

Variance d'une distribution de données

Hugues GENVRIN

December 2025

Soit une série de données $(x_i)_{1 \leq i \leq n}$. On note $\bar{x}$ la moyenne arithmétique, $\bar{x}_q = \sum_{i=1}^n \frac{x_i^2}{n}$ la moyenne quadratique, et on appelle $V_n = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$ la variance.

Montrons que la variance ne converge pas systématiquement vers zéro.

Démonstration. Montrons tout d'abord que (V_n) converge.

Soit $V_{n+1} = \sum_{i=1}^{n+1} \frac{(x_i - \bar{x})^2}{n+1} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n+1} \times \frac{n}{n+1} + \frac{(x_{n+1} - \bar{x})^2}{n+1} = V_n \times \frac{1}{1 + \frac{1}{n}} + \frac{(x_{n+1} - \bar{x})^2}{n+1}$.

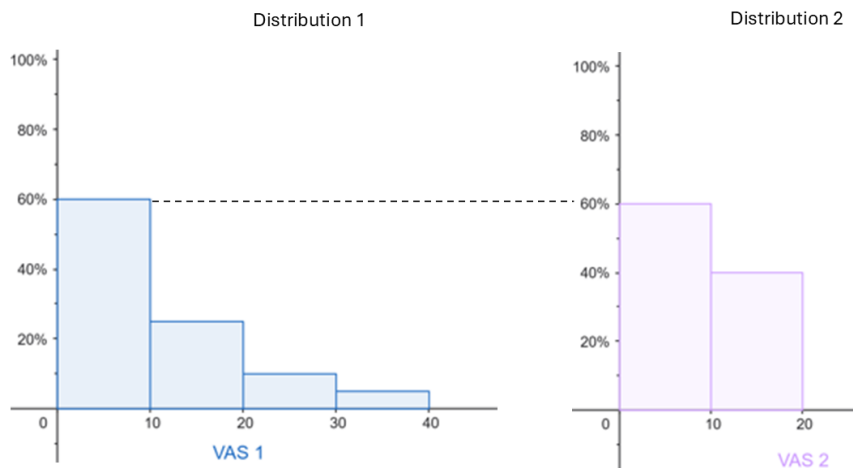
Comme $(x_i)_{1 \leq i \leq n}$ est bornée, lorsque n tend vers l'infini, l'expression de la variance $V_{n+1} \rightarrow V_n \Leftrightarrow V_{n+1} - V_n \rightarrow 0$. Donc, $(V_i)_{1 \leq i \leq n}$ est convergente vers une valeur réelle positive. En conséquence l'écart-type $(\sigma_i)_{1 \leq i \leq n} = \sqrt{V_i}$ est convergent.

Montrons que l'expression de $(V_i)_{1 \leq i \leq n}$ ne converge pas systématiquement vers zéro. Soit $V_n = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_{i=1}^n \frac{x_i^2}{n} - 2\bar{x} \sum_{i=1}^n \frac{x_i}{n} + \bar{x}^2 = \bar{x}_q - 2\bar{x} \bar{x} + \bar{x}^2 = \bar{x}_q - \bar{x}^2$. Donc, une condition nécessaire pour que la variance converge vers zéro est que la moyenne quadratique vaut la moyenne arithmétique au carré.

Considérons un nombre pair de données. On suppose que $x_i = Sup$ pour la moitié des données et $x_i = Inf$ pour l'autre moitié (pour maximiser la variance). Alors $\bar{x} = \frac{1}{2}(Sup + Inf)$, tandis que $\sqrt{\bar{x}_q} = \sqrt{\frac{n(Sup^2 + Inf^2)}{2n}}$.

En élevant au carré de part et d'autre, et en simplifiant les expressions. Il vient $\frac{1}{4}Sup^2 + \frac{1}{4}Inf^2 + \frac{1}{2} \times Sup \times Inf = \frac{1}{2}Sup^2 + \frac{1}{2}Inf^2 \Leftrightarrow Sup \times Inf = \frac{1}{4}Sup^2 + \frac{1}{4}Inf^2$, ce qui est manifestement faux pour une multitude de cas. On conclut que la variance et donc l'écart-type ne convergent pas systématiquement vers zéro. $\square$

Interprétons ce contre exemple qui nous servira de cas de base. Montrons que la variance est systématiquement finie dans le cas de la chirurgie. Interprétons le cas d'une distribution où des effectifs avec $\frac{N}{2}Inf$ ou $\frac{N}{2}Sup$ (Où N définit le nombre de patients inclus) n'est pas exceptionnel. On pense au cas d'une évaluation post-opératoire avec un recul de 3 ou 12 mois. La distribution des résultats peut témoigner d'un mode à 0, renchérit par un effet plancher. Prenons le cas où il y aurait 60 % des effectifs qui n'auraient plus de douleur post-opératoire. Alors c'est légèrement supérieur à $\frac{N}{2}Inf$. Par ailleurs, l'amplitude de la série ne sera pas 100, car on suppose qu'aucun patient n'est 100 en post opératoire. Si bien que Sup est proche de Inf en pratique.



La variance de la série 1 est supérieure à la variance de la série 2, car la dispersion par rapport à la moyenne est plus élevée dans la série 1 que dans la série 2, où l'effectif total est constant.

On observera que la fréquence de la première classe des deux distributions est constante .

FIGURE 1 – Exemple d'une distribution du cas de base.