



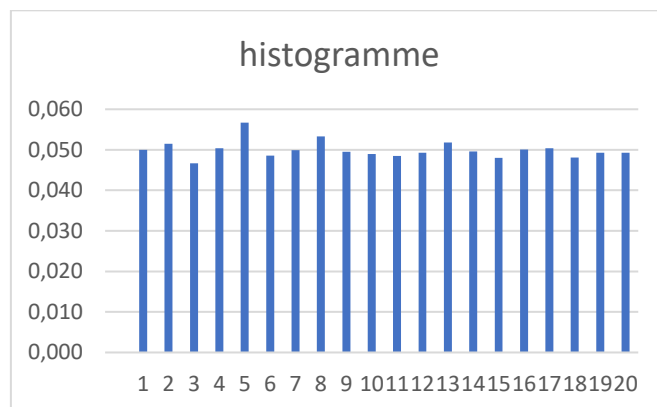
Présentation canonique d'un histogramme

par Bernard Beauzamy

Juillet 2022, rev. août 2026

1. Position du problème

Il est commode de constituer un histogramme à partir d'un ensemble de données. Un histogramme (appelé aussi "diagramme en bâtons") est fait d'un certain nombre de classes consécutives, de même taille ; on compte combien de données tombent dans chacune des classes. A priori, la largeur des classes et le point de départ de la première sont à la disposition de l'utilisateur, mais certains choix sont plus pratiques que d'autres, et favorisent la compréhension. En particulier, nous sommes habitués au système décimal, et nous préférons les classes qui s'appuient sur ce système. En outre, le nombre de classes ne doit pas être trop élevé : 20 est un maximum. Voici un exemple, avec 20 classes :



Nous conviendrons que les classes sont toutes de la forme $a \leq x < b$ (fermé à gauche, ouvert à droite), sauf la dernière qui est de la forme $a \leq x \leq b$ (fermé à gauche et à droite). Enfin, évidemment, les classes sont consécutives et disjointes. Leur réunion recouvre toutes les valeurs recueillies.

2. Notations et choix

Notons respectivement m, M la plus petite et la plus grande des valeurs présentes parmi les données. La taille des classes, notée w , est déterminée par la précision attendue dans la présentation des résultats : elle dépend donc de considérations extérieures aux données elles-mêmes. Selon les cas, on peut souhaiter une précision de l'ordre de l'année ; dans d'autres cas, de l'ordre de la milliseconde.

Il sera commode, pour la compréhension, de choisir des classes de largeur $w = 10^\alpha$, où α peut être positif ou négatif : nous sommes habitués à la numération décimale. Dire que les classes ont largeur 100 ou 1/100 est plus parlant que dire qu'elles ont largeur 0.1234.

Nous pouvons dire que la taille des classes s'impose de l'extérieur. Nous avons à choisir A , début de la première classe ; il doit satisfaire :

$$A \leq m < A + w.$$

Les classes seront :

$$[A, A + w[, [A + w, A + 2w[, \dots, [A + (k - 2)w, A + (k - 1)w[, [A + (k - 1)w, A + kw]$$

avec bien sûr $A + (k - 1)w \leq M \leq A + kw$ (la dernière classe contient M , la plus grande valeur).

Si le début A de la première classe est fixé, le nombre de classes k s'en déduit, puisque w est connu. En effet, des inégalités $A + (k - 1)w \leq M \leq A + kw$ on déduit :

$$k - 1 \leq \frac{M - A}{w} \leq k,$$

et donc :

- Si $\frac{M - A}{w}$ est entier, $k = \frac{M - A}{w}$;
- Si $\frac{M - A}{w}$ n'est pas entier, $k = \text{int}\left(\frac{M - A}{w}\right) + 1$, où $\text{int}(x)$, noté aussi $[x]$, désigne le plus grand entier inférieur ou égal à x .

Reste à déterminer A , début de la première classe. En théorie, on peut prendre n'importe quel A vérifiant $A \leq m < A + w$. En pratique, certains choix sont plus simples et plus faciles à retenir. Si $m = 0.1234$ et $w = 1$, on choisira naturellement $A = 0$.

De manière générale, on choisira A , approximation décimale la plus proche de m . Cela conduit au choix :

$$A = \text{int}\left(\frac{m}{w}\right) \times w.$$

Si par exemple $m = 2.451$ et $w = 0.05$, $\frac{m}{w} = 49.02$, $\text{int}\left(\frac{m}{w}\right) = 49$ et $A = 2.45$.

On dispose ainsi d'une méthode "canonique" pour présenter un histogramme :

On retient m, M , plus petite et plus grande valeur recueillies ;

On détermine w , taille des classes, en fonction de la précision souhaitée (et non, en principe, en fonction du nombre de données) ;

On détermine A , début de la première classe, par la formule $A = \text{int}\left(\frac{m}{w}\right) \times w$, où $\text{int}(x)$ est le plus grand entier inférieur ou égal à x ;

On détermine le nombre de classes, k , par les formules :

Si $\frac{M - A}{w}$ est entier, $k = \frac{M - A}{w}$; si $\frac{M - A}{w}$ n'est pas entier, $k = \text{int}\left(\frac{M - A}{w}\right) + 1$.