La Variance

Jean-François Berger-Lefébure

Contents

1 Pourquoi la formule de la variance est-elle définie ainsi ?

1.1 Introduction

La formule de la variance :

$$Var(Y) = E[(Y - E(Y))^2]$$

a été conçue pour mesurer la **dispersion** ou la **variabilité** d'une variable aléatoire autour de sa moyenne.
—

1.2 Pourquoi mesurer la dispersion ?

Connaître uniquement la moyenne $E(Y)$ d'une variable aléatoire ne suffit pas pour décrire sa distribution. Deux ensembles de données peuvent avoir la même moyenne mais des dispersions très différentes :

- Exemple 1 : $Y = [8, 9, 10, 11, 12]$
- Exemple 2 : $Y = [2, 5, 10, 15, 18]$

Ces deux ensembles ont une moyenne de 10, mais leurs dispersions sont clairement différentes. Nous avons donc besoin d'une mesure pour capturer **l'éloignement des valeurs par rapport à la moyenne**.
—

1.3 Pourquoi utiliser l'écart à la moyenne ?

Pour mesurer la dispersion, on calcule d'abord l'écart entre chaque valeur et la moyenne :

$$D_i = Y_i - E(Y)$$

Mais la somme de ces écarts est toujours **nulle** :

$$\sum(Y_i - E(Y)) = 0$$

Cela empêche de mesurer correctement la dispersion.
—

1.4 Pourquoi élever au carré ?

Pour éviter l'annulation des écarts positifs et négatifs, on **élève au carré** :

$$(Y - E(Y))^2$$

- Les valeurs négatives deviennent positives.
- Les écarts extrêmes ont un poids plus élevé, car $4^2 = 16$ pèse plus que $2^2 = 4$.
- Ce choix correspond aussi à des mesures physiques comme la distance euclidienne.

—

1.5 Pourquoi prendre l'espérance ?

Pour généraliser à l'ensemble des valeurs possibles, on calcule la **moyenne des carrés des écarts** :

$$E[(Y - E(Y))^2]$$

Cela permet de :

- Avoir une **mesure moyenne de la dispersion** sur toute la distribution de la variable.
- Obtenir une mesure **robuste et cohérente** pour comparer des distributions différentes.

—

1.6 Fondement mathématique

La formule satisfait trois propriétés essentielles :

- **Non-négativité** : $Var(Y) \geq 0$.
- **Invariance par translation** : Ajouter une constante n'affecte pas la variance.
- **Homogénéité par changement d'échelle** : Multiplier Y par une constante a multiplie la variance par a^2.

—

1.7 Résumé

La formule de la variance a été définie pour :

- Quantifier la **dispersion réelle** des données.
- Corriger l'annulation des écarts positifs et négatifs.
- Donner plus d'importance aux valeurs éloignées en les élevant au carré.
- Fournir une mesure moyenne généralisée sur toute la distribution.

Ainsi, la définition formelle :

$$Var(Y) = E[(Y - E(Y))^2]$$

est un outil fondamental pour étudier la variabilité des données et la manière dont elles se répartissent autour de leur moyenne.

—

2 Démonstration compléte

2.1 Définition de la variance

La variance mesure la dispersion des valeurs d'une variable aléatoire autour de sa moyenne. La formule source est :

$$Var(Y) = E[(Y - E(Y))^2].$$

Explications :

- $E(Y)$: Espérance (ou moyenne) de Y.
- $(Y - E(Y))$: Écart entre Y et sa moyenne.
- $(Y - E(Y))^2$: Carré de l'écart $\rightarrow$ Mesure l'intensité de l'écart (pas de signe négatif).
- $E[\ldots]$: Moyenne de tous ces carrés $\rightarrow$ Mesure la dispersion globale des valeurs autour de la moyenne.

—

2.2 Problème posé : Transformation linéaire avec une constante

Supposons que :

$$Y = aX,$$

où :

- a est une constante.
- X est une variable aléatoire avec : $E(X) = \mu_X$ et $Var(X) = \sigma_X^2$.

On veut prouver que :

$$Var(Y) = a^2 Var(X).$$

2.3 Calcul détaillé de l'espérance

Pour prouver cette propriété, calculons d'abord l'espérance de Y :

$$E(Y) = E(aX).$$

Propriété de l'espérance :

$$E(aX) = aE(X).$$

Donc :

$$E(Y) = a\mu_X.$$

—

2.4 Calcul détaillé de la variance

Par définition :

$$Var(Y) = E[(Y - E(Y))^2].$$

Substitution de $Y = aX$ et $E(Y) = aE(X)$:

$$Var(Y) = E[(aX - aE(X))^2].$$

Facteur a :

$$Var(Y) = E[(a(X - E(X)))^2].$$

Développement du carré :

$$Var(Y) = E[a^2(X - E(X))^2].$$

Constante au carré (a^2) : L'espérance d'une constante multipliée est :

$$E(cZ) = cE(Z).$$

On peut donc sortir a^2 :

$$Var(Y) = a^2 E[(X - E(X))^2].$$

Par définition de la variance de X :

$$Var(X) = E[(X - E(X))^2].$$

Substitution finale :

$$Var(Y) = a^2 Var(X).$$

—

2.5 Intuition simple : Pourquoi a^2 ?

- **Écart multiplié par a :** Si la variable initiale varie de ± 1, la nouvelle variable varie de $\pm a$.
- **Carré des écarts :** En passant au carré, on amplifie l'effet :

$$(a \cdot 1)^2 = a^2.$$

- **Variance = dispersion quadratique moyenne :** Comme la variance mesure une aire (écart quadratique moyen), toute transformation par une constante se répercute au carré.

—

2.6 Exemple numérique pour illustrer

Données :

- X suit une loi normale avec : $E(X) = 10$ et $Var(X) = 4$.
- Transformons $Y = 3X$.

Calculs :

- Espérance de Y :

$$E(Y) = 3E(X) = 3(10) = 30.$$

- Variance de Y :

$$Var(Y) = 3^2 Var(X) = 9(4) = 36.$$

Conclusion : La variance est multipliée par le carré de la constante ($a^2 = 9$). Cela reflète le fait que les écarts à la moyenne sont également multipliés par a, donc leur dispersion au carré est amplifiée par a^2.
—

2.7 Conclusion

La formule :

$$Var(aX) = a^2 Var(X)$$

provient directement des propriétés :

- L'espérance se multiplie linéairement par la constante a.
- La variance mesure des écarts quadratiques $\rightarrow$ Toute constante a appliquée sur les valeurs est au carré dans la variance.