Calcul du nombre de classes dans un histogramme

par Dorsaf Mathlouthi

octobre 2022

Résumé opérationnel : nous donnons des estimations pour le nombre de classes qui seront utilisées pour construire un histogramme. L'intérêt pratique est double : aide à la programmation, contrôle a posteriori.

Notations

Soient x_n, $n=1,\dots,N$, des nombres réels à partir desquels on veut construire un histogramme. Nous notons $m=\min(x_n)$, $M=\max(x_n)$; soient en outre k, le nombre de classes, w leur taille (la même pour toutes), α le début de la première classe. Les classes seront donc de la forme $[\alpha,\alpha+w[,[\alpha+w,\alpha+2w[,\dots$ Par convention, chaque classe est fermée à gauche et ouverte à droite.

L'objet du présent travail est la détermination de k à partir de m, M, w.

Proposition 1. – Le nombre de classes est invariant par translation et par homothétie de rapport $\lambda>0$.

Ceci est évident : si on ajoute un même nombre à tous les x_n et à toutes les bornes de classes, le nombre de classes est inchangé. De même si tous les x_n et toutes les bornes de classes sont multipliés par un même $\lambda>0$.

Il en résulte que, pour déterminer le nombre de classes k, on peut toujours supposer $m=0$ (ce qui revient à dire que tous les x_n sont ≥ 0 et que $w=1$).

Proposition 2. – Avec les notations précédentes, on a :

$$k=\left[\frac{M-\alpha}{w}\right]+1$$

où $[\ ]$ désigne la partie entière.

Démonstration de la Proposition 2

Par définition, la dernière classe contient M, ce qui se traduit par :

$\alpha+(k-1)w \le M < \alpha+kw$.

et donc :

$k-1 \le \dfrac{M-\alpha}{w} < k$

ce qui prouve notre Proposition.

L'utilisateur a à sa disposition la suite des x_n et il choisit w, taille des classes. Il aimerait disposer d'une formule déterminant le nombre de classes k, à partir de M, m, w. Comme nous allons le voir, deux valeurs sont possibles et deux seulement.

Commençons par un exemple : $m=0, w=1, k=3, \alpha=-0.5$. Les classes sont : [-0.5, 0.5[, [0.5,1.5[, [1.5, 2.5[.

Pour $M \ge 2$, on aura : $\left[\dfrac{M-m}{w}\right]=2=k-1$

Pour $1 \le M < 2$, on aura : $\left[\dfrac{M-m}{w}\right]=1=k-2$.

Proposition 3. – Avec les notations précédentes, on a :

$$k=\left[\frac{M-m}{w}\right]+1 \text{ ou } k=\left[\frac{M-m}{w}\right]+2.$$

Démonstration de la Proposition 3

Par définition de la première et de la dernière classe, on a :

$\alpha \le m < \alpha+w$; $\alpha+(k-1)w \le M < \alpha+kw$

Puisque l'ensemble des classes contient $[m, M]$:

$M-m \le kw$

Il n'est pas possible d'avoir égalité, car la dernière classe ne peut avoir M pour borne de droite, puisqu'elle est ouverte à droite. Donc :

$M - m < kw$

Si l'on retire la première et la dernière classe, ce qui reste est contenu dans le segment $[m, M]$ et par conséquent :

$$(k-2)w \leq M - m$$

Il n'est pas possible d'avoir égalité, car la première classe ne peut avoir m pour borne de droite, puisqu'elle est ouverte à droite. Donc :

$$(k-2)w < M - m$$

En résumé :

$$k - 2 < \frac{M-m}{w} < k$$

On en déduit, pour la partie entière :

$$k - 2 \leq \left[\frac{M-m}{w}\right] < k$$

ce qui prouve la Proposition.

La valeur exacte de k peut être déterminée en utilisant la valeur de $\alpha,$ début de la première classe, comme nous l'avons vu à la Proposition 2 :

$$k = \left[\frac{M-\alpha}{w}\right] + 1.$$