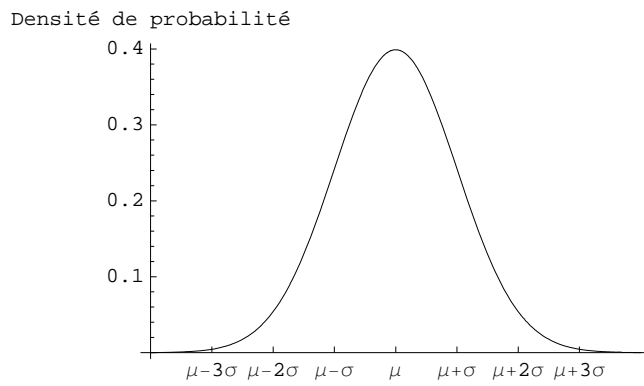
Si $k p \geq 5$ et $k(1 - p) \geq 5$ alors la distribution normale constitue une bonne approximation de la distribution binomiale.

■ Distribution normale (μ, σ)

Si Z désigne la distribution normale $(0, 1)$ c'est-à-dire avec $E(Z) = 0$ et $V(Z) = 1$, alors la variable $X = \sigma Z + \mu$ a la distribution normale (μ, σ) car

$$\begin{aligned} E(X) &= E(\sigma Z + \mu) = \sigma E(Z) + \mu = \sigma \cdot 0 + \mu = \mu \\ V(X) &= V(\sigma Z + \mu) = \sigma^2 V(Z) = \sigma^2 \cdot 1 = \sigma^2 \end{aligned}$$

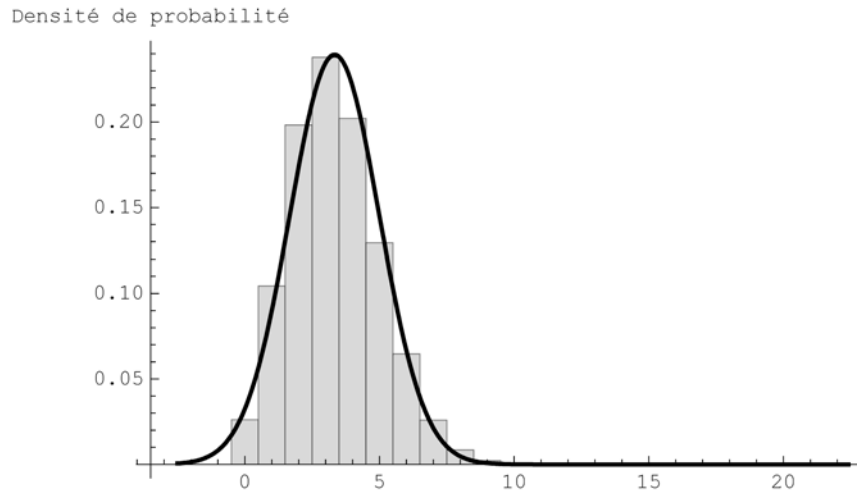
Les distributions normales constituent une famille dont on peut choisir librement l'espérance mathématique μ et l'écart-type théorique σ .



Puisque cette courbe est une densité, l'aire de la surface comprise entre la courbe et l'axe horizontal est 1.

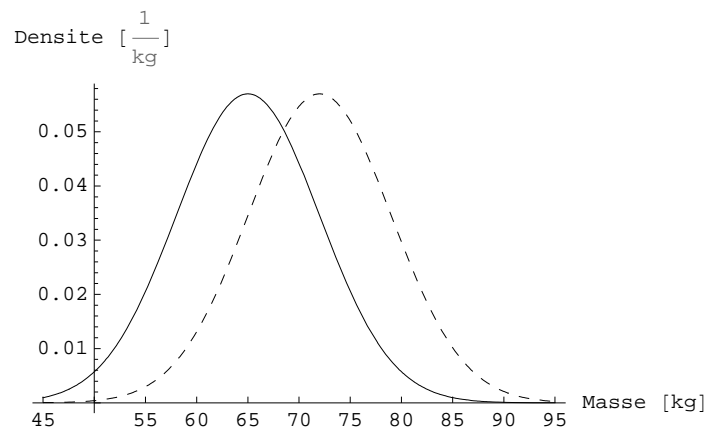
A titre d'exemple, superposons cette courbe avec la distribution binomiale $(20; \frac{1}{6})$ interpolée:

```
k = 20; p = 1/6;
b = Range[-1/2, k + 1/2];
freq = Table[PDF[BinomialDistribution[k, p], c], {c, 0, k}];
μ = k p;
σ = Sqrt[k p (1 - p)];
Show[histogramme[b, freq], Plot[PDF[NormalDistribution[μ, σ], x],
  {x, First[b] - 2, Last[b] + 2}, PlotStyle -> {Black, Thick}],
  AxesLabel -> {None, "Densité de probabilité"}, AxesOrigin -> {-3.5, 0}]
```



Rôle de $\mu = E(X)$

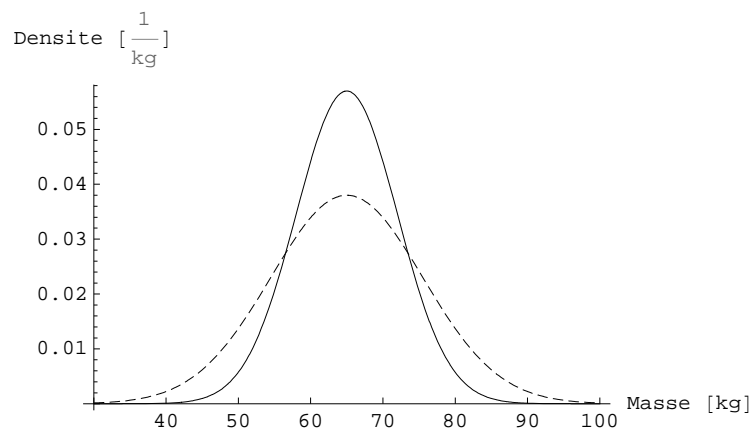
Graphiquement, une augmentation de l'espérance mathématique se traduit par une translation de la densité vers la droite (courbe en traitillé)



Rôle de $\sigma = \sqrt{V(X)}$

Graphiquement, une augmentation de l'écart-type théorique se traduit par un étalement de la densité c'est-à-dire une augmentation de la dispersion (courbe en traitillé). Rappelons que l'aire de la surface emprisonnée entre la courbe et l'axe horizontal est égale à 1.

```
distr3 = NormalDistribution[μ,  $\frac{3\sigma}{2}$ ];
Plot[{PDF[distr, t], PDF[distr3, t]}, {t, 30, 100},
  PlotStyle → {{Black, Dashing[{]}}, {Black, Dashing[{0.02, 0.01}]}},
  Ticks → {Range[30, 100, 10], Automatic}, ImageSize → {400, 200},
  AxesLabel → {"Masse [kg]", "Densité [ $\frac{1}{\text{kg}}$ ]"}]
```



■ Probabilité d'un intervalle

Rappels

Pour les distributions empiriques, nous avons vu, dans le § 2.1, que la distribution est liée à la fréquence par les relations

$$F(x) = f(-\infty, x]$$

$$f(a, b] = F(b) - F(a)$$

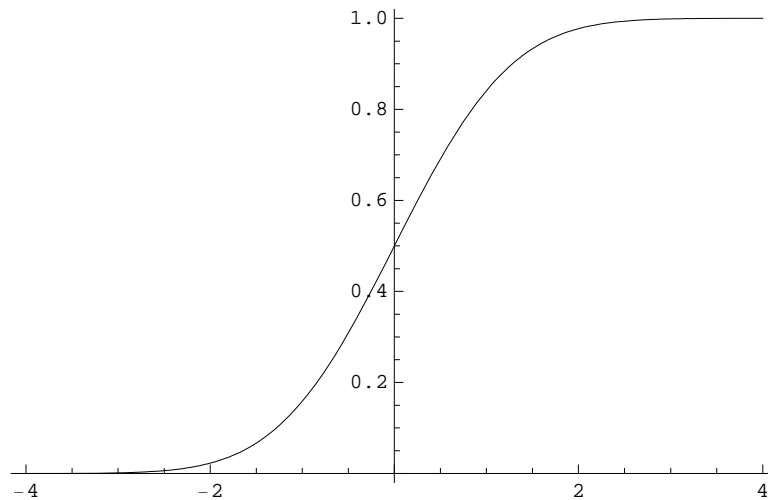
Cette dernière relation peut s'interpréter comme représentant l'aire de la surface délimitée par la fonction densité au-dessus de l'intervalle $]a, b]$ (voir fig. dans le § 2.1).

Distribution et densité normales

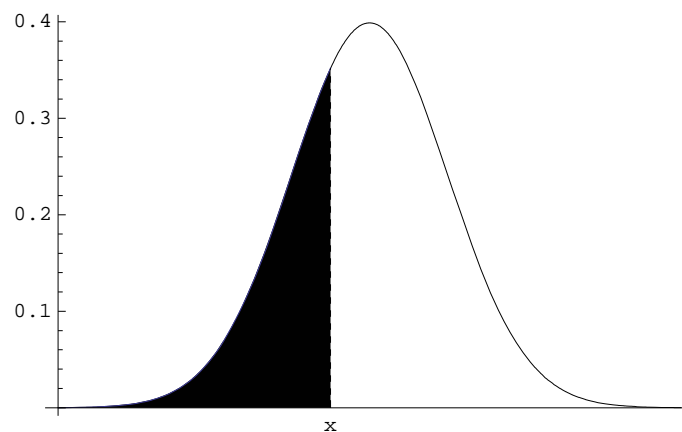
Les distributions théoriques entretiennent les mêmes relations avec les probabilités. Notons Φ la fonction de distribution normale et ϕ la fonction densité normale. La valeur de la distribution en x est égale à la probabilité de l'événement $] -\infty, x]$.

$$\Phi(x) = P[-\infty, x]$$

```
Plot[CDF[NormalDistribution[0, 1], x], {x, -4, 4}, PlotStyle -> Black]
```



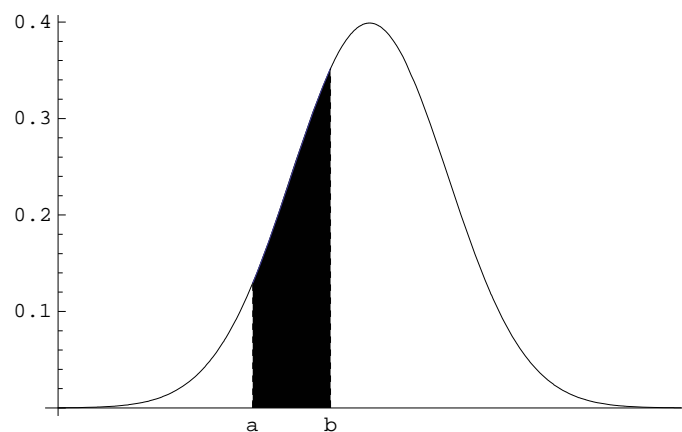
Graphiquement, la fonction de distribution en x s'interprète comme l'aire de la surface située au-dessus de l'intervalle $]-\infty, x]$ et délimitée par la fonction densité.



La probabilité d'un intervalle peut se déduire de la fonction distribution

$$P(]a, b]) = \Phi(b) - \Phi(a)$$

Graphiquement, la probabilité d'un intervalle s'interprète comme l'aire de la surface située au-dessus de l'intervalle et délimitée par la fonction densité.

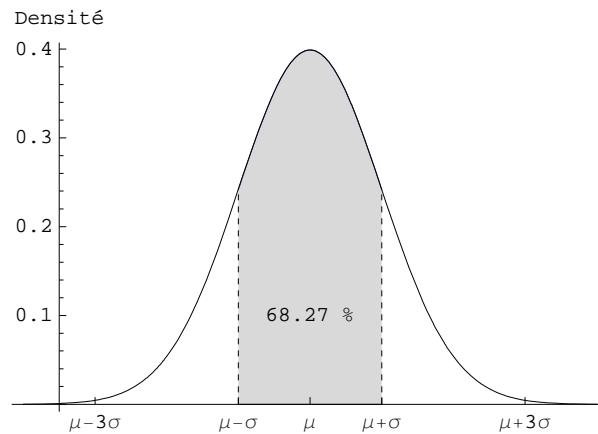


Calculons la probabilité de l'événement X appartient à l'intervalle $]\mu - \sigma, \mu + \sigma]$:

$$P] \mu - \sigma, \mu + \sigma] = P] - \infty, \mu + \sigma] - P] - \infty, \mu - \sigma] = \Phi(\mu + \sigma) - \Phi(\mu - \sigma)$$

```
distr = NormalDistribution[0, 1];
N[CDF[distr, 1] - CDF[distr, -1]]
0.682689
```

Cette probabilité est d'environ 68.27 %.

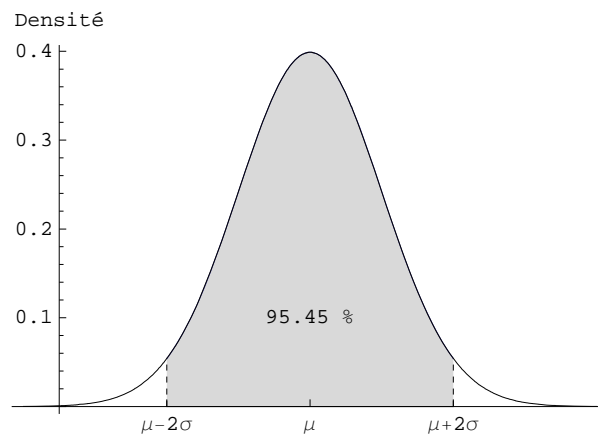


Calculons la probabilité de l'événement X appartient à l'intervalle $]\mu - 2\sigma, \mu + 2\sigma]$

$$P] \mu - 2\sigma, \mu + 2\sigma] = P] - \infty, \mu + 2\sigma] - P] - \infty, \mu - 2\sigma] = \Phi(\mu + 2\sigma) - \Phi(\mu - 2\sigma)$$

```
distr = NormalDistribution[0, 1];
N[CDF[distr, 2] - CDF[distr, -2]]
0.9545
```

Cette probabilité est d'environ 95.45 %.

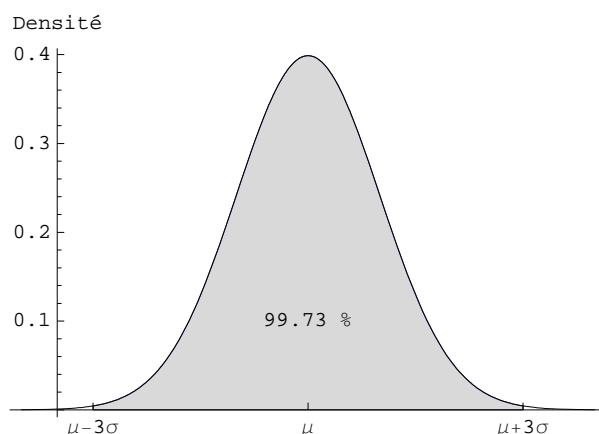


Calculons la probabilité de l'événement X appartient à l'intervalle $]\mu - 3\sigma, \mu + 3\sigma]$

$$P] \mu - 3\sigma, \mu + 3\sigma] = P] - \infty, \mu + 3\sigma] - P] - \infty, \mu - 3\sigma] = \Phi(\mu + 3\sigma) - \Phi(\mu - 3\sigma)$$

```
distr = NormalDistribution[0, 1];
N[CDF[distr, 3] - CDF[distr, -3]]
0.9973
```

Cette probabilité est d'environ 99.73 %.



Retenons la règle des 3 σ

la quasi totalité des observations sont situées dans l'intervalle $]\mu - 3\sigma, \mu + 3\sigma]$.

■ Exercice 4.2 - 2 [Avec *Mathematica*]

Nous nous proposons de comparer une distribution binomiale et une distribution normale.

On lance k pièces de monnaie ($k = 32$) et on compte le nombre de "pile" obtenus.

Dressez un diagramme à bâtons des fréquences observées.

Pour $p = \frac{1}{2}$, dessinez le graphique de la densité de probabilité de la distribution normale $\mu = k p$ et $\sigma = \sqrt{k p (1 - p)}$.

Superposez les deux graphiques dans un même repère (Remarque : lorsque tous les écarts entre deux modalités successives sont égaux à 1, on peut interpréter le diagramme à bâtons comme un histogramme.)

■ Exercice 4.2 - 3 [Avec *Mathematica*]

a) Considérons le nombre moyen de six obtenus en lançant 100 fois le dé.

Cette variable aléatoire, désignée par M_{100} , est $M_{100} = \frac{1}{100} S$ où

S est la distribution binomiale $(100, \frac{1}{6})$.

Tracez l'histogramme de la distribution théorique. Comparez avec la distribution normale.

Remarque : pour la variable M_{100} , l'écart entre deux modalités successives est $\frac{1}{100}$, raison pour laquelle on ne peut pas simplement superposer le diagramme à bâtons et la fonction de densité. Le procédé d'interpolation contourne la difficulté.

b) Effectuons maintenant une simulation de M_{100} . On demande de

- construire un échantillon de taille n ;
- grouper les données en les traitant comme une variable continue;
- superposer l'histogramme et la distribution normale.

■ Exercice 4.2 - 4 [Avec *Mathematica*]

Pour illustrer graphiquement la notion d'écart-type, considérons deux variables aléatoires continues qui obéissent à des distributions normales. Imaginons deux populations (infinies) dont on mesure la masse de chaque individu. Les moyennes des deux populations sont égales à 60 kg. L'écart-type de la première est de 2 kg tandis que l'écart-type de la deuxième est de 10 kg.

a) Dans un même graphique, superposez les densités des deux populations.

- b) Effectuez un tirage aléatoire de 100 personnes de chaque population.

■ **Exercice 4.2 - 5** [Avec *Mathematica*]

- a) On donne un échantillon a dont les données ont été groupées:

Classes	Effectifs
0 - 0.5	113
0.5 - 1	277
1 - 1.5	278
1.5 - 2	191
2 - 2.5	97
2.5 - 3	34
3 - 3.5	10
3.5 - 4	0

Une distribution normale est-elle un modèle acceptable pour a ?

Indication : utilisez une méthode graphique.

- b) On donne un échantillon b dont les données ont été groupées:

Classes	Effectifs
0 - 8	8
8 - 16	48
16 - 24	172
24 - 32	362
32 - 40	272
40 - 48	115
48 - 56	21
56 - 64	1

Une distribution normale est-elle un modèle acceptable pour b ?

■ **Exercice 4.2 - 6** [Avec *Mathematica*]

Le diamètre intérieur moyen d'un échantillon de rondelles produites par une machine est de 1.275 cm et l'écart-type de 0.013 cm. L'usage que l'on en fait nécessite que le diamètre soit situé entre les bornes de tolérance de 1.260 à 1.290 cm sans quoi les rondelles sont considérées comme défectueuses.

Déterminez le pourcentage de rondelles défectueuses en supposant que les diamètres sont normalement distribués.

■ **Exercice 4.2 - 7** [Avec *Mathematica*]

On a lancé 500 fois une pièce de monnaie parfaitement équilibrée.

- a) Répondez à la question suivante en approximant la distribution binomiale au moyen d'une distribution normale: quelle est la probabilité que le nombre de faces soit situé entre 240 et 260 inclusivement ?
- b) Comparez avec l'exercice 3.2-3.

§ 4.3 Autres distributions

■ Exercice 4.3-1 [Avec *Mathematica*]

Une fabrique produit deux modèles d'appareils:

30 % de modèle A

70 % de modèle B.

Le modèle A a une durée de vie uniformément distribué sur l'intervalle [3 ans; 8 ans[tandis que le modèle B a une durée de vie uniformément distribué sur l'intervalle [2 ans; 6 ans[.

- Calculez l'espérance mathématique μ .
Représentez la fonction densité de la distribution théorique
- A partir d'un grand échantillon, calculez la moyenne et l'écart-type empiriques de la variable aléatoire "durée de vie d'un appareil".

Indications:

$\frac{3}{10}$ des appareils dans [3; 8[et $\frac{7}{10}$ dans [2; 6[équivaut à

$\frac{7}{10} \cdot \frac{1}{4}$ des appareils dans [2; 3[

$\frac{3}{10} \cdot \frac{3}{5} + \frac{7}{10} \cdot \frac{3}{4}$ des appareils dans [3; 6[

$\frac{3}{10} \cdot \frac{2}{5}$ des appareils dans [6, 8[

Il s'agit donc de définir une variable aléatoire qui est

uniformément distribuée dans [2, 3[avec une probabilité de $\frac{7}{40}$

uniformément distribuée dans [3, 6[avec une probabilité de $\frac{141}{200}$

uniformément distribuée dans [6, 8[avec une probabilité de $\frac{3}{25}$.

Pour choisir dans lequel des trois cas on tombe, on commence par tirer une variable aléatoire uniformément distribuée sur l'intervalle [0, 1[. Pour une question d'efficacité, on commence par traiter le cas le plus fréquent:

cas [3, 6] avec la prob. $\frac{141}{200} = 0.705$

cas [3, 6] ou [2, 3] avec la prob. $\frac{141}{200} + \frac{7}{40} = 0.88$

```

va := Module[{z},
  z = Random[UniformDistribution[{0, 1}]];
  Which[z < 0.705, Random[UniformDistribution[{3, 6}]],
    z < 0.88, Random[UniformDistribution[{2, 3}]],
    True, Random[UniformDistribution[{6, 8}]]];

  n = 40;

  x = Table[va[], {n}]

{5.46646, 2.03241, 4.98581, 3.49952, 4.85049, 2.60339, 3.33505, 4.09055,
 2.91219, 3.56597, 7.9615, 3.44513, 3.47579, 2.11596, 4.48998, 5.84837,
 7.75966, 4.03819, 5.30444, 7.96509, 5.56785, 5.38167, 5.62561, 4.93654,
 2.71661, 4.58865, 3.65984, 4.74028, 4.02035, 2.23403, 7.14394, 6.50297,
 5.14806, 4.37279, 2.84082, 5.43625, 6.24842, 6.56507, 5.7128, 2.70244}

```

§ 4.4 Estimation des paramètres

Dans ce paragraphe, nous nous demandons si les paramètres empiriques m et s permettent d'estimer les paramètres théoriques μ et σ .

Revenons à notre problème "Masses corporelles d'étudiants" du § 2.1. A en juger par son histogramme, il paraît

plausible que la distribution théorique soit normale. Pour passer d'une distribution empirique à la distribution théorique normale, il reste encore à estimer les paramètres μ et σ .

■ Estimation de l'espérance mathématique

Nous montrerons, dans *Statistique II*, § 2.5, qu'un bon estimateur de l'espérance mathématique est la moyenne empirique. Nous avons vu, dans le § 2, que l'on obtient des moyennes un peu différentes selon les groupements effectués. En prenant la moyenne des données brutes:

$$\hat{\mu} = m = \frac{1}{n} \sum_{i=1}^n x_i$$

```
x = {69, 59, 70, 72, 67, 49, 69, 67, 62, 65, 60, 68, 71, 75, 62, 77, 74, 59, 65, 62, 67,
74, 54, 63, 54, 69, 61, 72, 65, 54, 62, 71, 71, 74, 80, 61, 80, 67, 65, 65, 69,
69, 77, 62, 73, 61, 58, 77, 59, 73, 73, 66, 62, 57, 55, 61, 62, 67, 57, 55, 61,
67, 79, 69, 64, 70, 68, 68, 59, 67, 67, 64, 77, 73, 67, 57, 66, 68, 72, 78, 75,
62, 55, 64, 62, 71, 66, 67, 70, 68, 52, 77, 60, 65, 61, 57, 58, 70, 69, 66, 76,
69, 67, 63, 77, 77, 69, 72, 66, 60, 65, 62, 65, 61, 66, 72, 73, 76, 61, 63, 66,
64, 65, 61, 64, 61, 70, 66, 60, 65, 67, 56, 67, 66, 73, 70, 73, 73, 68, 64};
```

```
n = Length[x]
```

```
140
```

```
μ = Mean[x] ; N[μ]
```

```
66.2786
```

■ Exercice 4.4 - 1 [Sans ordinateur]

Démontrez que la valeur de t pour laquelle la fonction suivante est minimale

$$t \mapsto \frac{1}{n} \sum_{i=1}^n (x_i - t)^2$$

est la moyenne arithmétique

$$t = m = \frac{1}{n} \sum_{i=1}^n x_i$$

Indication : la fonction à étudier est du deuxième degré en t .

■ Estimation de l'écart-type théorique

L'écart-type empirique est calculée par rapport à la moyenne empirique

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m)^2}$$

Parmi tous les nombres m , c'est la moyenne arithmétique qui rend l'expression précédente minimale. Si maintenant on veut calculer l'écart type par rapport à l'espérance mathématique, on obtiendra un résultat plus grand

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$$

L'écart entre m et μ peut être important, particulièrement pour de petits échantillons. L'écart-type empirique s est généralement plus petit que l'écart-type théorique σ . On dit que s est un estimateur biaisé de σ .

Nous montrerons, dans *Statistique II*, § 2.5, qu'on obtient un estimateur non biaisé en remplaçant n par $(n - 1)$ dans la formule, c'est-à-dire

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2}$$

La correction est sensible pour de petits échantillons mais elle est négligeable pour les grands.

Il est nécessaire de calculer l'écart-type sur les mêmes données que celles utilisées pour calculer la moyenne (les deux paramètres avec les données brutes ou les deux paramètres avec les données groupées).

```
σ = StandardDeviation[x]; N[σ]
```

```
6.36093
```

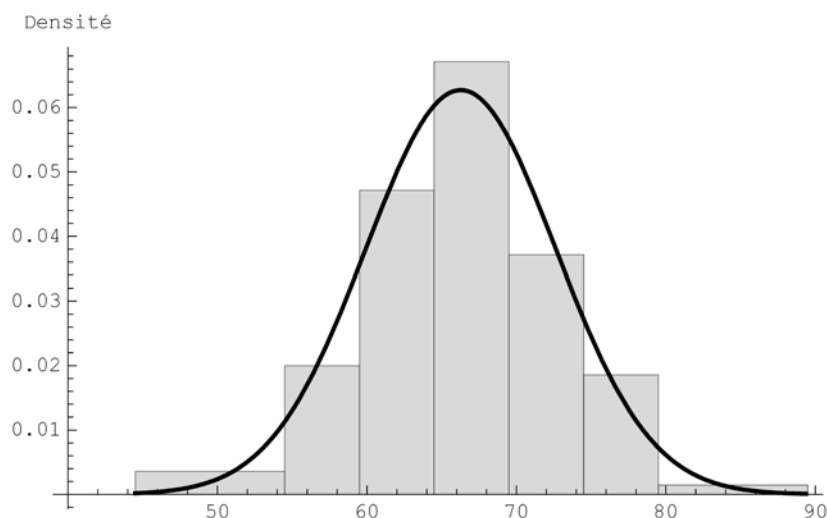
■ Modèle théorique

Pour l'exemple "Masses corporelles d'étudiants" du § 2.1, la distribution théorique est la suivante

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

Superposons les graphiques empiriques et théoriques

```
b = {44.5`, 54.5`, 59.5`, 64.5`, 69.5`, 74.5`, 79.5`, 89.5`};
effectifs = {5, 14, 33, 47, 26, 13, 2};
freq =  $\frac{\text{effectifs}}{n}$ ;
Show[histogramme[b, freq],
Plot[PDF[distr, t], {t, Min[b], Max[b]}, PlotStyle -> {Black, Thick}],
AxesLabel -> {None, "Densité"}, PlotRange -> All, AxesOrigin -> {40, 0}]
```



Puisque la superposition est excellente, l'hypothèse "la distribution est normale" est raisonnable. On dit alors que l'on a ajusté les données par une loi normale.

■ Exercice 4.4 - 2

On donne l'échantillon suivant:

$$x = \{95, 84, 76, 82, 87\}$$

- [Sans ordinateur] Calculez la moyenne et l'écart-type empiriques.
- [Avec Mathematica] Calculez la moyenne et l'écart-type empiriques.

- c) *[Sans ordinateur]* Estimez l'espérance $\hat{\mu}$ et l'écart-type théorique $\hat{\sigma}$.
- d) *[Avec Mathematica]* Estimez l'espérance $\hat{\mu}$ et l'écart-type théorique $\hat{\sigma}$.