# Plaquette 3 – Estimer une probabilité par une fréquence

## Échantillonnage et fluctuation — Seconde

Corrections détaillées

**Boîte à outils**

**Deux façons de choisir un modèle.**

— Par une hypothèse d'**équiprobabilité** : pièce, dé ou roue « équilibrés », tirage « au hasard » dans une population.

— Par l'**expérience** : on répète l'expérience un grand nombre de fois et on prend la fréquence observée comme valeur de la probabilité.

$$p \approx f = \frac{k}{n} \qquad \text{quand } n \text{ est grand}$$

**Fiabilité d'une estimation.**

— Plus l'échantillon est grand, plus la fréquence est, sauf exception, proche de la probabilité.

— Un grand échantillon **mal choisi** (biais de sélection) ne corrige rien : l'échantillon doit être tiré au hasard dans la population étudiée.

— Regrouper deux échantillons de la même population donne une estimation plus fiable : on additionne les effectifs et les tailles.

**Prévoir avec le modèle.** Si l'événement a la probabilité $p$, sur $N$ répétitions on peut s'attendre à **environ** $N \times p$ réalisations, la valeur exacte fluctuant d'une série à l'autre.

## Correction de l'exercice 1 – Le lancer de punaise

a) Non. Pour une pièce, la symétrie permet de supposer que les deux faces ont la même chance. La punaise n'est pas symétrique : rien ne permet de supposer les deux positions équiprobables, et aucun calcul simple ne donne la probabilité.

b) On construit le modèle à partir de l'expérience. La fréquence observée est $f = \dfrac{612}{1\,000} = 0{,}612$ ; d'après la loi des grands nombres, sur un grand nombre de lancers, la fréquence est, sauf exception, proche de la probabilité. On prend donc $p \approx 0{,}61$.

c) Sur 250 lancers, on peut prévoir **environ**

$$250 \times 0{,}612 = 153$$

lancers pointe en haut. Le résultat réel fluctuera autour de cette valeur.

**Réponse.** a) Non, la punaise n'est pas symétrique ; b) $p \approx 0{,}61$ (loi des grands nombres) ; c) environ 153.

## Correction de l'exercice 2 – Garçons et filles à la naissance

a) Sur $105 + 100 = 205$ naissances, 105 sont des garçons. La probabilité est estimée par la fréquence :

$$p \approx \frac{105}{205} \approx 0{,}512$$

b) Dans la maternité, $f = \dfrac{1\,236}{2\,400} = 0{,}515$. L'écart avec 0,512 n'est que de 0,003 : sur 2 400 naissances, il s'explique facilement par la fluctuation d'échantillonnage.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

c) Le modèle équiprobable donnerait 0,5. Les fréquences observées sur des millions de naissances, qui sont stables d'une année à l'autre autour de 0,512, montrent que ce modèle n'est pas tout à fait exact. Pour la plupart des calculs, 0,5 reste une bonne approximation, mais le modèle le plus fidèle à la réalité est $p \approx 0{,}512$.

**Réponse.** a) $p \approx 0{,}512$ ; b) $f = 0{,}515$, très proche ; c) l'équiprobabilité est une approximation ; la réalité est $\approx 0{,}512$.

## Correction de l'exercice 3 – Un dé déformé

a) Avec un dé équilibré, chaque face a la probabilité $\frac{1}{6}$ : sur 6 000 lancers, on attend environ 1 000 apparitions de chaque face.

b) On estime chaque probabilité par la fréquence observée, ce que justifie le grand nombre de lancers :

| Face | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Fréquence | 0,15 | 0,16 | 0,16 | 0,14 | 0,15 | 0,24 |

Par exemple $\frac{1\,447}{6\,000} \approx 0{,}241$.

c) La face 6 est apparue 1 447 fois au lieu d'environ 1 000 : sur 6 000 lancers, un tel écart n'est pas une simple fluctuation. Le modèle équiprobable n'est pas adapté. On peut proposer le modèle

$$P(6) \approx 0{,}24 \qquad \text{et} \qquad P(1) \approx \ldots \approx P(5) \approx 0{,}15$$

dont la somme vaut bien $5 \times 0{,}15 + 0{,}24 = 0{,}99 \approx 1$, l'écart venant des arrondis.

**Réponse.** a) Environ 1 000 ; b) 0,15 ; 0,16 ; 0,16 ; 0,14 ; 0,15 ; 0,24 ; c) non : $P(6) \approx 0{,}24$, les autres faces $\approx 0{,}15$.

## Correction de l'exercice 4 – Deux sondages, deux tailles

a) Pour l'enquête A, $f_A = \frac{12}{30} = 0{,}4$ ; pour l'enquête B,

$$f_B = \frac{555}{1\,500} = 0{,}37$$

b) Les deux échantillons sont tirés au hasard dans la même population. Le premier ne compte que 30 personnes : sa fréquence peut fluctuer fortement d'un échantillon à l'autre, et quelques réponses suffisent à la déplacer de plusieurs points (1 personne pèse $\frac{1}{30} \approx 0{,}033$). Le second compte 1 500 personnes : d'après la loi des grands nombres, sa fréquence est, sauf exception, proche de la proportion réelle. On retient donc l'estimation de l'enquête B, environ 37 % de personnes favorables.

**Réponse.** a) 0,4 et 0,37 ; b) l'enquête B, beaucoup plus grande.

## Correction de l'exercice 5 – Un échantillon mal choisi

a) La fréquence observée vaut $f = \frac{736}{800} = 0{,}92$, soit 92 %.

b) Non. La loi des grands nombres ne s'applique qu'à des échantillons tirés **au hasard** dans la population étudiée. Ici, les personnes interrogées sortent d'une salle de sport : elles font presque toutes du sport, par construction. L'échantillon n'est pas représentatif des adultes de la ville ; on parle de biais de sélection. Interroger 8 000 personnes au même endroit ne corrigerait rien.

c) Il faudrait tirer au sort les personnes interrogées dans l'ensemble des adultes de la

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

ville (par exemple sur une liste d'habitants), à différents moments, de sorte que chaque adulte ait la même chance d'être interrogé, qu'il fasse du sport ou non.

**Réponse.** a) 0,92 ; b) non : échantillon biaisé (sortie d'une salle de sport) ; c) tirer les personnes au hasard dans toute la population.

## Correction de l'exercice 6 – Compter les poissons d'un étang

a) Dans la seconde pêche, $f = \dfrac{12}{150} = 0{,}08$.

b) Dans l'étang, la proportion de poissons marqués est $\dfrac{200}{N}$. En l'égalant à la fréquence observée :

$$\frac{200}{N} \approx 0{,}08 \Leftrightarrow N \approx \frac{200}{0{,}08} = 2\,500$$

On estime qu'il y a environ 2 500 poissons dans l'étang. Ce n'est qu'une estimation : une autre pêche de 150 poissons donnerait sans doute un autre nombre de poissons marqués, donc une autre estimation.

c) Il faut que les poissons marqués aient le temps de se mélanger aux autres, sinon ils resteraient groupés près du lieu de pêche et l'échantillon ne serait pas tiré « au hasard » : la fréquence surestimerait alors la proportion de poissons marqués.

**Réponse.** a) 0,08 ; b) $N \approx 2\,500$ ; c) pour que les poissons marqués se mélangent aux autres.

## Correction de l'exercice 7 – Regrouper deux échantillons

a) Le lundi, $f_1 = \dfrac{12}{400} = 0{,}03$ ; le mardi, $f_2 = \dfrac{9}{500} = 0{,}018$.

b) On additionne les effectifs et les tailles des deux échantillons :

$$f = \frac{12 + 9}{400 + 500} = \frac{21}{900} \approx 0{,}023$$

Cette estimation, fondée sur 900 bouteilles, est plus fiable que chacune des deux précédentes.

c) La moyenne des deux fréquences, $\dfrac{0{,}03 + 0{,}018}{2} = 0{,}024$, donne le même poids aux deux jours, alors que le mardi compte plus de bouteilles. La bonne fréquence globale est une moyenne pondérée par les tailles : ici les deux résultats sont proches, mais ils peuvent différer beaucoup si les tailles sont très inégales.

**Réponse.** a) 0,03 et 0,018 ; b) $\dfrac{21}{900} \approx 0{,}023$ ; c) les deux échantillons n'ont pas la même taille.

## Correction de l'exercice 8 – Le feu tricolore

a) Sur 600 passages, la fréquence observée est $\dfrac{222}{600} = 0{,}37$. Le nombre de passages étant grand, on prend $p \approx 0{,}37$.

b) Sur 30 trajets, on peut prévoir environ

$$30 \times 0{,}37 = 11{,}1$$

soit une dizaine de feux verts, à quelques unités près.

c) En septembre, la fréquence est $\dfrac{15}{30} = 0{,}5$. Sur seulement 30 passages, la fluctuation est importante : un mois « chanceux » n'est pas surprenant et ne suffit pas à remettre en cause une estimation fondée sur 600 observations. Il faudrait que de tels écarts se

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

répètent plusieurs mois de suite pour soupçonner un changement du réglage du feu.

**Réponse.** a) $p \approx 0{,}37$ ; b) environ 11 ; c) non, 30 passages, c'est peu.

## Correction de l'exercice 9 – Prévoir des sinistres

a) La fréquence observée sur un très grand nombre de contrats est

$$f = \frac{1\,600}{50\,000} = 0{,}032$$

On estime donc la probabilité d'un sinistre à $p \approx 0{,}032$, soit 3,2 %.

b) Sur 2 000 contrats, on peut prévoir environ $2\,000 \times 0{,}032 = 64$ sinistres.

c) Non. La prévision donne un ordre de grandeur, pas une valeur exacte : le nombre de sinistres d'une agence fluctue d'une année à l'autre autour de 64. Avec 71 sinistres, la fréquence vaut $\frac{71}{2\,000} = 0{,}0355$, assez proche de 0,032 ; un tel écart est ordinaire pour un échantillon de 2 000 contrats.

**Réponse.** a) $p \approx 0{,}032$ ; b) environ 64 ; c) non, simple fluctuation ($f = 0{,}0355$).

## Correction de l'exercice 10 – Remettre en cause un modèle

a) La fréquence observée vaut $f = \frac{5\,200}{10\,000} = 0{,}52$.

b) Pour une pièce équilibrée, la simulation montre que la fréquence sur 10 000 lancers reste très proche de 0,5 : jamais, sur 300 séries, elle n'a dépassé 0,5122. La fréquence 0,52 est en dehors de tout ce qui a été observé, ce qui serait un hasard extraordinaire. Le modèle « pièce équilibrée » ne semble pas convenir : la pièce, sans doute usée ou mal frappée, favorise « face ».

c) On construit un nouveau modèle à partir de l'expérience : avec 10 000 lancers, la loi des grands nombres permet de prendre

$$p \approx 0{,}52$$

**Réponse.** a) 0,52 ; b) non : 0,52 est hors de la plage [0,4871 ; 0,5122] observée pour une pièce équilibrée ; c) $p \approx 0{,}52$.

## Correction de l'exercice 11 – Groupes sanguins

a) On divise chaque effectif par 500 :

| Groupe | O | A | B | AB |
|---|---|---|---|---|
| Fréquence | 0,414 | 0,456 | 0,092 | 0,038 |

b) Les écarts avec la répartition nationale sont faibles : 0,6 point pour O, 0,6 point pour A, 0,2 point pour B et 0,2 point pour AB. Sur un échantillon de 500 personnes, ils s'expliquent par la fluctuation d'échantillonnage : les donneurs de cette collecte ressemblent à la population française.

c) En prenant la proportion nationale 0,04 comme probabilité qu'un donneur soit du groupe AB, on peut espérer environ

$$1\,200 \times 0{,}04 = 48$$

donneurs du groupe AB, le nombre réel fluctuant autour de cette valeur.

**Réponse.** a) 0,414 ; 0,456 ; 0,092 ; 0,038 ; b) écarts de moins d'un point : compatibles ; c) environ 48.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

## Correction de l’exercice 12 – « 30 % de risque de pluie »

a) L’annonce signifie que, parmi un grand nombre de journées ayant des conditions météorologiques semblables, il a plu environ 30 % du temps : le modèle attribue la probabilité 0,3 à l’événement « il pleut demain ».

b) Ni l’un ni l’autre ne prouve que la prévision était fausse. Une probabilité de 0,3 laisse la place aux deux issues : il ne pleut pas dans environ 70 % des cas semblables, et il pleut dans environ 30 % des cas. Une seule journée ne permet pas de juger un modèle.

c) On rassemble toutes les journées pour lesquelles le service a annoncé « 30 % » sur plusieurs années, et on calcule la fréquence des journées où il a effectivement plu. Si ces journées sont nombreuses, la loi des grands nombres dit que cette fréquence devrait être proche de 0,3 si les prévisions sont fiables.

**Réponse.** a) Il pleut dans environ 30 % des situations semblables ; b) non dans les deux cas ; c) fréquence de pluie sur toutes les journées annoncées à 30 %.

## Correction de l’exercice 13 – Les retards d’une ligne de train

a) Sur 1 200 trains, la fréquence des retards est

$$f = \frac{78}{1\,200} = 0{,}065$$

On estime la probabilité d’un retard à $p \approx 0{,}065$.

b) Sur 40 trajets, il peut prévoir environ $40 \times 0{,}065 = 2{,}6$ retards, soit 2 ou 3 retards dans un mois « ordinaire ».

c) En octobre, la fréquence vaut $\frac{6}{40} = 0{,}15$. C’est plus que prévu, mais 40 trajets forment un petit échantillon : quelques retards de plus suffisent à doubler la fréquence. Un seul mois ne permet pas de conclure ; il faudrait observer la ligne sur plusieurs mois, ou sur un grand nombre de trains, pour savoir si la probabilité de retard a vraiment augmenté.

**Réponse.** a) $p \approx 0{,}065$ ; b) environ 2,6, soit 2 ou 3 ; c) pas forcément : 40 trajets, c’est peu.

## Correction de l’exercice 14 – La chute d’un gobelet

a) On divise l’effectif « sur le côté » par le nombre de lancers : Inès $\frac{34}{50} = 0{,}68$, Malo $\frac{151}{200} \approx 0{,}76$, Sami $\frac{552}{750} \approx 0{,}74$.

b) On additionne effectifs et nombres de lancers, en supposant que les trois élèves ont lancé le même gobelet de la même façon :

$$f = \frac{34 + 151 + 552}{50 + 200 + 750} = \frac{737}{1\,000} = 0{,}737$$

On estime donc la probabilité à environ 0,74.

c) Inès n’a fait que 50 lancers : sa fréquence fluctue beaucoup d’une série à l’autre. L’estimation regroupée repose sur 1 000 lancers ; d’après la loi des grands nombres, elle est, sauf exception, bien plus proche de la probabilité.

**Réponse.** a) 0,68 ; 0,76 ; 0,74 ; b) 0,737 ; c) elle repose sur 1 000 lancers au lieu de 50.

## Correction de l’exercice 15 – Synthèse : des graines qui germent

a) La fréquence est $\frac{17}{25} = 0{,}68$, soit 68 %. Elle est inférieure aux 80 % annoncés, mais sur 25 graines seulement, la fluctuation est forte : quelques graines de plus ou de moins font

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

varier la fréquence de 4 points chacune. Ce résultat ne suffit pas à accuser le fabricant.

b) La fréquence vaut $\frac{318}{400} = 0{,}795$, soit 79,5 %.

c) En regroupant, $\frac{17+318}{25+400} = \frac{335}{425} \approx 0{,}788$. Sur 425 graines, la fréquence observée est très proche de 0,8 : l'indication du fabricant est crédible, et le premier semis n'était qu'une fluctuation.

d) Si chaque graine germe avec une probabilité d'environ 0,8, semer $N$ graines donne environ $0{,}8N$ plants. On cherche $N$ tel que

$$0{,}8\,N = 1\,200 \Leftrightarrow N = \frac{1\,200}{0{,}8} = 1\,500$$

Il doit semer environ 1 500 graines, et même un peu plus s'il veut être sûr d'avoir ses 1 200 plants, le nombre réel de germinations fluctuant autour de la prévision.

**Réponse.** a) 0,68, fluctuation sur 25 graines ; b) 0,795 ; c) ≈ 0,788, crédible ; d) environ 1 500 graines.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).