Chapitre 7.
Loi à densité : la loi normale

Yann Barsamian

École Européenne de Bruxelles 1

Année scolaire 2020–2021

Uccle

Schola Europaea • Bruxelles I

Plan du chapitre

- Révisions sur la loi binomiale et les intégrales
- Une nouvelle notion : les probabilités continues
- La loi normale

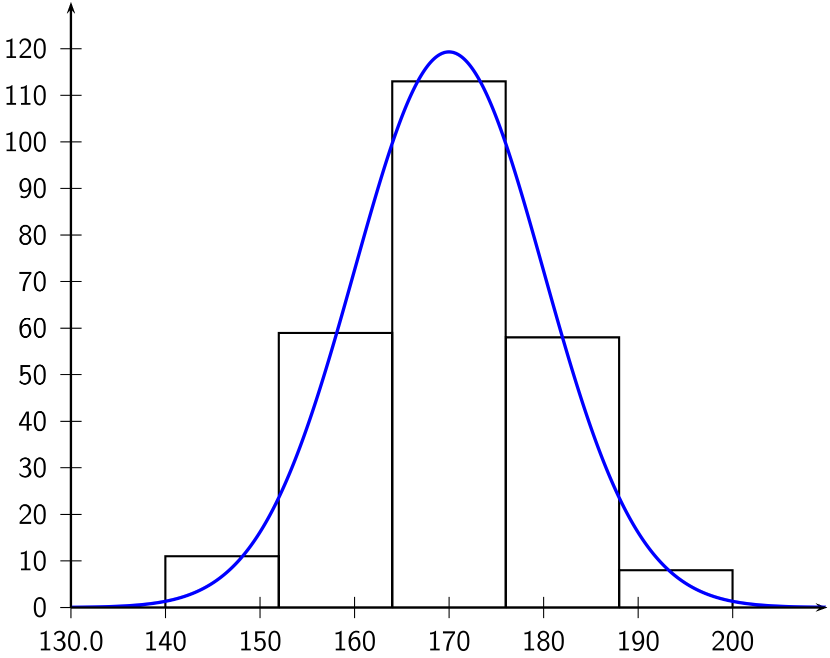
Introduction

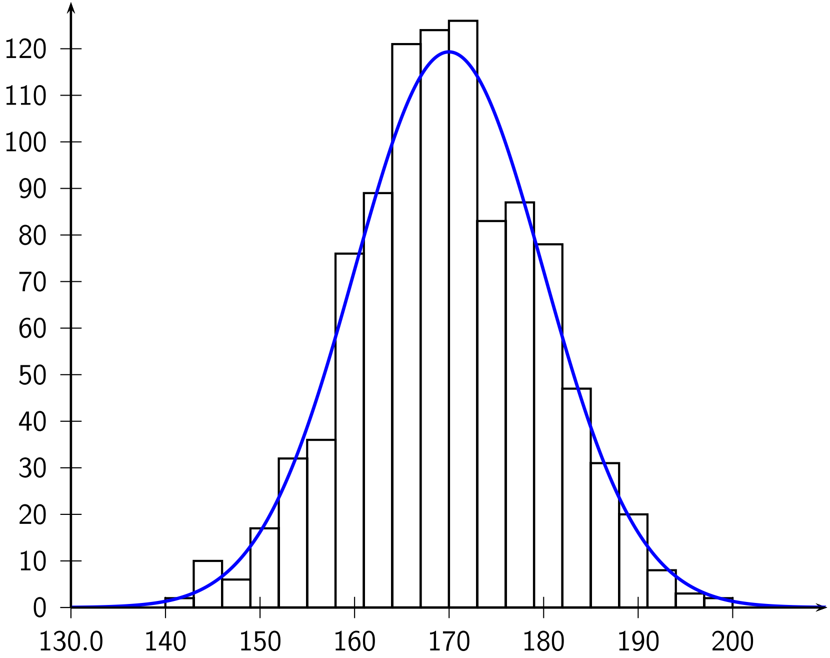
Lors du chapitre 5, on s'est intéressés à variables aléatoires discrètes : il n'y avait qu'un nombre fini de valeurs possibles.

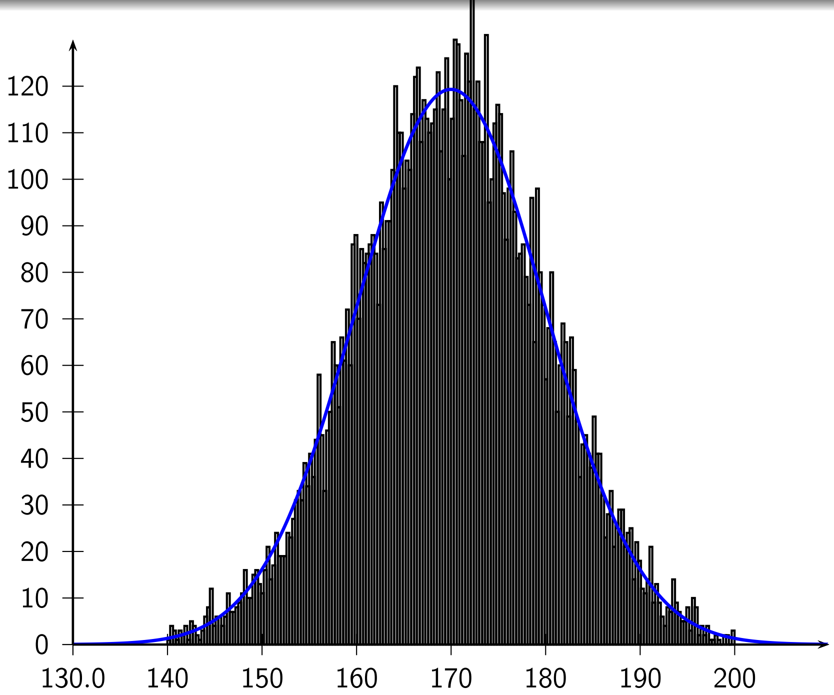
La loi binomiale de paramètres $n = 100$ et $p = 0,5$ modélise par exemple le nombre de "pile" que l'on peut faire sur 100 lancers : elle n'est définie que pour des valeurs entre 0 et 100.

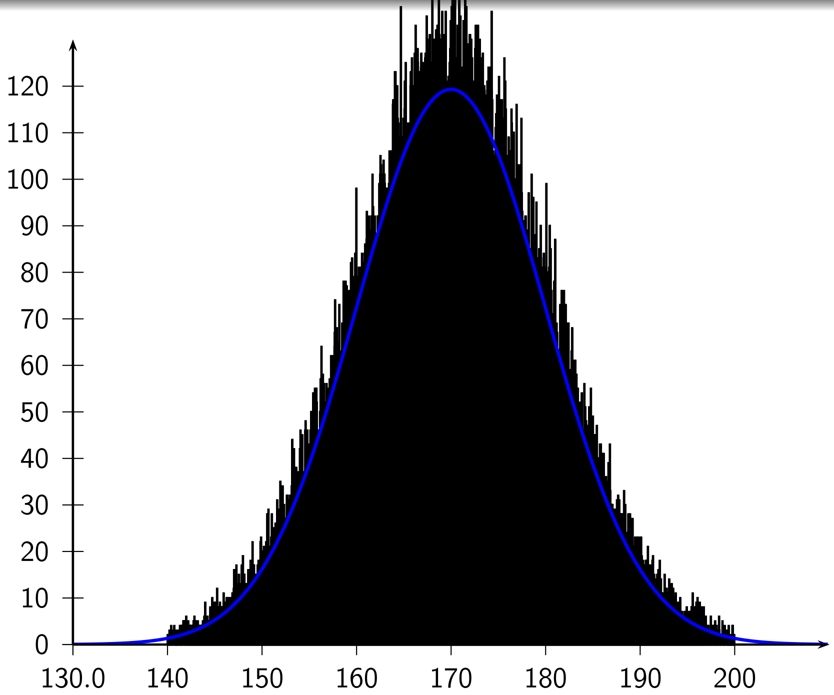
Dans ce chapitre, on va étudier des variables aléatoires continues : elles peuvent prendre n'importe quelle valeur dans un intervalle donné.

Notre premier exemple va être de regarder les tailles de différentes personnes. On va supposer que, sur notre échantillon de personnes, la moyenne est de 170 cm et l'écart-type est de 10 cm.

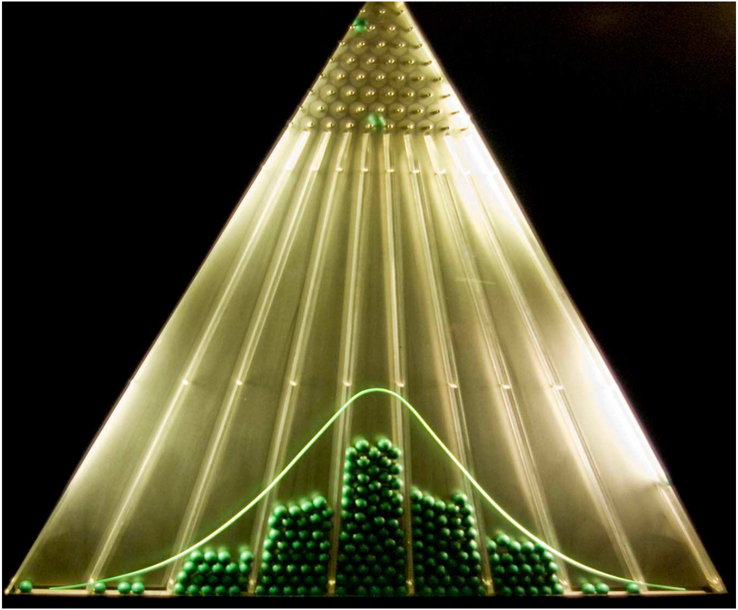

120
110
100
90
80
70
60
50
40
30
20
10
0
130.0
140
150
160
170
180
190
200


120
110
100
90
80
70
60
50
40
30
20
10
0
130.0
140
150
160
170
180
190
200


120
110
100
90
80
70
60
50
40
30
20
10
0
130.0
140
150
160
170
180
190
200


120
110
100
90
80
70
60
50
40
30
20
10
0
130.0
140
150
160
170
180
190
200

Un second exemple est la planche de Galton :



De très nombreux autres phénomènes sont modélisables par une loi normale (courbe de Gauss) :

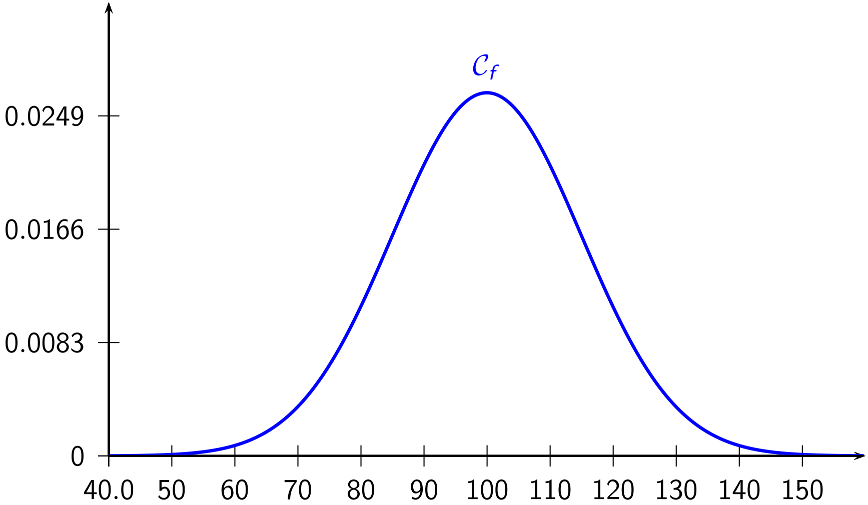
- courbe de quotient intellectuel (QI)
- les points d'impact de différents boulets tirés par un même canon
- la taille
- le bruit dans la transmission d'un signal électrique
- ...

Dans ce chapitre on s'intéresse à des événements aléatoires continus, et pas à des événements discrets. Cela veut dire qu'on ne regarde plus des probabilités du style $P(X = 160)$ mais des probabilités du style $P(140 \leq X \leq 160)$.

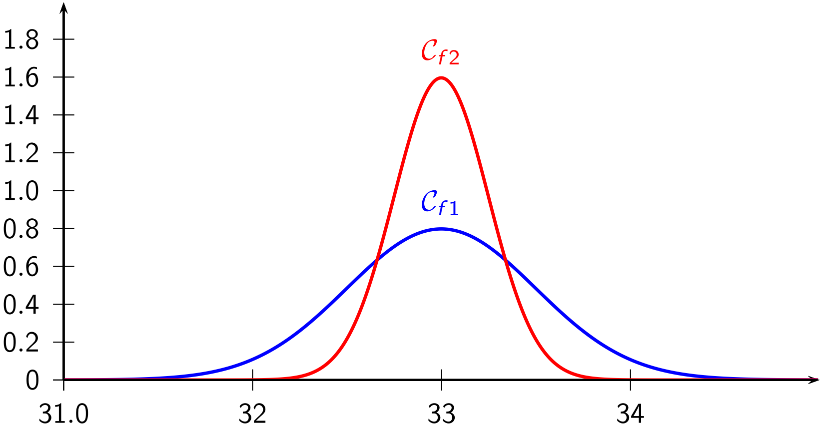
Remarque : cela a du sens de calculer $P(X = \text{quelque chose})$, mais ces probabilités valent 0 pour des variables aléatoires continues.

De nombreux phénomènes de la vie courante donnent des répartitions qui sont des courbes "en cloche". Ces courbes sont l'objet de notre cours.

Par exemple, si on regarde la répartition des QI dans la population, on observe la répartition suivante (moyenne 100, écart-type 15) :



Un second exemple est la fabrication de canette de 33 cL par une machine. On observera probablement des répartitions de ce type-là :

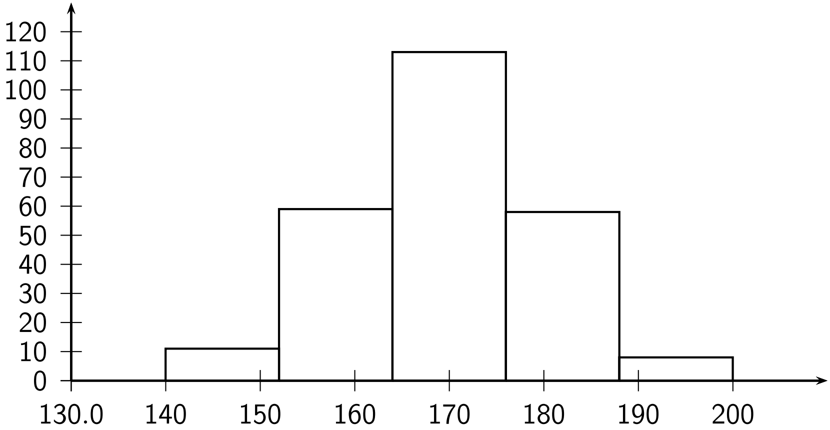


On voit que la machine 2 est plus précise que la machine 1 : l'écart-type 1 est plus grand que le 2e, puisque les valeurs sont plus étalées.

Tous les phénomènes aléatoires continus ne suivent pas une loi normale. S'il pleut et qu'on étudie la répartition des gouttes sur le terrain de sport de la cour, les gouttes vont se répartir de manière équitable sur toute la surface : il s'agit d'une loi <u>uniforme</u> (cela correspond à une situation d'équiprobabilité).

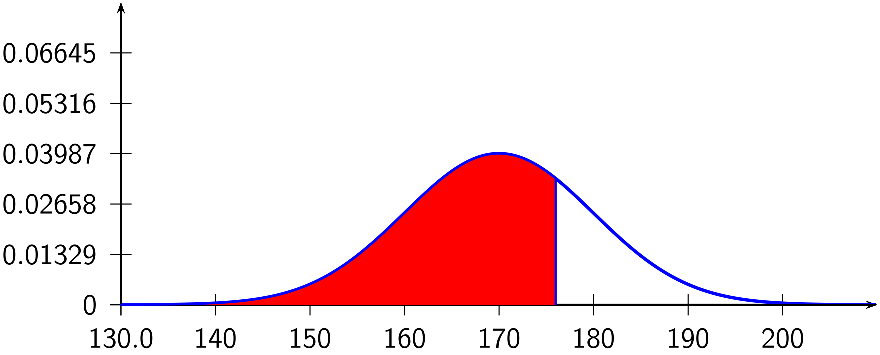
Le terrain de sport est composé de deux terrains de football de surface équivalente, donc la probabilité qu'une goutte de pluie tombe sur le terrain de football de gauche est de 50%.

Si on nous donne une répartition par classes, ici la taille de 250 personnes :



Pour calculer la probabilité qu'un individu mesure moins que 176 cm, on calcule : $\dfrac{11+59+113}{250} \approx 0,73$.

Avec une loi à densité, on n'a pas de rectangles et on se sert de l'aire sous la courbe. La fonction de répartition est définie sur $\mathbb{R}$ (donc de $-\infty$ à $+\infty$) et l'aire totale sous la courbe vaut $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$.



Par exemple, pour la loi normale de moyenne $\mu = 170$ et d'écart-type $\sigma = 10$, on calcule $P(X \leq 176) = \int_{-\infty}^{176} f(x)\,\mathrm{d}x$ (c'est l'aire rouge).

À la calculatrice : $\mathrm{normCdf}(-\infty, 176, 170, 10) \approx 0,73$.

La loi normale de moyenne μ est d'écart-type σ se note $\mathcal{N}\left(\mu\ ;\ \sigma^2\right)$.
Sa fonction de répartition a une allure de courbe en cloche, symétrique par rapport à la droite d'équation $x = \mu$.

- 68% des valeurs dans $[\mu - \sigma; \mu + \sigma]$ (en noir)
- $95,4\%$ des valeurs dans $[\mu - 2\sigma; \mu + 2\sigma]$ (noir + bleu)
- $99,7\%$ des valeurs dans $[\mu - 3\sigma; \mu + 3\sigma]$ (noir + bleu + rouge)