

Probabilités et échantillonnage

Cours de maths - 2de

Pile ou face, loterie, météo, résultats d'un sondage : le hasard est partout. Les **probabilités** permettent de mesurer les chances qu'un événement se produise, et l'**échantillonnage** d'expliquer pourquoi, sur 100 lancers, on n'obtient presque jamais exactement 50 piles. Dans ce chapitre, tu apprends à décrire une expérience aléatoire, à calculer des probabilités avec des tableaux et des arbres, puis à décider si une fréquence observée est « normale » ou surprenante.

1. Expérience aléatoire, issues et univers

Expérience aléatoire Une expérience est **aléatoire** quand on connaît tous ses résultats possibles, mais qu'on ne peut pas prévoir lequel va se produire. Chaque résultat possible s'appelle une **issue**. L'ensemble de toutes les issues est l'**univers**, noté Ω .

Exemple On lance un dé équilibré à 8 faces numérotées de 1 à 8. Les issues sont les nombres 1, 2, 3, 4, 5, 6, 7, 8. L'univers est $\Omega = \{1 ; 2 ; 3 ; 4 ; 5 ; 6 ; 7 ; 8\}$. Il contient 8 issues.

Pour décrire correctement un univers, il faut être précis. Si on lance **deux fois** une pièce, les issues sont des *listes* : PP, PF, FP, FF (P = pile, F = face). L'ordre compte : PF et FP sont deux issues différentes.

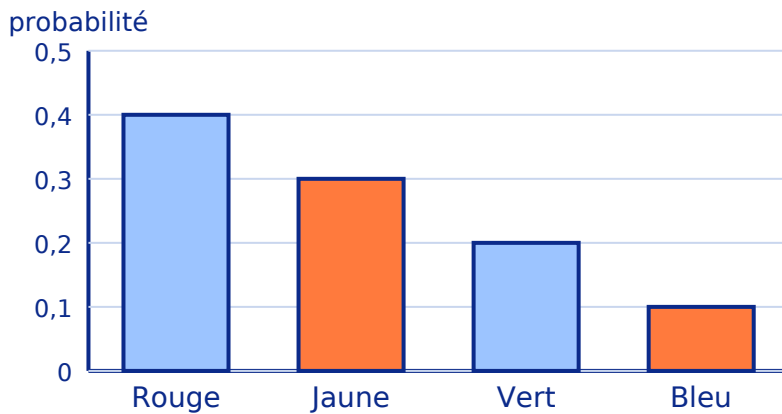
Piège Pour deux lancers de pièce, il y a **4** issues et non 3. « Un pile et un face » n'est pas une issue : c'est un *événement* qui regroupe PF et FP.

2. Loi de probabilité

Loi de probabilité Définir une loi de probabilité sur un univers $\Omega = \{x_1 ; x_2 ; \dots ; x_n\}$, c'est associer à chaque issue x_i un nombre p_i compris entre 0 et 1, de sorte que **$p_1 + p_2 + \dots + p_n = 1$** .

On présente souvent une loi dans un tableau ou dans un diagramme en barres. Voici une roue de loterie à quatre couleurs :

Couleur	Rouge	Jaune	Vert	Bleu	Total
Probabilité	0,4	0,3	0,2	0,1	1



Les quatre probabilités sont entre 0 et 1, et leur somme vaut $0,4 + 0,3 + 0,2 + 0,1 = 1$: c'est bien une loi de probabilité. Ici les issues **n'ont pas** la même probabilité.

Équiprobabilité Quand toutes les issues ont la **même probabilité**, on dit qu'il y a **équiprobabilité** (dé équilibré, pièce équilibrée, tirage « au hasard » dans un sac). Avec n issues, chacune a pour probabilité **$1/n$** .

Exemple Dé équilibré à 8 faces : chaque face a pour probabilité **$1/8$** . Somme : $8 \times 1/8 = 1$.

Ne confonds pas deux idées : la probabilité est une valeur *théorique* choisie par le modèle ; la **fréquence** est une valeur *observée* après avoir répété l'expérience. Quand on répète très souvent l'expérience, les fréquences se rapprochent des probabilités.

3. Événements et probabilité d'un événement

Événement Un **événement** est une partie de l'univers, c'est-à-dire un groupe d'issues. Un événement qui contient une seule issue est dit **élémentaire**. L'événement qui contient toutes les issues est **certain** (probabilité 1) ; celui qui n'en contient aucune est **impossible** (probabilité 0).

Probabilité d'un événement La probabilité d'un événement A , notée $p(A)$, est la **somme des probabilités des issues** qui le réalisent. Toujours $0 \leq p(A) \leq 1$. En situation d'équiprobabilité :

$p(A) = \text{nombre d'issues favorables à } A \div \text{nombre total d'issues}$.

Exemple 1 (équiprobabilité) Dé à 8 faces, A : « obtenir un nombre premier ». Les issues favorables sont 2, 3, 5, 7 : il y en a 4 sur 8. Donc $p(A) = 4/8 = 1/2$.

Exemple 2 (roue non équilibrée) Avec la roue du paragraphe 2, E : « obtenir rouge ou vert ». $p(E) = p(\text{rouge}) + p(\text{vert}) = 0,4 + 0,2 = 0,6$.

Méthode : calculer la probabilité d'un événement

1. Écris l'univers (liste des issues).
2. Vérifie s'il y a équiprobabilité.
3. Liste les issues qui réalisent l'événement.
4. Divise leur nombre par le nombre total (ou additionne leurs probabilités).
5. Simplifie la fraction et conclus par une phrase.

4. Événement contraire

Événement contraire Le **contraire** d'un événement A, noté $\bar{A}$, est l'événement formé de toutes les issues qui **ne réalisent pas** A. A et $\bar{A}$ n'ont aucune issue en commun, et à elles deux elles couvrent tout l'univers.

Propriété $p(\bar{A}) = 1 - p(A)$.

Exemple Dans un sac de 20 jetons, 6 sont rouges. A : « tirer un jeton rouge », $p(A) = \frac{6}{20} = 0,3$. L'événement $\bar{A}$: « ne pas tirer un jeton rouge » a pour probabilité $1 - 0,3 = 0,7$.

Astuce « au moins un » Quand une question contient « au moins un », passe par le contraire : « au moins un » a pour contraire « aucun », souvent bien plus simple à calculer. Exemple : en lançant 3 fois une pièce, $p(\text{au moins un pile}) = 1 - p(\text{aucun pile}) = 1 - \frac{1}{8} = \frac{7}{8}$.

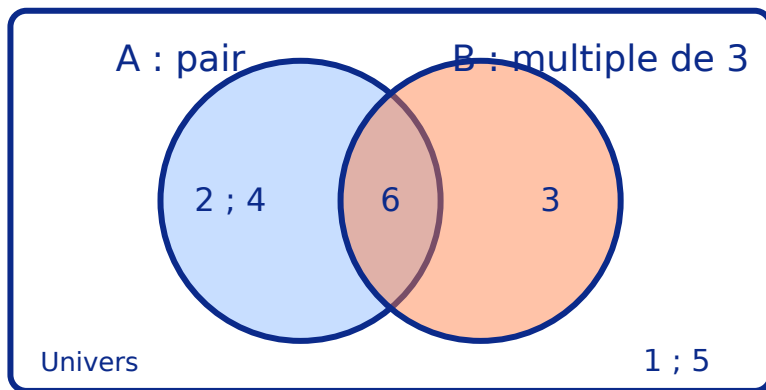
5. Intersection et réunion de deux événements

Intersection et réunion Soient A et B deux événements.

- **$A \cap B$** (« A inter B », « A et B ») : les issues qui réalisent **à la fois** A et B.
- **$A \cup B$** (« A union B », « A ou B ») : les issues qui réalisent **A, ou B, ou les deux**.
Le « ou » est inclusif en mathématiques.

Si $A \cap B$ est vide, on dit que A et B sont **incompatibles**.

Exemple Dé équilibré à 6 faces. A : « pair » = {2 ; 4 ; 6}. B : « multiple de 3 » = {3 ; 6}. Alors $A \cap B = \{6\}$ et $A \cup B = \{2 ; 3 ; 4 ; 6\}$. Donc $p(A \cap B) = \frac{1}{6}$ et $p(A \cup B) = \frac{4}{6} = \frac{2}{3}$.



Formule de la réunion Pour deux événements A et B quelconques :

$$p(A \cup B) = p(A) + p(B) - p(A \cap B).$$

Si A et B sont incompatibles : $p(A \cup B) = p(A) + p(B)$.

Démonstration. Quand on additionne $p(A)$ et $p(B)$, les issues de $A \cap B$ sont comptées deux fois (une fois dans A, une fois dans B). Il faut donc les retirer une fois : $p(A) + p(B) - p(A \cap B)$.

Vérification sur l'exemple : $3/6 + 2/6 - 1/6 = 4/6$. On retrouve bien $p(A \cup B)$.

Piège Additionner simplement $p(A) + p(B)$ donne un résultat faux dès que A et B ont des issues communes (on peut même dépasser 1 !).

6. Tableaux à double entrée et arbres

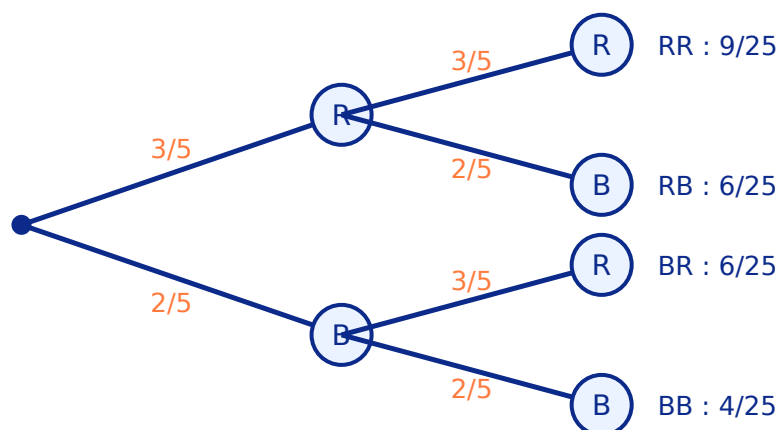
Quand on étudie deux caractères en même temps, un **tableau à double entrée** permet de lire toutes les probabilités. Dans un lycée, on interroge 150 élèves :

	Demi-pensionnaire	Externe	Total
Fille	52	28	80
Garçon	36	34	70
Total	88	62	150

On choisit un élève au hasard. Notons F : « c'est une fille » et D : « il est demi-pensionnaire ».

- $p(F) = 80/150 = 8/15$ (total de la ligne).
- $p(D) = 88/150 = 44/75$ (total de la colonne).
- $p(F \cap D) = 52/150 = 26/75$ (case à l'intersection).
- $p(F \cup D) = 80/150 + 88/150 - 52/150 = 116/150 = 58/75$.

Pour une expérience qui se déroule **en plusieurs étapes**, on utilise un **arbre pondéré**. Une urne contient 3 boules rouges (R) et 2 boules bleues (B). On tire une boule, on la remet, puis on en tire une seconde.



Règles de l'arbre

- Sur les branches issues d'un même nœud, la somme des probabilités vaut 1.
- La probabilité d'un chemin (une issue) est le **produit** des probabilités le long du chemin.
- La probabilité d'un événement est la **somme** des probabilités des chemins qui le réalisent.

Ici $p(RR) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25}$. « Une boule de chaque couleur » regroupe RB et BR : $\frac{6}{25} + \frac{6}{25} = \frac{12}{25}$. « Au moins une rouge » a pour contraire BB : $1 - \frac{4}{25} = \frac{21}{25}$.

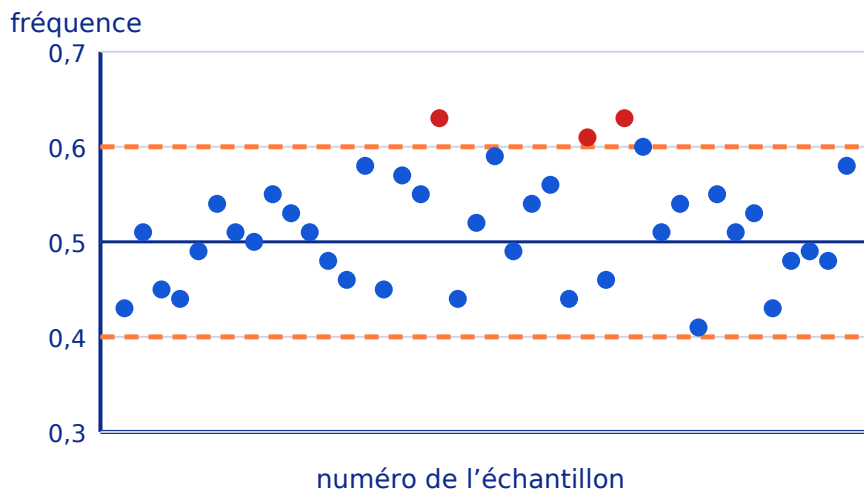
Contrôle Vérifie toujours que la somme des probabilités de toutes les feuilles de l'arbre fait 1 : ici $\frac{9}{25} + \frac{6}{25} + \frac{6}{25} + \frac{4}{25} = \frac{25}{25} = 1$.

7. Échantillonnage et fluctuation

Échantillon Un **échantillon de taille n** est la liste des résultats obtenus en répétant n fois la même expérience aléatoire, de façon indépendante (n lancers d'un dé, n personnes interrogées au hasard...). La **fréquence** d'un événement dans l'échantillon est : $f = \text{nombre de fois où il se produit} \div n$.

Si on lance 100 fois une pièce équilibrée, on s'attend à environ 50 piles, mais on en obtient rarement exactement 50. Deux échantillons de même taille donnent des fréquences **différentes** : c'est la **fluctuation d'échantillonnage**.

Nous avons simulé 40 échantillons de taille 100 (pièce équilibrée, $p = 0,5$). Les nombres de piles des 10 premiers échantillons sont : 43 ; 51 ; 45 ; 44 ; 49 ; 54 ; 51 ; 50 ; 55 ; 53. Sur l'ensemble des 40 échantillons, les fréquences vont de 0,41 à 0,63.



Chaque point est la fréquence de piles d'un échantillon. Les traits orange pointillés marquent 0,4 et 0,6. Les points bleus sont dans la bande, les points rouges en dehors (il y en a 3 sur 40).

Deux constats importants

- Les fréquences **fluctuent** autour de la probabilité p .
- Plus la taille n de l'échantillon est **grande**, moins les fréquences fluctuent : elles se resserrent autour de p .

8. Intervalle de fluctuation et prise de décision

Intervalle de fluctuation (seuil de 95 %) Soit un événement de probabilité p . Si $n \geq 25$ et $0,2 \leq p \leq 0,8$, alors, pour au moins 95 % des échantillons de taille n , la fréquence f de l'événement vérifie :

$$f \in [p - 1/\sqrt{n} ; p + 1/\sqrt{n}].$$

Cet intervalle s'appelle l'**intervalle de fluctuation** de la fréquence.

Exemple Pièce équilibrée, $p = 0,5$, $n = 100$ lancers. $1/\sqrt{100} = 1/10 = 0,1$. L'intervalle est $[0,5 - 0,1 ; 0,5 + 0,1] = [0,4 ; 0,6]$. On s'attend à ce que, dans au moins 95 % des échantillons de 100 lancers, la fréquence de piles soit entre 0,4 et 0,6. Dans notre simulation, 37 échantillons sur 40 sont dans cette bande.

Méthode : prendre une décision à partir d'un échantillon

1. Identifie la probabilité p annoncée (l'hypothèse) et la taille n de l'échantillon. Vérifie $n \geq 25$ et $0,2 \leq p \leq 0,8$.
2. Calcule $1/\sqrt{n}$ puis les bornes $p - 1/\sqrt{n}$ et $p + 1/\sqrt{n}$.
3. Calcule la fréquence observée f .
4. Compare : si f est dans l'intervalle, l'observation est **compatible** avec l'hypothèse (on ne la rejette pas) ; sinon on la **remet en cause**, en acceptant un risque d'environ 5 % de se tromper.

Exemple résolu Un fabricant affirme que 40 % de ses bonbons sont à la fraise. On en prélève 225 au hasard et on en compte 110 à la fraise. Ici $p = 0,4$ et $n = 225$, avec $0,2 \leq 0,4 \leq 0,8$ et $225 \geq 25$. $1/\sqrt{225} = 1/15 \approx 0,067$. L'intervalle est environ $[0,333 ; 0,467]$. La fréquence observée est $f = 110/225 \approx 0,489$, qui est **en dehors** de l'intervalle. On met en doute l'affirmation du fabricant.

Attention Rester dans l'intervalle ne *prouve* pas que l'hypothèse est vraie : on dit seulement que l'observation est compatible avec elle. De même, sortir de l'intervalle peut arriver par malchance (environ 5 % des cas).

9. Simuler une expérience avec Python

Un ordinateur peut répéter une expérience aléatoire des milliers de fois. La fonction **random()** du module random renvoie un nombre décimal « au hasard » dans l'intervalle $[0 ; 1[$. Pour simuler un événement de probabilité p , on regarde si ce nombre est inférieur à p : cela arrive avec une probabilité de p .

```
from random import random

def frequence(n, p):
    """Fréquence d'un événement de probabilité p
    dans un échantillon de taille n."""
    compteur = 0
    for i in range(n):
        if random() < p:
            compteur = compteur + 1
    return compteur / n

print(frequence(100, 0.5))
```

À chaque exécution, le programme affiche une fréquence différente, proche de 0,5 : c'est la fluctuation d'échantillonnage. Pour étudier la fluctuation, on appelle la fonction un grand nombre de fois et on compte les fréquences sortant de l'intervalle $[p - 1/\sqrt{n} ; p + 1/\sqrt{n}]$.

Mathbot « Relance mon programme plusieurs fois : le résultat change toujours un peu ! Avec $n = 10\,000$, il change beaucoup moins qu'avec $n = 100$. Teste-le ! »

À retenir

- Une loi de probabilité associe à chaque issue un nombre entre 0 et 1, la somme valant 1.
- En équiprobabilité : $p(A) = \text{nombre d'issues favorables} \div \text{nombre d'issues}$.

- Événement contraire : $p(\bar{A}) = 1 - p(A)$. Pour « au moins un », passe par « aucun ».
- Réunion : $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.
- Arbre pondéré : on multiplie le long d'un chemin, on additionne les chemins favorables.
- Les fréquences fluctuent d'un échantillon à l'autre, de moins en moins quand n grandit.
- Intervalle de fluctuation ($n \geq 25$, $0,2 \leq p \leq 0,8$) : $[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}]$ pour au moins 95 % des échantillons.
- Une fréquence hors de l'intervalle remet en cause l'hypothèse sur p ; dedans, elle est compatible.