Marcel Délèze
Edition 2017

Supplément facultatif du § 2.1

Lien vers la page mère: Distribution empirique d'une variable continue
https://www.deleze.name/marcel/sec2/applmaths/csud/statistique_1/2-stat_I.pdf

Polygone de densité

La représentation de la densité par un histogramme n'est pas satisfaisante dans le cas où on désire arriver à une densité continue. Par exemple, lorsque la forme de la distribution théorique est inconnue, on peut s'intéresser à améliorer la distribution empirique.

Nous poursuivons l'exemple "*Masses corporelles d'étudiants*" sur les données déjà groupées (voir *Statistique I, § 2.1*).

```
Needs["Statistique`",
nécessite
  "C:\\Program Files\\Wolfram Research\\Mathematica\\perso\\Statistique.m"]
  constante C
```

```
b = {44.5, 54.5, 59.5, 64.5, 69.5, 74.5, 79.5, 89.5};
```

$$\text{freq} = \left\{\frac{1}{28}, \frac{1}{10}, \frac{33}{140}, \frac{47}{140}, \frac{13}{70}, \frac{13}{140}, \frac{1}{70}\right\};$$

Le polygone de densité

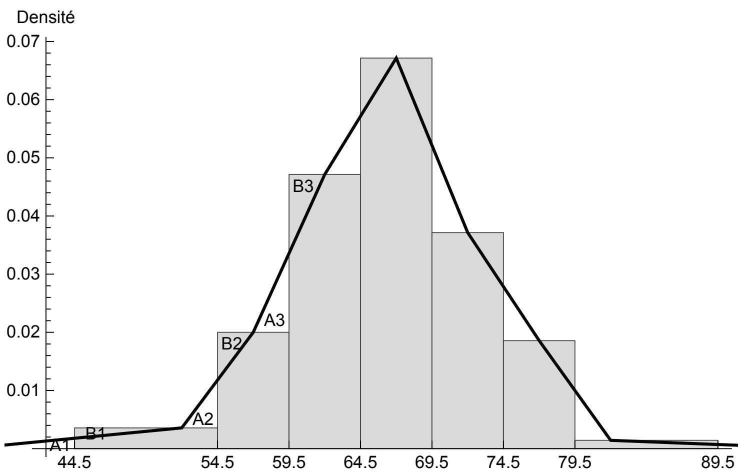
Dans le cas où les classes ont la même amplitude a, excepté peut-être les deux classes extrêmes,

$$b_2 - b_1 = b_3 - b_2 = \ldots = b_{k-1} - b_{k-2} = a$$

$$b_1 - b_0 > \frac{a}{2} \quad \text{et} \quad b_k - b_{k-1} > \frac{a}{2}$$

on peut aisément poursuivre la construction dans le but d'obtenir une densité continue.

L'histogramme est remplacé par une ligne polygonale de telle manière que les fréquences sur certains intervalles soient conservées.



$$\text{aire}(A1) = \text{aire}(B1), \quad \text{aire}(A2) = \text{aire}(B2), \quad \text{aire}(A3) = \text{aire}(B3), \quad \ldots$$

Pour réaliser cet objectif, on définit une liste de nombres w_0, ..., w_{k+1} définis comme suit. Des classes centrales de même largeur, on retient les centres de classes

$$w_2 = c_2 = \frac{b_1 + b_2}{2}, \quad \ldots, \quad w_{k-1} = c_{k-1} = \frac{b_{k-2} + b_{k-1}}{2}$$

Aux extrémités, on définit les nouvelles valeurs telles que $w_2 - b_1 = \frac{a}{2} = b_1 - w_1$, $w_1 - b_0 = b_0 - w_0$, ...

$$w_1 = b_1 - \frac{a}{2}$$

$$w_0 = 2\, b_0 - w_1$$

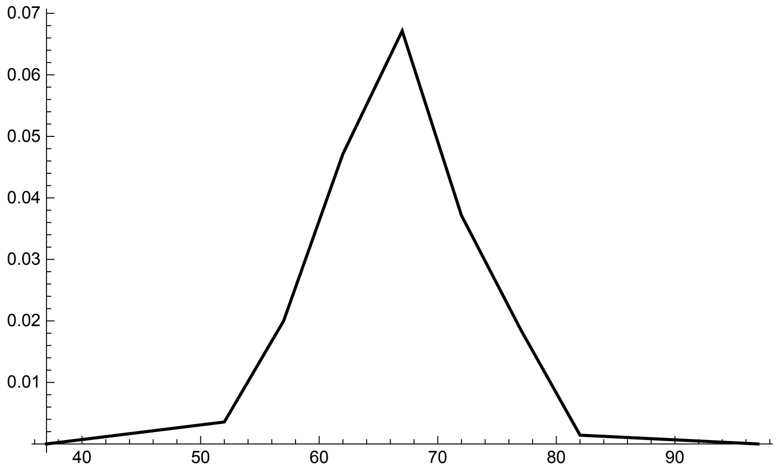
$$w_k = b_{k-1} + \frac{a}{2}$$

$$w_{k+1} = 2\, b_k - w_k$$

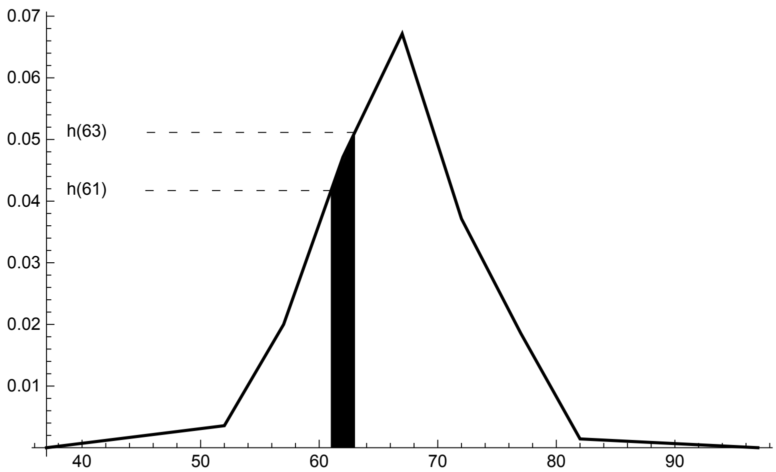
La fonction ainsi obtenue s'appelle le *polygone des densités*. Cette densité est *continue*.
L'aire de la surface délimitée par la densité vaut 1.

```
Show[polygoneDeDensite[b, freq]]
```

montre



Dans le but d'interpréter le graphique précédent, répondons à la question : "Quelle est la fréquence de l'événement *la masse appartient à l'intervalle [61; 63[*".



Conformément à la relation *fréquence = densité moyenne * amplitude*, on a

$$f\,[61;\ 63[\ = \frac{h(61) + h(63)}{2}\,(63\ \mathrm{kg} - 61\ \mathrm{kg})$$

Du point de vue géométrique, pour le trapèze marqué en noir,

*aire = hauteur moyenne * largeur*

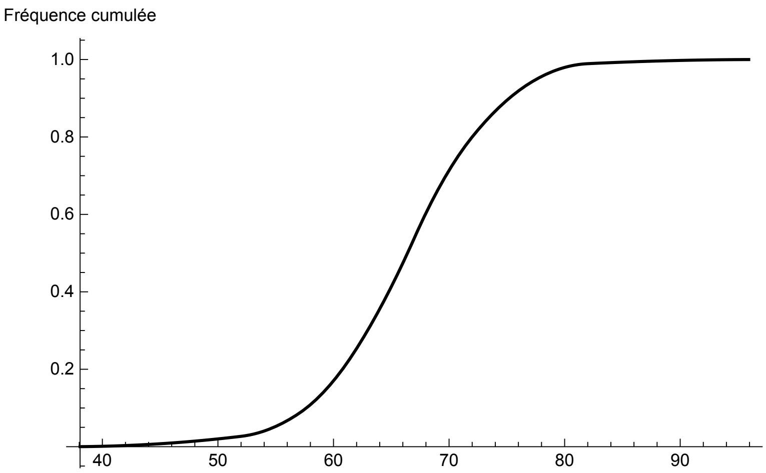
L'aire en noir représente donc la fréquence de l'événement *la masse appartient à l'intervalle [61; 63[*.

La fréquence cumulée lissée

La fonction de distribution qui correspond à la densité continue, notée $F_l : \mathbb{R} \longrightarrow \mathbb{R}$, est nommée *fréquence cumulée lissée*. Le calcul de F_l se ramène à la somme d'aires de trapèzes. La fonction F_l est un polynôme de degré 2 par morceaux.

```
frequenceCumuleeLissee[b, freq, AxesLabel → {None, "Fréquence cumulée"}]
```

titre d'axe aucun



Les paramètres empiriques de cette distribution peuvent différer un peu de ceux qui proviennent de l'histogramme (voir § 2.1); par exemple, pour la médiane

```
quantileLisse[b, freq, 1/2]
```

```
66.4043
```