1. Rappels de 2nde : échantillonnage et loi des grands nombres

Ex. 1 ★☆☆ **Automatismes : fréquences et intervalles**

(a) $f = \frac{26}{80} = 0{,}325$, soit 32,5 %.

(b) $0{,}075 = \frac{7{,}5}{100}$, donc $f = 7{,}5\,\%$.

(c) $42\,\% = \frac{42}{100} = 0{,}42$.

(d) $0{,}36 \times 250 = 90$: 90 personnes ont un chien.

(e) $\frac{1}{\sqrt{900}} = \frac{1}{30} \approx 0{,}033$.

(f) $\frac{1}{\sqrt{100}} = 0{,}1$, donc l'intervalle est $[0{,}2\,;0{,}4]$.

(g) $\frac{1}{\sqrt{2\,500}} = \frac{1}{50} = 0{,}02$, donc l'intervalle est $[0{,}58\,;0{,}62]$.

Ex. 2 ★☆☆ **Automatismes : fréquence ou probabilité ?**

(a) Faux. $60 \times \frac{1}{6} = 10$ est seulement le nombre de six « attendu ». À cause de la fluctuation d'échantillonnage, on obtient souvent 8, 11, 13…

(b) Faux. $\frac{15}{60} = 0{,}25$ est la fréquence observée sur cet échantillon. La probabilité $p = \frac{1}{6}$ reste fixe : c'est f qui fluctue d'un échantillon à l'autre.

(c) Faux. Leurs fréquences sont $\frac{9}{60} = 0{,}15$ et $\frac{12}{60} = 0{,}2$. Deux échantillons de même taille donnent en général des fréquences différentes : c'est la fluctuation d'échantillonnage, pas une erreur.

(d) Faux. f se stabilise autour de $\frac{1}{6}$: l'écart devient en général très petit, mais f n'est pas pour autant égale à $\frac{1}{6}$.

(e) Faux. On a $\frac{1}{\sqrt{100}} = 0{,}1$, mais l'inégalité est vraie seulement pour environ 95\,% des échantillons, pas pour 100\,%.

(f) Vrai. $\frac{1}{\sqrt{3\,600}} = \frac{1}{60}$: c'est la propriété de 2nde, vraie pour environ 95\,% des échantillons.

Ex. 3 ★★☆ **Vingt échantillons sur un nuage**

(a) $\frac{1}{\sqrt{100}} = 0{,}1$, donc l'intervalle est $[0{,}3\,;0{,}5]$.

(b) 2 fréquences : 0,29 (échantillon 4) et 0,52 (échantillon 13).

(c) $\frac{18}{20} = 0{,}9$, soit 90 % des échantillons.

Ex. 3 (suite)

(d) Non. La propriété dit « environ 95\,% des échantillons ». Avec seulement 20 échantillons, cette proportion fluctue elle aussi. 90 % est une valeur tout à fait normale.

Ex. 4 ★★☆ **Contrôle qualité : corriger un programme**

(a) `k` compte les pièces défectueuses de l'échantillon. `random()` tombe dans $[0\,;0{,}08[$ avec la probabilité 0,08 : c'est la probabilité qu'une pièce soit défectueuse.

(b) `return k / n` est indenté dans la boucle : la fonction s'arrête dès le premier tour. Il faut le désindenter (au niveau de `for`) pour qu'il s'exécute après la boucle.

(c) $0{,}092\,8 \times 625 = 58$ pièces défectueuses.

(d) Non. La fréquence change d'un échantillon à l'autre, c'est la fluctuation d'échantillonnage ($0{,}067\,2 \times 625 = 42$ pièces). La probabilité $p = 0{,}08$, elle, reste fixe.

(e) $\frac{1}{\sqrt{625}} = \frac{1}{25} = 0{,}04$, donc l'intervalle est $[0{,}04\,;0{,}12]$. On a $0{,}04 \leqslant 0{,}067\,2 \leqslant 0{,}092\,8 \leqslant 0{,}12$: oui, les deux fréquences sont dans l'intervalle, comme pour environ 95\,% des échantillons.

Ex. 5 ★★★ **Tickets à gratter : la fréquence se stabilise**

(a) $f = \frac{4}{10} = 0{,}4$; $\frac{21}{100} = 0{,}21$; $\frac{262}{1\,000} = 0{,}262$; $\frac{2\,493}{10\,000} = 0{,}249\,3$.

(b) Les écarts sont 0,15 ; 0,04 ; 0,012 ; 0,000 7.

(c) L'écart diminue : f se stabilise autour de 0,25. Mais aucune fréquence n'est égale à 0,25 : la stabilisation ne veut pas dire égalité.

(d) Non. Sur un petit échantillon, la fluctuation d'échantillonnage est forte : 4 gagnants au lieu des $10 \times 0{,}25 = 2{,}5$ « attendus », ce n'est pas anormal. Pour $n = 10\,000$, $f = 0{,}249\,3$ est très proche de 0,25 : c'est un grand échantillon qui renseigne sur p.

Ex. 6 ★★★ **Sondage vélo : 95 %, pas 100 %**

(a) $f = \frac{119}{400} = 0{,}297\,5$, soit 29,75 % de cyclistes dans l'échantillon.

(b) $\sqrt{400} = 20$, donc $\frac{1}{\sqrt{400}} = 0{,}05$. L'intervalle est $[0{,}35 - 0{,}05\,;0{,}35 + 0{,}05] = [0{,}30\,;0{,}40]$. Comme $0{,}297\,5 < 0{,}30$, f n'appartient pas à l'intervalle.

Ex. 6 **(suite)**

(c) Non. L'intervalle contient f pour environ 95\,% des échantillons de taille 400, et pas pour tous : environ 5\,% des échantillons donnent une fréquence en dehors. L'observation d'Ambre est surprenante pour le modèle $p = 0{,}35$, mais elle ne prouve rien avec certitude.

(d) Environ 5\,% des échantillons donnent une fréquence hors de l'intervalle : $0{,}05 \times 60 = 3$. On s'attend à environ 3 sondages en dehors (ce nombre fluctue lui aussi). L'échantillon d'Ambre peut très bien être l'un d'eux.

(e) $\frac{1}{\sqrt{1\,600}} = \frac{1}{40} = 0{,}025$: l'intervalle est $[0{,}325\,;0{,}375]$, de largeur 0,05 contre 0,1 pour $n = 400$. Quand n est multiplié par 4, $\sqrt{n}$ est multiplié par 2 : la largeur est donc divisée par 2.

Ex. 7 ★★★ **Pile ou face : pas de compensation**

(a) Non. Les lancers sont indépendants : la pièce n'a pas de mémoire. Au 9e lancer, la probabilité de pile vaut toujours $1 - 0{,}5 = 0{,}5$, quels que soient les résultats précédents.

(b) La somme $x_1 + \cdots + x_n$ ajoute 1 pour chaque face et 0 pour chaque pile : elle vaut le nombre k de faces. Donc $m = \frac{k}{n} = f$.

(c) Les excédents valent $56 - 50 = 6$, $1\,280 - 1\,250 = 30$ et $20\,160 - 20\,000 = 160$: l'excédent de faces augmente, les piles n'ont rien compensé.

(d) $f - 0{,}5 = \frac{k}{n} - \frac{n/2}{n} = \frac{k - n/2}{n}$. On obtient $\frac{6}{100} = 0{,}06$, $\frac{30}{2\,500} = 0{,}012$ et $\frac{160}{40\,000} = 0{,}004$. L'excédent grandit moins vite que n : il est « dilué », donc f se stabilise autour de 0,5.

(e) $\frac{1}{\sqrt{40\,000}} = \frac{1}{200} = 0{,}005$: l'intervalle est $[0{,}495\,;0{,}505]$. $f = \frac{20\,160}{40\,000} = 0{,}504$ est dans l'intervalle, comme pour environ 95\,% des échantillons de cette taille. Un excédent de 160 faces n'a rien de surprenant.

Ex. 8 ★★★ **La roue de la kermesse**

(a) `randint(1, 10)` donne chaque entier de 1 à 10 avec la probabilité 0,1. $X = 8$ si $r = 1$: probabilité 0,1. $X = 2$ si $r \in \{2, 3, 4\}$: probabilité 0,3. $X = 0$ si $r \in \{5, \ldots, 10\}$: probabilité 0,6. Tableau : x_i : 0 ; 2 ; 8 et $P(X = x_i)$: 0,6 ; 0,3 ; 0,1
(total 1).

(b) $\mu = 0 \times 0{,}6 + 2 \times 0{,}3 + 8 \times 0{,}1 = 0{,}6 + 0{,}8 = 1{,}4$ €.

Ex. 8 **(suite)**

(c)
```
def moyenne(n):
    s = 0
    for k in range(n):
        s = s + gain()
    return s / n
```
On place `s = 0` avant la boucle et `return s / n` après la boucle, sans indentation sous `for`. Indenté dans la boucle, il s'exécuterait dès le premier tour et renverrait $\frac{x_1}{n}$.

(d) $|1{,}9 - 1{,}4| = 0{,}5$; $|1{,}27 - 1{,}4| = 0{,}13$; $|1{,}418 - 1{,}4| = 0{,}018$. Aucune moyenne n'est égale à $\mu = 1{,}4$, mais l'écart $|m - \mu|$ diminue en général quand la taille n augmente.

(e) Sur un grand échantillon, le gain moyen m d'un joueur est proche de $\mu = 1{,}4$ €. Chaque partie rapporte donc environ $1{,}5 - 1{,}4 = 0{,}1$ € à la kermesse, soit environ $2\,000 \times 0{,}1 = 200$ €. Ce n'est qu'une estimation : avec $m = 1{,}418$, on aurait $2\,000 \times (1{,}5 - 1{,}418) = 164$ €.

2. *Simuler une variable aléatoire en Python*

Ex. 9 ★☆☆ **Probabilité qu'un test soit vrai**

(a) u tombe dans $[0\,;0{,}35[$, de longueur 0,35 : la probabilité vaut 0,35.

(b) u tombe dans $[0{,}2\,;0{,}65[$, de longueur $0{,}65 - 0{,}2 = 0{,}45$: la probabilité vaut 0,45.

(c) u tombe dans $[0{,}72\,;1[$, de longueur $1 - 0{,}72 = 0{,}28$: la probabilité vaut 0,28.

(d) `randint(1, 6)` a 6 valeurs possibles et 4 d'entre elles passent le test (de 1 à 4) : $\frac{4}{6} = \frac{2}{3}$.

(e) Il y a 5 valeurs, de 0 à 4, et 0, 1 passent le test : $\frac{2}{5}$, et non $\frac{1}{5}$ ni $\frac{1}{4}$.

(f) Il y a 20 valeurs. Le test strict garde 1, 2, 3 et 4 : $\frac{4}{20} = \frac{1}{5}$.

(g) Il y a $12 - 5 + 1 = 8$ valeurs, 12 compris, donc la probabilité vaut $\frac{1}{8}$.

Ex. 10 ★☆☆ **Seuils cumulés**

(a) Les seuils sont 0,15, puis $0{,}15 + 0{,}45 = 0{,}6$.

(b) Les seuils sont 0,25, puis $0{,}25 + 0{,}6 = 0{,}85$: le programme renvoie 0 pour $u \in [0{,}25\,;0{,}85[$, un intervalle de longueur 0,6.

(c) $X = 20$ pour $u \in [0{,}08\,;0{,}66[$: $P(X = 20) = 0{,}66 - 0{,}08 = 0{,}58$. $X = 50$ dans le else, pour $u \in [0{,}66\,;1[$: $P(X = 50) = 1 - 0{,}66 = 0{,}34$.

Ex. 10 (suite)

(d) Les seuils sont 0,1, puis $0{,}1 + 0{,}35 = 0{,}45$, puis $0{,}45 + 0{,}4 = 0{,}85$.

(e) La 2e valeur correspond à $[0{,}24\,;0{,}46[$, de longueur $0{,}46 - 0{,}24 = 0{,}22$ au lieu de 0,46. Il fallait écrire le seuil cumulé $0{,}24 + 0{,}46 = 0{,}7$.

(f) Les seuils sont 0,4 puis 0,75. La dernière valeur, 9, va dans le else : elle correspond à $u \in [0{,}75\,;1[$, de longueur 0,25.

Ex. 11 ★★☆ Retrouver la loi d'un jeu de fléchettes

(a) Non. La fonction renvoie 2 lorsque $0{,}12 \leqslant u < 0{,}4$. Cet intervalle a pour longueur $0{,}4 - 0{,}12 = 0{,}28$, donc $P(X = 2) = 0{,}28$. Le seuil 0,4 est cumulé.

(b) $P(X = 5) = 0{,}12$ (intervalle $[0\,;0{,}12[$) ; $P(X = 2) = 0{,}28$; $P(X = 0) = 1 - 0{,}4 = 0{,}6$
(le else : $[0{,}4\,;1[$).

(c) $\mu = 5 \times 0{,}12 + 2 \times 0{,}28 + 0 \times 0{,}6 = 0{,}6 + 0{,}56 = 1{,}16$ point.

(d) Les nouveaux seuils sont 0,2, puis $0{,}2 + 0{,}28 = 0{,}48$: on écrit `if u < 0.2:` puis `elif u < 0.48:`. Le else ne change pas.

Ex. 12 ★★☆ Corriger le programme de Malik

(a) On arrive au elif si le premier tirage vaut au moins 0,65 (probabilité 0,35), et il faut que le second soit $< 0{,}25$ (probabilité 0,25) : $0{,}35 \times 0{,}25 = 0{,}0875$, au lieu de 0,25.

(b) Le premier tirage vaut au moins 0,65 et le second au moins 0,25 : $0{,}35 \times 0{,}75 = 0{,}2625$, au lieu de 0,1. Deux tirages faussent la loi.

(c) Si on arrive au elif, c'est que $u \geqslant 0{,}65$. Le test $u < 0{,}25$ est alors toujours faux, donc la probabilité vaut 0 : le seuil n'est pas cumulé.

(d) Un seul tirage, avec les seuils 0,65 puis $0{,}65 + 0{,}25 = 0{,}9$:

```
from random import random
def pannes():
    u = random()
    if u < 0.65:
        return 0
    elif u < 0.9:
        return 1
    else:
        return 2
```

Ex. 13 ★★☆ Pièces défectueuses

(a)

```
from random import random
def defaut():
    if random() < 0.08:
        return 1
    else:
        return 0
```

On a bien $P(\text{random}() < 0{,}08) = 0{,}08$.

(b) Oui. Le test retient 8 valeurs sur les 100 entiers de 1 à 100 : la probabilité vaut $\frac{8}{100} = 0{,}08$.

(c) Non. `randint(0, 100)` a 101 valeurs, 100 compris, et 8 passent le test (de 0 à 7) : $\frac{8}{101} \approx 0{,}0792 \neq 0{,}08$.

(d) =SI(ALEA()<0,08;1;0), car =ALEA()<0,08 est VRAI avec la probabilité 0,08. On recopie la formule sur 500 lignes.

(e) D suit une loi de Bernoulli de paramètre $p = 0{,}08$: $\mu = 0{,}08$ et $\sigma = \sqrt{0{,}08 \times 0{,}92} = \sqrt{0{,}0736} \approx 0{,}2713$.

Ex. 14 ★★★ La roue de la kermesse

(a) `randint(1, 20)` renvoie l'un des 20 entiers de 1 à 20, chacun avec la probabilité $\frac{1}{20}$. Le test `k <= 3` est vrai pour 1, 2 et 3 : $P(X = 10) = \frac{3}{20} = 0{,}15$. On n'arrive au `elif` que si $k \geqslant 4$, et il est vrai pour k de 4 à 9, soit 6 entiers : $P(X = 2) = \frac{6}{20} = 0{,}3$. Le `else` regroupe les 11 entiers de 10 à 20 : $P(X = 0) = \frac{11}{20} = 0{,}55$.

(b) $E(X) = 10 \times 0{,}15 + 2 \times 0{,}3 + 0 \times 0{,}55 = 2{,}1$. En moyenne, un joueur gagne 2,10 € pour un ticket à 2 €. Sur un grand nombre de parties, la kermesse perd donc en moyenne environ 0,10 € par partie : le jeu lui fait perdre de l'argent.

(c) `k < 3` est vrai seulement pour 1 et 2 :
$P(X = 10) = \frac{2}{20} = 0{,}1$. `k < 9` est vrai pour k de 3 à 8, soit 6 entiers : $P(X = 2) = \frac{6}{20} = 0{,}3$. Le `else` regroupe les 12 entiers de 9 à 20 : $P(X = 0) = \frac{12}{20} = 0{,}6$. Son espérance vaut $10 \times 0{,}1 + 2 \times 0{,}3 = 1{,}6$, et non 2,1 : passer de <= à < change la loi.

(d) Les seuils cumulés sont 0,15 puis $0{,}15 + 0{,}3 = 0{,}45$.

```
from random import random
def roue():
    u = random()
    if u < 0.15:
        return 10
    elif u < 0.45:
        return 2
    else:
        return 0
```

$[0\,;0{,}15[$ a pour longueur 0,15 et $[0{,}15\,;0{,}45[$ a pour longueur 0,3. Le `else` correspond à $[0{,}45\,;1[$, de longueur 0,55.

Ex. 15 ★★★ **Le programme de Chloé à réparer**

(a) La fonction renvoie 0 si $u \in [0\,;0{,}1[$, avec la probabilité 0,1 (c'est correct). Elle renvoie 1 si $u \in [0{,}1\,;0{,}25[$, avec la probabilité 0,15 au lieu de 0,25. Elle renvoie 2 si $u \in [0{,}25\,;0{,}4[$, avec la probabilité 0,15 au lieu de 0,4. Le dernier test n'est examiné que si $u \geqslant 0{,}4$, et alors `u < 0.25` est toujours faux : elle renvoie 3 avec la probabilité 0.

(b) Quand $u \geqslant 0{,}4$, aucun test n'est vrai et il n'y a pas de `else` : la fonction ne renvoie rien. Cela arrive avec la probabilité $1 - 0{,}4 = 0{,}6$. Première erreur : Chloé a recopié les probabilités au lieu de les cumuler. Seconde erreur : la dernière valeur doit aller dans un `else`.

(c) Les seuils cumulés sont 0,1, puis $0{,}1 + 0{,}25 = 0{,}35$, puis $0{,}35 + 0{,}4 = 0{,}75$ (toujours après `from random import random`) :

```
def buts():
    u = random()
    if u < 0.1:
        return 0
    elif u < 0.35:
        return 1
    elif u < 0.75:
        return 2
    else:
        return 3
```

Les intervalles ont pour longueurs 0,1 ; 0,25 ; 0,4, et le `else` correspond à $[0{,}75\,;1[$, de longueur 0,25.

(d) On utilise les mêmes seuils cumulés : `=SI(A1<0,1;0;SI(A1<0,35;1;SI(A1<0,75;2;3)))`. La valeur 3 joue le rôle du `else` : c'est le reste de $[0\,;1[$.

Ex. 16 ★★★ **Lancers francs : choisir le bon test**

(a) `randint(1, 20) <= 7` : 7 entiers sur 20, soit $\frac{7}{20} = 0{,}35$. Il convient. `randint(0, 20) <= 7` : il y a 21 entiers de 0 à 20, et 8 d'entre eux vérifient le test (de 0 à 7). On obtient $\frac{8}{21} \approx 0{,}381$: il ne convient pas. `randint(0, 19) < 7` : 7 entiers (0 à 6) sur 20, soit 0,35. Il convient. `randint(1, 20) < 7` : 6 entiers sur 20, soit 0,3. Il ne convient pas.

(b) Les deux tests conviennent. `random() > 0.65` est vrai si u tombe entre 0,65 et 1, dans un intervalle de longueur $1 - 0{,}65 = 0{,}35$. `random() <= 0.35` correspond à $[0\,;0{,}35]$, de longueur 0,35. Ajouter le seul nombre 0,35 ne change pas la probabilité.

(c) $0{,}355 \times 20 = 7{,}1$ n'est pas un entier : aucun nombre d'entiers parmi 20 ne donne la probabilité 0,355. On peut écrire `randint(1, 1000) <= 355` : il est vrai pour 355 entiers sur 1 000, soit avec la probabilité 0,355.

Ex. 16 **(suite)**

(d) `=SI(ALEA.ENTRE.BORNES(1;20)<=7;1;0)`. Le test est vrai pour 7 entiers sur 20, donc avec la probabilité $\frac{7}{20} = 0{,}35$. Recopiée sur n lignes, cette formule donne un échantillon de taille n de la loi de R.

3. *Moyenne d'un échantillon de taille n*

Ex. 17 ★☆☆ **Calculer la moyenne d'un échantillon**

(a) Ici $n = 8$ et la somme compte les 1 : $m = \frac{5}{8} = 0{,}625$. C'est la fréquence des paiements sans contact dans l'échantillon.

(b) C'est une moyenne pondérée par les effectifs. On a $n = 12 + 6 + 2 = 20$, et non 3. $m = \frac{1\times12+2\times6+5\times2}{20} = \frac{34}{20} = 1{,}7$.

(c) La somme vaut bien 60, mais l'échantillon contient $n = 5$ valeurs : on divise par n, pas par $n - 1$. Donc $m = \frac{60}{5} = 12$ minutes.

(d) La moyenne d'une variable 0/1 est une fréquence : $m = \frac{26}{400} = 0{,}065$.

(e) $|m - \mu| = |1{,}38 - 1{,}5| = |-0{,}12| = 0{,}12$.

(f) Une distance est toujours positive : $|m - \mu| = |-0{,}4| = 0{,}4$.

Ex. 18 ★☆☆ **Lire un programme de moyenne**

(a) `randint(1, 4)` renvoie un entier entre 1 et 4, bornes comprises, de façon équiprobable : X prend les valeurs 1, 2, 3, 4, chacune avec la probabilité $\frac{1}{4}$.

(b) $\mu = 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} = \frac{10}{4} = 2{,}5$.

(c) `range(30)` fait 30 tours (k va de 0 à 29) : on additionne 30 valeurs.

(d) La ligne `s = s + randint(1, 4)` devient `s = s + randint(1, 6)`.

(e) La fonction s'arrêterait au premier tour et renverrait $\frac{x_1}{30}$, c'est-à-dire la première valeur divisée par 30.

(f) Non. $\mu = 2{,}5$ est fixe, mais la moyenne m change d'un échantillon à l'autre : c'est la fluctuation d'échantillonnage. Obtenir deux résultats différents est normal.

Ex. 19 ★★☆ **Le programme d'Inès qui ne marche pas**

(a) X vaut 1 si la graine germe (quand `random() < 0.35`, avec la probabilité 0,35) et 0 sinon. C'est une variable de Bernoulli de paramètre 0,35, donc $\mu = 0{,}35$.

Ex. 19 **(suite)**

(b) `s = 0` est dans la boucle : l'accumulateur est remis à zéro à chaque tour. À la fin, `s` vaut 1 si la dernière graine a germé et 0 sinon. Il faut placer `s = 0` avant la boucle.

(c) On doit diviser par n, pas par $n-1$: on a additionné n valeurs. Programme corrigé :

```
from random import random

def frequence(n):
    s = 0
    for k in range(n):
        if random() < 0.35:
            s = s + 1
    return s / n
```

(d) `s` compte les graines qui ont germé, et $m = \frac{s}{200} = 0{,}315$. Donc $s = 0{,}315 \times 200 = 63$. $|m - \mu| = |0{,}315 - 0{,}35| = 0{,}035$.

Ex. 20 ★★☆ **Attente à la cantine**

(a) $\mu = 2 \times 0{,}5 + 5 \times 0{,}4 + 10 \times 0{,}1 = 1 + 2 + 1 = 4$ minutes.

(b) Ici $n = 28 + 17 + 5 = 50$ jours. $m = \frac{2\times 28 + 5 \times 17 + 10 \times 5}{50} = \frac{191}{50} = 3{,}82$ minutes.

(c) $|m - \mu| = |3{,}82 - 4| = 0{,}18$ minute.

(d) Non. Même si la loi est la bonne, m fluctue d'un échantillon à l'autre et vaut en général autre chose que μ. Un écart de 0,18 minute ne suffit pas à prouver que le modèle est faux.

Ex. 21 ★★☆ **Estimer la part des cyclistes**

(a) La variable vaut 0 ou 1. La somme $x_1 + \cdots + x_n$ compte donc les cyclistes k, et $m = \frac{k}{n} = f$.

(b) $0{,}226 \times 500 = 113$ cyclistes et $0{,}214 \times 2\,000 = 428$ cyclistes.

(c) On choisit $f = 0{,}214$, obtenue avec le plus grand échantillon ($n = 2\,000$). L'écart $|m - p|$ est en général plus petit quand n augmente, donc cette estimation est la plus fiable.

(d) Faux. 0,214 décrit un échantillon. C'est une estimation de p, pas sa valeur exacte : un autre échantillon de taille $2\,000$ donnerait en général une autre fréquence.

(e) Faux. Ces personnes ne sont pas choisies au hasard : elles prennent sans doute le train plutôt que le vélo, et l'échantillon risque de sous-estimer p.

Ex. 22 ★★★ **La roue de la kermesse**

(a) $\mu = 0 \times 0{,}55 + 2 \times 0{,}3 + 5 \times 0{,}15 = 0 + 0{,}6 + 0{,}75 = 1{,}35$ €.

(b) Le `return` est indenté dans la boucle : il s'exécute dès le premier tour et la fonction s'arrête. Un seul gain x_1 est simulé, et la fonction renvoie $\frac{x_1}{20}$ au lieu de la moyenne. En plus, `s = 0` dans la boucle remettrait le total à zéro à chaque tour.

(c)

```
def moyenne(n):
    s = 0
    for k in range(n):
        s = s + roue()
    return s / n
```

`s = 0` passe avant la boucle : on initialise le total une seule fois. `s = s + roue()` reste dans la boucle : on ajoute un gain par tour, donc n gains en tout. `return s / n` passe après la boucle : on ne divise par n qu'une fois les n gains additionnés.

(d) $|1{,}6 - 1{,}35| = 0{,}25$ et $|1{,}37 - 1{,}35| = 0{,}02$. Les deux moyennes diffèrent de μ : c'est normal, car μ est fixe alors que m fluctue d'un échantillon à l'autre. Mais la moyenne de l'échantillon de taille 2 000 est bien plus proche de μ : quand n augmente, m se stabilise autour de μ.

(e) Sur un grand nombre de parties, le gain moyen d'un joueur est proche de $\mu = 1{,}35$ €. L'organisateur peut donc espérer environ $1{,}50 - 1{,}35 = 0{,}15$ € par partie. Les 20 parties ne le montraient pas : avec $m = 1{,}6$, il aurait versé en moyenne $1{,}6 - 1{,}50 = 0{,}10$ € de plus que la mise par partie, donc perdu de l'argent. Avec $n = 2\,000$, il gagne $1{,}50 - 1{,}37 = 0{,}13$ € par partie, ce qui est proche de 0,15 €.

Ex. 23 ★★★ **Des bulles dans les carreaux**

(a) Sacha fait la moyenne des quatre valeurs possibles, sans tenir compte des effectifs. Or l'échantillon compte 80 carreaux : chaque valeur doit être comptée autant de fois qu'elle apparaît (moyenne pondérée). $m = \frac{0\times 38 + 1\times 25 + 2\times 12 + 3\times 5}{80} = \frac{64}{80} = 0{,}8$ bulle par carreau.

(b) La somme des valeurs de Y compte les 1, c'est-à-dire les 38 carreaux sans bulle. La moyenne vaut donc $\frac{38}{80} = 0{,}475$. C'est la fréquence des carreaux sans bulle dans l'échantillon : une fréquence est la moyenne d'une variable qui vaut 0 ou 1.

(c) $\mu = 0 \times 0{,}5 + 1 \times 0{,}3 + 2 \times 0{,}15 + 3 \times 0{,}05 = 0{,}75$. Ensuite, $|m - \mu| = |0{,}8 - 0{,}75| = 0{,}05$.

Ex. 23 (suite)

(d) Y vaut 1 quand $X = 0$: c'est une variable de Bernoulli de paramètre $P(X = 0) = 0{,}5$, donc $E(Y) = 0{,}5$. La distance vaut $|0{,}475 - 0{,}5| = 0{,}025$.

(e) Les 80 carreaux de Sacha portent $0{,}8 \times 80 = 64$ bulles, et ceux du collègue $0{,}65 \times 120 = 78$ bulles. La moyenne des 200 carreaux vaut $m = \frac{64+78}{200} = \frac{142}{200} = 0{,}71$. Ce n'est pas $\frac{0{,}8+0{,}65}{2} = 0{,}725$: les deux échantillons n'ont pas la même taille, et chaque carreau doit compter une fois (moyenne pondérée).

Ex. 24 ★★★ **Le dé à huit faces de Wassim**

(a) Le dé est équilibré : X prend les valeurs $1, 2, \dots, 8$, chacune avec la probabilité $\frac{1}{8}$. $\mu = \frac{1+2+\cdots+8}{8} = \frac{36}{8} = 4{,}5$.

(b) `range(10)` fait 10 tours, avec k qui prend les valeurs $0, 1, \dots, 9$. À chaque tour, on ajoute k et non un résultat du dé : $s = 0 + 1 + \cdots + 9 = 45$. L'appel renvoie $\frac{45}{10} = 4{,}5$.

(c) La moyenne d'un vrai échantillon fluctue d'un appel à l'autre. Ici, on obtient toujours le même résultat parce que `randint` n'est jamais appelée : le programme n'a rien d'aléatoire, et il ne tombe sur μ que par coïncidence. D'ailleurs, `moyenne(100)` renvoie $\frac{0+1+\cdots+99}{100} = \frac{4\,950}{100} = 49{,}5$. C'est impossible pour un dé dont les faces vont de 1 à 8. Correction : la ligne de la boucle devient `s = s + randint(1, 8)`, qui ajoute un lancer par tour.

(d) $|5{,}2 - 4{,}5| = 0{,}7$; $|4{,}37 - 4{,}5| = 0{,}13$; $|4{,}541 - 4{,}5| = 0{,}041$; $|4{,}4968 - 4{,}5| = 0{,}0032$. Une distance est toujours positive, même quand $m < \mu$. Aucune moyenne n'est égale à 4,5, mais l'écart diminue nettement quand la taille n augmente.

(e) Faux. $\mu = E(X) = 4{,}5$ ne dépend que de la loi du dé : elle est fixe. Ce qui change d'un échantillon à l'autre, c'est la moyenne m. Ici, elle varie entre 4,18 et 4,86 autour de 4,5 : c'est la fluctuation d'échantillonnage.

4. *L'écart $2\sigma/\sqrt{n}$*

Ex. 25 ★☆☆ **Calculer l'écart $\frac{2\sigma}{\sqrt{n}}$**

(a) $\frac{2\times5}{\sqrt{100}} = \frac{10}{10} = 1$.

(b) $\frac{2\times1{,}8}{\sqrt{36}} = \frac{3{,}6}{6} = 0{,}6$.

(c) On prend d'abord l'écart type : $\sigma = \sqrt{0{,}64} = 0{,}8$.
Donc $\frac{2\times0{,}8}{\sqrt{64}} = \frac{1{,}6}{8} = 0{,}2$.

(d) Léo a divisé par n au lieu de $\sqrt{n}$. On a $\frac{2\times4}{\sqrt{25}} = \frac{8}{5} = 1{,}6$.

Ex. 25 (suite)

(e) $\sigma = \sqrt{0{,}1 \times 0{,}9} = \sqrt{0{,}09} = 0{,}3$.
Donc $\frac{2\times0{,}3}{\sqrt{400}} = \frac{0{,}6}{20} = 0{,}03$.

(f) $\sigma = \sqrt{2{,}25} = 1{,}5$ (et non 2,25).
Donc $\frac{2\times1{,}5}{\sqrt{100}} = \frac{3}{10} = 0{,}3$.

(g) $\sigma = \sqrt{0{,}15 \times 0{,}85} = \sqrt{0{,}1275}$. On garde cette valeur exacte : $\frac{2\sqrt{0{,}1275}}{\sqrt{50}} \approx 0{,}101$. Si on arrondit σ à 0,36 avant la fin, on trouve 0,102 : c'est faux.

Ex. 26 ★★☆ **Écrire et lire l'intervalle centré sur μ**

(a) $[50 - 1{,}5\,;50 + 1{,}5] = [48{,}5\,;51{,}5]$.

(b) $|m - 7| \leqslant 0{,}3 \iff 7 - 0{,}3 \leqslant m \leqslant 7 + 0{,}3$, soit $6{,}7 \leqslant m \leqslant 7{,}3$.

(c) Le centre vaut $\frac{4{,}6+5{,}4}{2} = 5$ et la demi-largeur $\frac{5{,}4-4{,}6}{2} = 0{,}4$: $|m - 5| \leqslant 0{,}4$.

(d) On prend d'abord l'écart type : $\sigma = \sqrt{1{,}21} = 1{,}1$.
Donc $\frac{2\times1{,}1}{\sqrt{121}} = \frac{2{,}2}{11} = 0{,}2$, d'où l'intervalle $[4{,}3\,;4{,}7]$.

(e) $\mu = 0{,}5$ et $\sigma = \sqrt{0{,}5 \times 0{,}5} = 0{,}5$.
Donc $\frac{2\times0{,}5}{\sqrt{625}} = \frac{1}{25} = 0{,}04$, d'où l'intervalle $[0{,}46\,;0{,}54]$.

(f) L'élève a centré l'intervalle sur m. Il doit le centrer sur μ (le modèle) : $[10 - 0{,}4\,;10 + 0{,}4] = [9{,}6\,;10{,}4]$. Et $m = 10{,}6$ est en dehors.

(g) $\frac{2\times0{,}9}{\sqrt{81}} = \frac{1{,}8}{9} = 0{,}2$, d'où l'intervalle $[1{,}8\,;2{,}2]$. On a $1{,}8 \leqslant 2{,}15 \leqslant 2{,}2$: m est dans l'intervalle.

Ex. 27 ★★☆ **Le stand de la kermesse**

(a) $\mu = 2 \times 0{,}7 + 4 \times 0{,}2 + 10 \times 0{,}1 = 1{,}4 + 0{,}8 + 1 = 3{,}2$ €.

(b) $V(X) = 2^2 \times 0{,}7 + 4^2 \times 0{,}2 + 10^2 \times 0{,}1 - 3{,}2^2 = 16 - 10{,}24 = 5{,}76$.
Donc $\sigma = \sqrt{5{,}76} = 2{,}4$ €.

(c) Ici $n = 100$: $\frac{2\times2{,}4}{\sqrt{100}} = \frac{4{,}8}{10} = 0{,}48$. L'intervalle est $[3{,}2 - 0{,}48\,;3{,}2 + 0{,}48] = [2{,}72\,;3{,}68]$.

(d) $3{,}75 > 3{,}68$: la moyenne observée sort de l'intervalle. Dans environ 95 % des échantillons de 100 tickets, la valeur moyenne des lots est comprise entre 2,72 € et 3,68 €. Une moyenne de 3,75 € est donc surprenante pour ce modèle.

Ex. 28 ★★☆ **Le programme de Malo**

(a) La ligne 6 donne 1 avec probabilité 0,36 : elle simule une variable de Bernoulli de paramètre $p = 0{,}36$, et `s` compte les lycéens à vélo. Ici $n = 144$ (lycéens par échantillon) et $N = 1\,000$ (nombre d'échantillons).

Ex. 28 **(suite)**

(b) $\sigma = \sqrt{0{,}36 \times 0{,}64} = \sqrt{0{,}2304} = 0{,}48$.
Donc $\frac{2\times 0{,}48}{\sqrt{144}} = \frac{0{,}96}{12} = 0{,}08$.

(c) Malo a utilisé la variance 0,2304 au lieu de l'écart type 0,48. Ligne corrigée : `if abs(s / 144 - 0.36) <= 2 * 0.48 / 12:`. Le programme compte alors les fréquences dans $[0{,}28\,;0{,}44]$.

(d) Avec l'erreur, la demi-largeur vaut $\frac{2\times 0{,}2304}{12} = 0{,}0384$, ce qui est plus petit que 0,08. L'intervalle est plus étroit, il contient moins de fréquences : la proportion affichée serait nettement plus petite que 0,95.

(e) Chaque exécution tire 1 000 nouveaux échantillons au hasard : la proportion fluctue elle aussi. 0,951 signifie que, pour 951 échantillons de 144 lycéens sur 1 000, la fréquence observée vérifie $|f - 0{,}36| \leqslant 0{,}08$. C'est proche de 0,95, comme le constat le prévoit.

Ex. 29 ★★☆ **Tablettes de chocolat : deux tailles d'échantillon**

(a) $\frac{2\times 1{,}2}{\sqrt{36}} = \frac{2{,}4}{6} = 0{,}4$, d'où l'intervalle $[99{,}6\,;100{,}4]$ (en g).

(b) $\frac{2\times 1{,}2}{\sqrt{144}} = \frac{2{,}4}{12} = 0{,}2$, d'où l'intervalle $[99{,}8\,;100{,}2]$. On a $144 = 4 \times 36$ et $0{,}2 = \frac{0{,}4}{2}$: la demi-largeur est divisée par $\sqrt{4} = 2$.

(c) Faux. Les 95\,\% portent sur les échantillons : dans environ 95\,\% des échantillons de 36 tablettes, la masse moyenne m est dans $[99{,}6\,;100{,}4]$. Cela ne dit rien de la masse d'une tablette seule, qui varie davantage ($\sigma = 1{,}2$ g).

(d) Faux. En 1re, ce résultat est un constat fait par simulation : on l'observe sur un grand nombre d'échantillons, on ne le démontre pas.

(e) Pour $n = 144$, l'intervalle est $[99{,}8\,;100{,}2]$ et $100{,}3 > 100{,}2$: m sort de l'intervalle. Un tel écart n'arrive que dans environ 5\,\% des échantillons de 144 tablettes : la masse moyenne de 100,3 g est surprenante pour le modèle. Pourtant, elle serait dans l'intervalle trouvé pour $n = 36$.

Ex. 30 ★★★ **Les flèches de Lucie**

(a) $\mu = 0 \times 0{,}5 + 1 \times 0{,}2 + 3 \times 0{,}3 = 1{,}1$ point. $V(X) = 0^2 \times 0{,}5 + 1^2 \times 0{,}2 + 3^2 \times 0{,}3 - 1{,}1^2 = 2{,}9 - 1{,}21 = 1{,}69$, d'où $\sigma = \sqrt{1{,}69} = 1{,}3$ point.

(b) $\sqrt{169} = 13$, donc $\frac{2\sigma}{\sqrt{n}} = \frac{2\times 1{,}3}{13} = 0{,}2$. L'intervalle, centré sur $\mu = 1{,}1$, est $[0{,}9\,;1{,}3]$.

Ex. 30 **(suite)**

(c) $1{,}35 > 1{,}3$: la moyenne est hors de l'intervalle, car $|1{,}35 - 1{,}1| = 0{,}25 > 0{,}2$. Un écart aussi grand n'arrive que dans environ 5\,\% des échantillons de 169 flèches. Cette observation est donc surprenante pour le modèle : Lucie a peut-être progressé.

(d) Théo a utilisé la variance $V = 1{,}69$ au lieu de l'écart type $\sigma = 1{,}3$. Il trouve $\frac{2\times 1{,}69}{13} = 0{,}26$, donc l'intervalle $[0{,}84\,;1{,}36]$, trop large, qui contient 1,35. Avec le bon σ, l'intervalle est $[0{,}9\,;1{,}3]$ et 1,35 en sort : sa conclusion est fausse.

(e) Faux. Les 95\,\% portent sur les moyennes des échantillons, pas sur les flèches. Une flèche rapporte 0, 1 ou 3 points. La seule valeur dans $[0{,}9\,;1{,}3]$ est 1, et $P(X = 1) = 0{,}2$: seulement environ 20\,\% des flèches sont dans cet intervalle.

Ex. 31 ★★★ **Les graines d'Elsa**

(a) X suit une loi de Bernoulli de paramètre $p = 0{,}9$: $\mu = 0{,}9$ et $\sigma = \sqrt{0{,}9 \times 0{,}1} = \sqrt{0{,}09} = 0{,}3$. Attention, σ n'est pas 0,09.
Donc $\frac{2\sigma}{\sqrt{n}} = \frac{2\times 0{,}3}{\sqrt{100}} = \frac{0{,}6}{10} = 0{,}06$.

(b) $|f - 0{,}9| \leqslant 0{,}06 \iff 0{,}9 - 0{,}06 \leqslant f \leqslant 0{,}9 + 0{,}06 \iff 0{,}84 \leqslant f \leqslant 0{,}96$. L'échantillon est compté si sa fréquence est dans $[0{,}84\,;0{,}96]$, l'intervalle centré sur μ.

(c) La ligne `s = 0` est placée avant la boucle des échantillons : s n'est jamais remis à zéro. Dès le 2e échantillon, `s / 100` vaut environ 1,8, puis 2,7, etc., et le test échoue toujours. Seul le 1er échantillon peut être compté : $\frac{1}{200} = 0{,}005$. Correction : écrire `s = 0` dans la boucle `for j`, juste avant `for i in range(100):` (indentée de 4 espaces).

(d) Faux. Augmenter n resserre l'intervalle ($\frac{2\sigma}{\sqrt{n}}$ diminue), mais la proportion d'échantillons comptés reste d'environ 95\,\%. Les sorties fluctuent parce que $\frac{c}{N}$ n'est calculée que sur $N = 200$ échantillons. Il faut augmenter le nombre N d'échantillons (par exemple 2 000) : la proportion se stabilise, sans être exactement 0,95.

(e) Règle des carrés : pour diviser $\frac{2\sigma}{\sqrt{n}}$ par 3, on multiplie n par $3^2 = 9$, donc $n = 900$. Vérification : $\frac{2\times 0{,}3}{\sqrt{900}} = \frac{0{,}6}{30} = 0{,}02 = \frac{0{,}06}{3}$.

Ex. 32 ★★★ **Le temps de réaction de Hugo**

(a) L'intervalle est centré sur $\mu = 250$ et sa demi-largeur vaut $255 - 250 = 5$.
Donc $\frac{2\sigma}{\sqrt{64}} = \frac{2\sigma}{8} = 5$, d'où $2\sigma = 40$ et $\sigma = 20$ ms.

(b)
```
from math import sqrt
def d(n):
    return 2 * 20 / sqrt(n)
```
On obtient `d(256)` $= \frac{40}{16} = 2{,}5$ et `d(1024)` $= \frac{40}{32} = 1{,}25$. Chaque fois que n est multiplié par $4 = 2^2$, la demi-largeur est divisée par 2 : c'est la règle des carrés.

(c) Faux. m fluctue d'un échantillon à l'autre : elle n'est presque jamais exactement égale à 250. Le constat dit seulement que, dans environ $95\,\%$ des échantillons de 256 mesures, $|m - 250| \leqslant 2{,}5$, c'est-à-dire $m \in [247{,}5\,;252{,}5]$ (en ms) : m est proche de μ, pas égale à μ.

(d) Pour $n = 256$: $|253 - 250| = 3 > 2{,}5$. La moyenne sort de $[247{,}5\,;252{,}5]$: c'est surprenant, car cela n'arrive que dans environ $5\,\%$ des échantillons. Pour $n = 64$: $3 \leqslant 5$. La moyenne est dans $[245\,;255]$: c'est compatible avec le modèle. Quand n augmente, l'intervalle se resserre, et un même écart de 3 ms devient surprenant.

(e) $\frac{2\sigma}{\sqrt{500}} = \frac{40}{\sqrt{500}} \approx 1{,}79$ ms. Si l'on arrondit $\sqrt{500}$ à 22 en cours de calcul, on trouve 1,82, ce qui est faux. Dans environ $95\,\%$ des échantillons de 500 mesures, $m \in [248{,}21\,;251{,}79]$ environ (en ms).

5. *Effet de la taille de l'échantillon*

Ex. 33 ★☆☆ **Automatismes : demi-largeur, largeur, intervalle**

(a) $d(36) = \frac{2\times 1{,}5}{\sqrt{36}} = \frac{3}{6} = 0{,}5$.

(b) $d(81) = \frac{2\times 1{,}8}{\sqrt{81}} = \frac{3{,}6}{9} = 0{,}4$, donc la largeur vaut $2 \times 0{,}4 = 0{,}8$. Attention : 0,4 est la demi-largeur, pas la largeur.

(c) $\sigma = \sqrt{0{,}1 \times 0{,}9} = \sqrt{0{,}09} = 0{,}3$, puis $d(900) = \frac{2\times 0{,}3}{\sqrt{900}} = \frac{0{,}6}{30} = 0{,}02$.

(d) $d(400) = \frac{2\times 0{,}8}{\sqrt{400}} = \frac{1{,}6}{20} = 0{,}08$.

(e) $d(16) = \frac{2\times 4}{\sqrt{16}} = \frac{8}{4} = 2$, donc l'intervalle est $[48\,;52]$.

(f) $d(16) = \frac{8}{4} = 2$ encore, et l'intervalle est $[78\,;82]$. La demi-largeur ne change pas : elle ne dépend pas de μ, l'intervalle est seulement décalé.

Ex. 34 ★★☆ **Automatismes : la règle des carrés**

(a) $400 = 4 \times 100 = 2^2 \times 100$, donc $d(400) = \frac{0{,}6}{2} = 0{,}3$ (et pas $\frac{0{,}6}{4}$).

(b) $5\,000 = 100 \times 50 = 10^2 \times 50$, donc $d(5\,000) = \frac{0{,}9}{10} = 0{,}09$.

(c) $9n = 3^2 n$, donc $d(9n) = \frac{0{,}42}{3} = 0{,}14$.

(d) $d(2n) = \frac{2\sigma}{\sqrt{2}\sqrt{n}} = \frac{0{,}2}{\sqrt{2}} \approx 0{,}141$: doubler n ne divise pas d par 2.

(e) Pour diviser d par 5, on multiplie n par $5^2 = 25$.

(f) $1\,600 = 4^2 \times 100$, donc $d(1\,600) = \frac{d(100)}{4}$, d'où $d(100) = 4 \times 0{,}05 = 0{,}2$.

Ex. 35 ★★☆ **Vrai ou faux : les trajets d'Inès**

(a) Faux. $d(2n) = \frac{2\sigma}{\sqrt{2}\sqrt{n}} = \frac{d(n)}{\sqrt{2}}$: d n'est divisée que par $\sqrt{2} \approx 1{,}41$. Pour la diviser par 2, il faut 4 fois plus de trajets.

(b) Faux. $d(n) = \frac{2\sigma}{\sqrt{n}}$ ne dépend pas de μ : l'intervalle garde la même largeur et il est seulement décalé.

(c) Faux. $d(36) = \frac{2\times 2{,}4}{\sqrt{36}} = \frac{4{,}8}{6} = 0{,}8$: c'est la demi-largeur. La largeur vaut $2 \times 0{,}8 = 1{,}6$.

(d) Vrai. $d(36) = \frac{2\times 4{,}2}{6} = \frac{8{,}4}{6} = 1{,}4 > 0{,}8$: un σ plus grand donne un intervalle plus large, ici $\frac{4{,}2}{2{,}4} = 1{,}75$ fois plus large.

Ex. 36 ★★☆ **Contrôle des bouchons : un tableau de d(n)**

(a) Ligne 5 : `d = 2 * sigma / sqrt(n)` (avec la racine de n, pas n).

(b) Le programme affiche $n = 25, 100, 225, 400$, et donc $d = 0{,}2\,;\,0{,}1\,;\,0{,}0667\,;\,0{,}05$ (car $\frac{1}{5}, \frac{1}{10}, \frac{1}{15}, \frac{1}{20}$).

(c) d est divisée par 2 ($0{,}2 \to 0{,}1$), pas par 4 : $100 = 2^2 \times 25$, donc $d(100) = \frac{d(25)}{2}$.

(d) Ligne 2 : `sigma = 0.25`. d est proportionnelle à σ : diviser σ par 2 divise chaque $d(n)$ par 2. Le programme affiche $d = 0{,}1\,;\,0{,}05\,;\,0{,}0333\,;\,0{,}025$. Un σ plus petit donne un intervalle plus étroit.

Ex. 37 ★★☆ **Deux sondages sur une piste cyclable**

(a) X suit une loi de Bernoulli de paramètre 0,5 : $\sigma = \sqrt{0{,}5 \times 0{,}5} = \sqrt{0{,}25} = 0{,}5$.

(b) A : $d(400) = \frac{2\times 0{,}5}{20} = 0{,}05$. B : $d(2\,500) = \frac{2\times 0{,}5}{50} = 0{,}02$. Dans environ $95\,\%$ des échantillons, la fréquence observée est à au plus 0,05 de p pour A, et à au plus 0,02 de p pour B.

Ex. 37 (suite)

(c) Non. $\frac{2\,500}{400} = 6{,}25 = 2{,}5^2$, donc d n'est divisée que par $\sqrt{6{,}25} = 2{,}5$: on a bien $\frac{0{,}05}{0{,}02} = 2{,}5$.

(d) $f = \frac{1\,310}{2\,500} = 0{,}524$ et $|f - p| = 0{,}024 > 0{,}02 = d(2\,500)$: f est hors de $[0{,}48\,;0{,}52]$. Dans environ 95\,% des échantillons de taille 2 500, f est à au plus 0,02 de p : l'observation est surprenante pour le modèle. Pour A, un écart de $0{,}024 \leqslant 0{,}05$ serait banal.

Ex. 38 ★★★ Les pommes de trois vergers

(a) Verger A : $d(64) = \frac{2\times 12}{\sqrt{64}} = \frac{24}{8} = 3$. Verger B : $d(64) = \frac{2\times 12}{8} = 3$. Les deux valeurs sont égales parce que $d(n)$ dépend seulement de σ et de n, pas de l'espérance μ. Ici, les deux vergers ont le même $\sigma = 12$ et la même taille $n = 64$.

(b) Verger A : $[150 - 3\,;150 + 3] = [147\,;153]$. Verger B : $[210 - 3\,;210 + 3] = [207\,;213]$. Le nombre 3 est la demi-largeur $d(64)$. La largeur vaut $2d(64) = 6$ g. L'élève a confondu la largeur et la demi-largeur.

(c) Verger C : $d(64) = \frac{2\times 18}{8} = \frac{36}{8} = 4{,}5$. C'est plus que 3 : l'intervalle $[145{,}5\,;154{,}5]$ est plus large que celui du verger A. Les masses du verger C sont plus dispersées (σ plus grand). Avec la même taille d'échantillon, les moyennes fluctuent donc davantage et l'intervalle est plus large.

(d) $d(144) = \frac{2\times 18}{\sqrt{144}} = \frac{36}{12} = 3$. C'est la même valeur que pour le verger A avec $n = 64$. On a $\frac{18}{12} = 1{,}5$ et $\frac{144}{64} = 2{,}25 = 1{,}5^2$. Quand σ est multiplié par 1,5, il faut multiplier n par $1{,}5^2$: ainsi $\sqrt{n}$ est multiplié par 1,5 et d reste égale à 3.

(e) Pour le verger A, l'intervalle est $[147\,;153]$. Or $154{,}2 > 153$: $|m - \mu| = 4{,}2 > 3$, donc m est hors de l'intervalle. Dans environ 95\,% des échantillons de 64 pommes, la moyenne tombe dans $[147\,;153]$. Cette observation fait donc partie des environ 5\,% de cas rares : elle est surprenante pour le modèle $\mu = 150$, $\sigma = 12$.

Ex. 39 ★★★ Gauchers : démontrer puis exploiter la règle des carrés

(a) X suit une loi de Bernoulli de paramètre $p = 0{,}1$. On a donc $\mu = p = 0{,}1$ et $\sigma = \sqrt{p(1-p)} = \sqrt{0{,}1 \times 0{,}9} = \sqrt{0{,}09} = 0{,}3$.

(b) Soit $k \geqslant 1$. On a $\sqrt{k^2 n} = \sqrt{k^2} \times \sqrt{n} = k\sqrt{n}$, car $k > 0$. Donc $d(k^2 n) = \frac{2\sigma}{k\sqrt{n}} = \frac{1}{k} \times \frac{2\sigma}{\sqrt{n}} = \frac{d(n)}{k}$.

Ex. 39 (suite)

(c) $d(25) = \frac{2\times 0{,}3}{\sqrt{25}} = \frac{0{,}6}{5} = 0{,}12$. Ensuite : $100 = 2^2 \times 25$, donc $d(100) = \frac{0{,}12}{2} = 0{,}06$. Puis $2\,500 = 10^2 \times 25$, donc $d(2\,500) = \frac{0{,}12}{10} = 0{,}012$. Enfin $10\,000 = 20^2 \times 25$, donc $d(10\,000) = \frac{0{,}12}{20} = 0{,}006$.

(d) Non, Noé a tort. On a $10\,000 = 100 \times 100 = 10^2 \times 100$. La règle des carrés, avec $k = 10$, donne $d(10\,000) = \frac{d(100)}{10}$. Le tableau le confirme : 0,06 devient 0,006. L'intervalle est donc 10 fois plus étroit, et non 100 fois. La précision progresse lentement.

(e) Pour Ben : $d(400) = \frac{0{,}6}{20} = 0{,}03$, donc l'intervalle est $[0{,}07\,;0{,}13]$. Comme $|0{,}125 - 0{,}1| = 0{,}025 \leqslant 0{,}03$, f est dans l'intervalle. Pour Maëlle : $d(1\,600) = \frac{0{,}6}{40} = 0{,}015$, donc l'intervalle est $[0{,}085\,;0{,}115]$. Comme $0{,}025 > 0{,}015$, f est hors de l'intervalle. Pour Ben, cette fréquence est banale : c'est le cas d'environ 95\,% des échantillons de taille 400. Pour Maëlle, la même fréquence est surprenante, car son échantillon, plus grand, a un intervalle plus étroit.

Ex. 40 ★★★ L'attente à la boulangerie : un programme à corriger

(a) $\mu = 1 \times 0{,}18 + 3 \times 0{,}64 + 5 \times 0{,}18 = 0{,}18 + 1{,}92 + 0{,}9 = 3$. Ensuite, $V(X) = (1-3)^2 \times 0{,}18 + (3-3)^2 \times 0{,}64 + (5-3)^2 \times 0{,}18 = 0{,}72 + 0 + 0{,}72 = 1{,}44$. Donc $\sigma = \sqrt{1{,}44} = 1{,}2$.

(b) Le programme calcule $\frac{2\times 1{,}2}{25} = 0{,}096$: il divise par n au lieu de diviser par $\sqrt{n}$. Ligne 3 corrigée : `return 2*sigma/sqrt(n)`. On obtient alors $d(25) = \frac{2{,}4}{\sqrt{25}} = \frac{2{,}4}{5} = 0{,}48$.

(c) On a $100 = 2^2 \times 25$, donc $d(100) = \frac{0{,}48}{2} = 0{,}24$. On a $400 = 4^2 \times 25$, donc $d(400) = \frac{0{,}48}{4} = 0{,}12$. Avec n au lieu de $\sqrt{n}$, multiplier n par 4 divisait d par 4 : on obtenait 0,096, puis 0,024, puis 0,006. En réalité, multiplier n par 4 divise d seulement par 2.

(d) $N = 1\,000$ est le nombre d'échantillons simulés. n (25, 100 ou 400) est la taille de chaque échantillon. Les trois proportions sont proches de 0,95 et fluctuent un peu autour de cette valeur. Elles n'augmentent pas avec n : quelle que soit la taille, environ 95\,% des échantillons vérifient $|m - \mu| \leqslant d(n)$. Ce qui change avec n, c'est que l'intervalle se resserre.

Ex. 40 **(suite)**

(e) Pour $n = 400$, l'intervalle est $[3 - 0{,}12\,;3 + 0{,}12] = [2{,}88\,;3{,}12]$ et sa largeur vaut $2 \times 0{,}12 = 0{,}24$ min. L'affirmation de Lise est fausse : m tombe dans cet intervalle pour environ 95\,% des échantillons de taille 400 seulement, pas toujours. Dans environ 5\,% des cas, m est en dehors. La simulation le confirme : $0{,}951 < 1$.

6. *Proportion des cas sur N échantillons*

Ex. 41 ★☆☆ **Pots de confiture : le test en nombres**

(a) $\dfrac{2\sigma}{\sqrt{n}} = \dfrac{2\times 6}{\sqrt{36}} = \dfrac{12}{6} = 2$ g.

(b) $[250 - 2\,;250 + 2] = [248\,;252]$: environ 95\,\,% des échantillons de 36 pots ont une masse moyenne dans cet intervalle.

(c) $\dfrac{2\sigma}{\sqrt{n}} = \dfrac{12}{\sqrt{400}} = \dfrac{12}{20} = 0{,}6$. Le test s'écrit $|m - 250| \leqslant 0{,}6$, c'est-à-dire $249{,}4 \leqslant m \leqslant 250{,}6$.

(d) Il obtient $\frac{12}{36} \approx 0{,}33$ au lieu de 2 : il a divisé par n au lieu de $\sqrt{n}$. La bonne ligne est `e = 2 * sigma / sqrt(n)`.

(e) $900 = 25 \times 36 = 5^2 \times 36$: la demi-largeur est divisée par 5. On passe de 2 à $\frac{12}{\sqrt{900}} = \frac{12}{30} = 0{,}4$.

(f) $|251{,}3 - 250| = 1{,}3$ et $1{,}3 \leqslant 2$. Oui, l'échantillon passe le test : $251{,}3 \in [248\,;252]$.

Ex. 42 ★☆☆ **Péage : lire un appel de proportion(N, n)**

(a) $N = 400$ échantillons, chacun de taille $n = 25$ voitures.

(b) $N \times n = 400 \times 25 = 10\,000$ valeurs de X.

(c) $c = 0{,}9425 \times 400 = 377$ échantillons.

(d) La taille d'un échantillon est $n = 25$, pas $N = 400$: $\frac{2\times 0{,}9}{\sqrt{25}} = \frac{1{,}8}{5} = 0{,}36$.

(e) Faux. c/N est une proportion d'échantillons : 377 échantillons de 25 voitures sur 400 ont une moyenne m qui passe le test. Ce n'est pas une part des voitures.

(f) Faux. Le programme est aléatoire : la proportion renvoyée est elle-même une fréquence. Elle change d'une exécution à l'autre, en restant proche de 0,95.

Ex. 43 ★★☆ **Lancers francs : compléter le programme**

(a) X suit une loi de Bernoulli de paramètre $p = 0{,}36$. Donc $\mu = 0{,}36$ et $\sigma = \sqrt{0{,}36 \times 0{,}64} = \sqrt{0{,}2304} = 0{,}48$.

Ex. 43 **(suite)**

(b) `c = 0` (avant la boucle), `e = 2 * sigma / sqrt(n)` et `return c / N` (après la boucle : on divise par le nombre d'échantillons).

(c) `proportion(1000, 100, 0.36, 0.48)` : $N = 1\,000$ échantillons de taille $n = 100$.

(d) $\frac{2\times 0{,}48}{\sqrt{100}} = 0{,}096$. Le test $|m - 0{,}36| \leqslant 0{,}096$ s'écrit $0{,}264 \leqslant m \leqslant 0{,}456$.

(e) Sur 1 000 séries de 100 lancers, 962 ont une fréquence de réussite entre 0,264 et 0,456. Cette valeur est proche de 95\,\,%, comme prévu. Ce n'est pas la part des lancers réussis.

Ex. 44 ★★☆ **Jeu de grattage : le programme d'Inès**

(a) $\mu = 0 \times 0{,}5 + 1 \times 0{,}2 + 3 \times 0{,}3 = 1{,}1$. $V(X) = 0^2 \times 0{,}5 + 1^2 \times 0{,}2 + 3^2 \times 0{,}3 - 1{,}1^2 = 2{,}9 - 1{,}21 = 1{,}69$. Donc $\sigma = \sqrt{1{,}69} = 1{,}3$.

(b) 1) `c = 0` doit être placé avant la boucle : dans la boucle, le compteur est remis à zéro à chaque échantillon. 2) `e = 2 * sigma / sqrt(n)` et non `/ n`. 3) `return c / N` : on divise par le nombre d'échantillons, pas par leur taille.

(c) $\frac{2\times 1{,}3}{\sqrt{169}} = \frac{2{,}6}{13} = 0{,}2$. Le test s'écrit $|m - 1{,}1| \leqslant 0{,}2$, soit $m \in [0{,}9\,;1{,}3]$.

(d) Non. 0,954 est une proportion d'échantillons : sur 500 échantillons de 169 tickets, $0{,}954 \times 500 = 477$ ont un gain moyen m dans $[0{,}9\,;1{,}3]$. Ce n'est pas une probabilité sur le gain d'un ticket : un ticket rapporte 0, 1 ou 3 €, et $P(0{,}9 \leqslant X \leqslant 1{,}3) = P(X = 1) = 0{,}2$.

Ex. 45 ★★☆ **Graines : interpréter des exécutions**

(a) $\mu = p = 0{,}1$ et $\sigma = \sqrt{0{,}1 \times 0{,}9} = 0{,}3$. Pour $n = 400$: $\frac{2\times 0{,}3}{20} = 0{,}03$. Le test s'écrit $|m - 0{,}1| \leqslant 0{,}03$, soit $0{,}07 \leqslant m \leqslant 0{,}13$.

(b) $c = 0{,}98 \times 50 = 49$ échantillons de 400 graines.

(c) Non. La proportion vaut 1 : les 50 échantillons de 400 graines ont tous passé le test. Le cours annonce environ 95\,\,% des échantillons, pas exactement. Avec seulement 50 échantillons, la proportion renvoyée est une fréquence qui fluctue beaucoup.

Ex. 45 **(suite)**

(d) Faux. Dans les deux lignes, chaque échantillon contient $n = 400$ graines, et le test est le même : $0{,}07 \leqslant m \leqslant 0{,}13$. Seul le nombre N d'échantillons change : la proportion est calculée sur 2 000 échantillons au lieu de 50, donc elle fluctue moins. Le programme tire $2\,000 \times 400 = 800\,000$ valeurs au lieu de $50 \times 400 = 20\,000$.

Ex. 46 ★★★ **Hotline : corriger le programme de Noé**

(a) $\mu = 0 \times 0{,}1 + 1 \times 0{,}6 + 2 \times 0{,}3 = 1{,}2$. $V(X) = 0^2 \times 0{,}1 + 1^2 \times 0{,}6 + 2^2 \times 0{,}3 - 1{,}2^2 = 1{,}8 - 1{,}44 = 0{,}36$, donc $\sigma = \sqrt{0{,}36} = 0{,}6$. `mu = 1.2` est juste. `sigma = 0.36` est faux : c'est la variance. Il faut écrire `sigma = 0.6`.

(b) Chez Noé : $e = \frac{2 \times 0{,}36}{\sqrt{36}} = \frac{0{,}72}{6} = 0{,}12$. Valeur attendue : $\frac{2 \times 0{,}6}{\sqrt{36}} = \frac{1{,}2}{6} = 0{,}2$. Le test de Noé exige $1{,}08 \leqslant m \leqslant 1{,}32$. Cet intervalle est plus étroit que $[1\,;1{,}4]$, qui contient environ 95\,\% des moyennes. Moins d'échantillons passent le test : la proportion est nettement inférieure à 0,95.

(c) Le `return` est dans la boucle : la fonction s'arrête dès le premier échantillon. À ce moment, `c` vaut 1 si cet échantillon passe le test, et 0 sinon. Elle renvoie donc $\frac{1}{500} = 0{,}002$ ou 0. On place `return c / N` au même niveau que le `for`, après la boucle : on ne divise par N qu'une fois les N échantillons comptés.

(d) Le test corrigé s'écrit $|m - 1{,}2| \leqslant 0{,}2$, soit $1 \leqslant m \leqslant 1{,}4$. 0,958 est une proportion d'échantillons, pas de minutes : sur 500 échantillons de 36 minutes, 95,8 % ont un nombre moyen d'appels compris entre 1 et 1,4. Une minute isolée a d'ailleurs 0, 1 ou 2 appels, jamais 1,2.

(e) Le programme tire $N \times n = 500 \times 36 = 18\,000$ valeurs de X. $c = 0{,}958 \times 500 = 479$ échantillons passent le test, donc $500 - 479 = 21$ échantillons ont une moyenne hors de $[1\,;1{,}4]$.

Ex. 47 ★★★ **Projet de tram : une erreur sur mu**

(a) X suit une loi de Bernoulli de paramètre $p = 0{,}5$: $\mu = p = 0{,}5$ et $\sigma = \sqrt{0{,}5 \times 0{,}5} = 0{,}5$. Pour $n = 400$: $\frac{2 \times 0{,}5}{\sqrt{400}} = \frac{1}{20} = 0{,}05$. Le test s'écrit $|m - 0{,}5| \leqslant 0{,}05$, soit $0{,}45 \leqslant m \leqslant 0{,}55$.

Ex. 47 **(suite)**

(b) Non. La valeur renvoyée est elle-même une fréquence calculée sur 1 000 échantillons : elle fluctue d'une exécution à l'autre. Les valeurs 0,957, 0,948 et 0,962 sont proches de 0,95, ce qu'on attend : dans environ 95\,\% des échantillons, la fréquence est entre 0,45 et 0,55.

(c) Le test devient $|m - 0{,}6| \leqslant 0{,}05$, soit $0{,}55 \leqslant m \leqslant 0{,}65$. Or les fréquences simulées fluctuent autour de la vraie valeur 0,5, et environ 95\,\% d'entre elles sont entre 0,45 et 0,55. Très peu atteignent 0,55 : la proportion tombe à environ 0,03, loin de 0,95. La valeur 0,6 ne correspond pas à la situation simulée.

(d) Faux. Les deux appels tirent bien $1\,000 \times 400 = 400 \times 1\,000 = 400\,000$ valeurs. Mais `proportion(400, 1000)` simule 400 échantillons de 1 000 habitants. Le test devient $|m - 0{,}5| \leqslant \frac{2 \times 0{,}5}{\sqrt{1\,000}} \approx 0{,}032$, et la proportion porte sur 400 échantillons seulement. Ce n'est pas la même expérience.

Ex. 48 ★★★ **Téléphérique : de la loi à la proportion**

(a) `random()` renvoie un nombre de $[0\,;1[$. Les seuils 0,18 et 0,82 sont des probabilités cumulées : $P(X = 1) = 0{,}18$, $P(X = 3) = 0{,}82 - 0{,}18 = 0{,}64$, $P(X = 5) = 1 - 0{,}82 = 0{,}18$.

(b) $\mu = 1 \times 0{,}18 + 3 \times 0{,}64 + 5 \times 0{,}18 = 3$. $V(X) = 1^2 \times 0{,}18 + 3^2 \times 0{,}64 + 5^2 \times 0{,}18 - 3^2 = 10{,}44 - 9 = 1{,}44$, donc $\sigma = \sqrt{1{,}44} = 1{,}2$.

(c) $n = 9$: $\frac{2 \times 1{,}2}{\sqrt{9}} = \frac{2{,}4}{3} = 0{,}8$. $n = 36$: $\frac{2{,}4}{6} = 0{,}4$. Comme $144 = 4 \times 36 = 2^2 \times 36$, la règle des carrés divise la demi-largeur par 2 : $\frac{0{,}4}{2} = 0{,}2$.

(d) Faux. Les trois proportions (0,947 ; 0,958 ; 0,952) restent proches de 0,95 quel que soit n. Quand n augmente, c'est l'intervalle qui se resserre (0,8, puis 0,4, puis 0,2 de demi-largeur) : les moyennes sont plus proches de μ, mais environ 95\,\% des échantillons passent toujours le test.

(e) Pour $n = 144$: $[3 - 0{,}2\,;3 + 0{,}2] = [2{,}8\,;3{,}2]$. Or $3{,}35 > 3{,}2$: la moyenne observée est en dehors de l'intervalle. Environ 95\,\% des échantillons simulés de 144 attentes ont une moyenne dans $[2{,}8\,;3{,}2]$, donc cette observation est surprenante pour le modèle.