Plaquette 3 – Écart entre m et μ : la règle des 2σ/√n

Échantillonnage — Première spé maths

Corrections détaillées

Boîte à outils

L'écart de référence. Pour des échantillons de taille n d'une variable d'espérance μ et d'écart type σ :

$$|m-\mu| \leq \frac{2\sigma}{\sqrt{n}} \Leftrightarrow \mu - \frac{2\sigma}{\sqrt{n}} \leq m \leq \mu + \frac{2\sigma}{\sqrt{n}}$$

Ce que montrent les simulations. Cette condition est réalisée dans environ 95 % des échantillons ; un écart plus grand se produit dans environ 5 % des cas : il est **inhabituel**, sans être impossible.

Effet de la taille. $\frac{2\sigma}{\sqrt{n}}$ est divisé par 2 quand n est multiplié par 4, par 10 quand n est multiplié par 100. Pour garantir $\frac{2\sigma}{\sqrt{n}} \leq e$, il faut

$$n \geq \left(\frac{2\sigma}{e}\right)^2$$

Cas d'une fréquence. Pour une variable de Bernoulli de paramètre p, $\sigma = \sqrt{p(1-p)} \leq \frac{1}{2}$, donc $\frac{2\sigma}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}$.

Valeurs utiles. Dé équilibré : $\mu = 3{,}5$, $\sigma \approx 1{,}708$. Somme de deux dés : $\mu = 7$, $\sigma = \sqrt{\frac{35}{6}} \approx 2{,}415$.

Correction de l'exercice 1 – Calculer l'écart de référence

a) On calcule $2 \times 1{,}708 = 3{,}416$, puis on divise par $\sqrt{n}$: $\frac{3{,}416}{5} \approx 0{,}683$ pour $n = 25$, $\frac{3{,}416}{10} \approx 0{,}342$ pour $n = 100$ et $\frac{3{,}416}{20} \approx 0{,}171$ pour $n = 400$.

b) La condition équivaut à

$$3{,}5 - 0{,}342 \leq m \leq 3{,}5 + 0{,}342 \Leftrightarrow 3{,}158 \leq m \leq 3{,}842$$

c) $|3{,}9 - 3{,}5| = 0{,}4 > 0{,}342$: la condition n'est pas réalisée. D'après les simulations, cela n'arrive que dans environ 5 % des échantillons de 100 lancers d'un dé équilibré : c'est un résultat inhabituel, mais possible.

Réponse. a) ≈ 0,683 ; 0,342 ; 0,171 ; b) $3{,}158 \leq m \leq 3{,}842$; c) non (0,4 > 0,342).

Correction de l'exercice 2 – Environ 95 % des échantillons

a) La proportion reste proche de 95 % pour toutes les tailles, alors que l'écart de référence $\frac{2\sigma}{\sqrt{n}}$ diminue nettement quand n augmente. Autrement dit, les moyennes se resserrent autour de μ exactement au rythme de $\frac{1}{\sqrt{n}}$: la règle « dans environ 95 % des cas, $|m-\mu| \leq \frac{2\sigma}{\sqrt{n}}$ » s'applique à chaque taille.

b) Chaque élève a environ 5 % de chances d'obtenir une moyenne hors de l'intervalle. Sur 40 élèves, on peut s'attendre à environ

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).

$$40 \times 0{,}05 = 2$$

élèves dans ce cas. Ce n'est pas que leur simulation est fausse : c'est la part normale des échantillons « extrêmes ».

Réponse. a) Proportion toujours proche de 95 %, écart de référence qui diminue ; b) environ 2 élèves.

Correction de l'exercice 3 – Le cas d'une fréquence

a) On a $\sigma = \sqrt{0{,}3 \times 0{,}7} = \sqrt{0{,}21} \approx 0{,}458$, d'où

$$\frac{2\sigma}{\sqrt{100}} \approx \frac{0{,}917}{10} \approx 0{,}092$$

b) La condition s'écrit $0{,}3 - 0{,}092 \leq f \leq 0{,}3 + 0{,}092$, soit $0{,}208 \leq f \leq 0{,}392$. Sur 100 personnes, cela correspond à un nombre de personnes ayant le caractère compris entre 21 et 39.

c) La proportion observée est proche de 95 %, comme pour les moyennes de lancers de dé. Elle est un peu plus élevée car les fréquences ne prennent que des valeurs « en centièmes » : les effectifs 21 et 39, à la limite, sont comptés dedans. La règle des $\frac{2\sigma}{\sqrt{n}}$ s'applique donc aussi aux fréquences, qui sont des moyennes de variables de Bernoulli.

Réponse. a) $\approx 0{,}092$; b) $0{,}208 \leq f \leq 0{,}392$, de 21 à 39 personnes ; c) proche de 95 %.

Correction de l'exercice 4 – La somme de deux dés

a) Par linéarité, $E(S) = 3{,}5 + 3{,}5 = 7$, et $\sigma(S) = \sqrt{\frac{35}{6}} \approx 2{,}415$.

b) $\sqrt{50} \approx 7{,}071$, donc

$$\frac{2\sigma}{\sqrt{50}} \approx \frac{4{,}830}{7{,}071} \approx 0{,}683$$

L'intervalle est $[7 - 0{,}683\,;\, 7 + 0{,}683] = [6{,}317\,;\, 7{,}683]$.

c) La proportion d'échantillons dans l'intervalle, 94,4 %, est proche de 95 %, comme prévu. Les moyennes extrêmes 6,02 et 7,98 sont en dehors : elles font partie des quelque 5 % d'échantillons inhabituels, ici à près de 1 point de l'espérance, soit environ 2,9 écarts types $\frac{\sigma}{\sqrt{n}}$.

Réponse. a) $E(S) = 7$, $\sigma \approx 2{,}415$; b) $\approx 0{,}683$, intervalle $[6{,}317\,;\, 7{,}683]$; c) proche de 95 %, les extrêmes sont des cas inhabituels.

Correction de l'exercice 5 – Un jeu sur cent parties

a) $\frac{2 \times 3{,}666}{10} \approx 0{,}733$. Le gain moyen par partie est donc, dans environ 95 % des cas, compris entre $0{,}4 - 0{,}733 = -0{,}333$ € et $0{,}4 + 0{,}733 = 1{,}133$ €.

b) Oui : l'intervalle contient des valeurs négatives. Même en restant dans les 95 % de cas « habituels », le joueur peut finir en perte, par exemple avec un gain moyen de $-0{,}2$ € par partie, soit une perte de 20 € sur 100 parties.

c) Il faut $0{,}4 - \frac{2\sigma}{\sqrt{n}} \geq 0$, soit $\frac{7{,}332}{\sqrt{n}} \leq 0{,}4$, c'est-à-dire

$$\sqrt{n} \geq \frac{7{,}332}{0{,}4} = 18{,}33 \Leftrightarrow n \geq 336$$

Au-delà d'environ 336 parties, le gain moyen est positif dans au moins 95 % des cas.

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).

Réponse. a) ≈ 0,733 ; intervalle $[-0{,}333\,;\,1{,}133]$; b) oui ; c) environ 336 parties.

Correction de l’exercice 6 – Programmer le calcul de la proportion

a) L contient les moyennes de N échantillons de taille n. *bons* est une liste en compréhension qui ne garde que les moyennes m vérifiant $|m-\mu| \le \dfrac{2\sigma}{\sqrt{n}}$.

b) La fonction renvoie la proportion d’échantillons de 25 lancers dont la moyenne est à moins de $\dfrac{2\sigma}{\sqrt{25}} \approx 0{,}683$ de 3,5, sur 1 000 échantillons simulés. D’après les simulations, on attend une valeur proche de 0,95.

c) On remplace la variable simulée et ses paramètres :

```
from random import random
```

```
from math import sqrt

def frequence(n, p):
    return sum([1 for i in range(n) if random() < p]) / n

def proportion(N, n, p):
    e = 2 * sqrt(p * (1 - p)) / sqrt(n)
    L = [frequence(n, p) for k in range(N)]
    return len([f for f in L if abs(f - p) <= e]) / N
```

La somme des « 1 » de la première liste compte les succès, et la fonction renvoie la proportion de fréquences proches de p.

Réponse. a) Les N moyennes, puis celles qui sont à moins de $\dfrac{2\sigma}{\sqrt{n}}$ de μ ; b) une proportion proche de 0,95 ; c) $\mu = p$ et $\sigma = \sqrt{p(1-p)}$.

Correction de l’exercice 7 – Choisir la taille d’un échantillon

a) On résout $\dfrac{3{,}416}{\sqrt{n}} \le 0{,}1$, soit $\sqrt{n} \ge 34{,}16$, donc $n \ge 34{,}16^2 \approx 1\,166{,}9$: il faut au moins 1 167 lancers.

b) Comme $\sigma = \sqrt{p(1-p)} \le \dfrac{1}{2}$, on a $\dfrac{2\sigma}{\sqrt{n}} \le \dfrac{1}{\sqrt{n}}$. Il suffit donc que

$$\frac{1}{\sqrt{n}} \le 0{,}02 \Leftrightarrow \sqrt{n} \ge 50 \Leftrightarrow n \ge 2\,500$$

c) Avec $n = 2\,500$, $\dfrac{1}{\sqrt{n}} = \dfrac{1}{50} = 0{,}02$: l’institut peut garantir que, dans environ 95 % des cas, sa fréquence est à moins de 2 points de la proportion réelle, quelle qu’elle soit.

Réponse. a) $n \ge 1\,167$; b) $n \ge 2\,500$; c) 0,02, soit 2 points.

Correction de l’exercice 8 – Démontrer une majoration

a) $g(p) = -p^2 + p$ est dérivable et $g'(p) = -2p + 1$, positive sur $\left[0\,;\,\dfrac{1}{2}\right]$ et négative sur $\left[\dfrac{1}{2}\,;\,1\right]$. La fonction g est croissante puis décroissante, et son maximum est

$$g\left(\frac{1}{2}\right) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

b) Pour tout $p \in [0\,;1]$, $0 \le p(1-p) \le \dfrac{1}{4}$; la fonction racine carrée étant croissante, $\sqrt{p(1-p)} \le \dfrac{1}{2}$, donc $2\sqrt{p(1-p)} \le 1$. En divisant par $\sqrt{n} > 0$, on obtient l’inégalité voulue.

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) Pour $p = 0{,}1$, $\sigma = \sqrt{0{,}09} = 0{,}3$, d'où $\dfrac{2\sigma}{\sqrt{400}} = \dfrac{0{,}6}{20} = 0{,}03$, alors que $\dfrac{1}{\sqrt{400}} = 0{,}05$. La majoration est d'autant plus grossière que p est éloigné de $\dfrac{1}{2}$ (proche de 0 ou de 1) ; elle est exacte pour $p = \dfrac{1}{2}$.

Réponse. a) Croissante puis décroissante, maximum $\dfrac{1}{4}$ en $\dfrac{1}{2}$; b) $2\sqrt{p(1-p)} \leq 1$; c) 0,03 contre 0,05 ; grossière quand p est loin de $\dfrac{1}{2}$.

Correction de l'exercice 9 – Le réglage d'une machine

a) $\dfrac{2 \times 4}{4} = 2$ mL. Pour une machine bien réglée, dans environ 95 % des prélèvements,

$$998 \leq m \leq 1\,002$$

b) Les moyennes 1 001,3 et 998,6 sont dans l'intervalle : rien d'inhabituel. Les moyennes 997,5 et 997,1 sont à 2,5 et 2,9 mL de 1 000, au-delà de l'écart de référence 2 mL.

c) Un prélèvement hors de l'intervalle arrive dans environ 5 % des cas avec une machine bien réglée ; deux de suite, du même côté et après une baisse régulière, c'est très inhabituel. Il est raisonnable de vérifier le réglage : la machine semble verser trop peu.

Réponse. a) 2 mL ; [998 ; 1 002] ; b) les deux premiers ; c) oui, deux prélèvements consécutifs hors de l'intervalle, en baisse.

Correction de l'exercice 10 – Vrai ou faux sur la règle des 2σ/√n

a) **Faux.** La condition n'est réalisée que dans environ 95 % des échantillons ; les autres, environ 5 %, s'écartent davantage.

b) **Faux.** $\dfrac{2\sigma}{\sqrt{4n}} = \dfrac{2\sigma}{2\sqrt{n}}$: l'écart de référence est divisé par 2, pas par 4.

c) **Vrai.** L'écart de référence $\dfrac{2\sigma}{\sqrt{n}}$ est proportionnel à σ.

d) **Vrai.** Environ 5 % des échantillons sont hors de l'intervalle, soit

$$1\,000 \times 0{,}05 = 50$$

échantillons, à quelques unités près.

Réponse. a) faux ; b) faux ; c) vrai ; d) vrai.

Correction de l'exercice 11 – Un sondage serré

a) $\sigma = \sqrt{0{,}52 \times 0{,}48} = \sqrt{0{,}2496} \approx 0{,}4996$, d'où

$$\frac{2\sigma}{\sqrt{1\,000}} \approx \frac{0{,}9992}{31{,}62} \approx 0{,}032$$

b) Dans environ 95 % des sondages, la fréquence est entre $0{,}52 - 0{,}032 = 0{,}488$ et $0{,}52 + 0{,}032 = 0{,}552$.

c) Oui : l'intervalle contient des valeurs inférieures à 0,5, et même parmi les sondages « habituels », certains donneront A sous les 50 %. Pour que l'intervalle reste au-dessus de 0,5, il faudrait $\dfrac{2\sigma}{\sqrt{n}} \leq 0{,}02$, soit $\sqrt{n} \geq \dfrac{0{,}9992}{0{,}02} \approx 50$, c'est-à-dire environ 2 500 électeurs interrogés.

Réponse. a) ≈ 0,032 ; b) entre 0,488 et 0,552 ; c) oui ; il faudrait environ 2 500 électeurs.

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 – Un histogramme de moyennes

a) Les classes de $[2{,}9\,;3{,}1[$ à $[3{,}9\,;4{,}1[$ totalisent

$$77 + 171 + 220 + 216 + 163 + 70 = 917$$

moyennes, soit 91,7 % des 1 000 échantillons.

b) L’intervalle $[2{,}9\,;4{,}1[$ s’écarte de 0,6 de chaque côté de 3,5, alors que l’intervalle de référence s’écarte de 0,683 : il va de 2,817 à 4,183. Il est plus large et contient donc plus de moyennes ; d’ailleurs, la simulation compte 952 moyennes dans l’intervalle de référence, soit 95,2 %. L’histogramme, « en cloche » centrée sur 3,5, montre que les moyennes se raréfient vite quand on s’éloigne de l’espérance.

Réponse. a) 91,7 % ; b) $[2{,}9\,;4{,}1[$ est plus étroit que l’intervalle de référence $[2{,}817\,;4{,}183]$.

Correction de l’exercice 13 – L’effet de l’écart type

a) Machine A : $\dfrac{2\times 2}{6} \approx 0{,}67$ g. Machine B : $\dfrac{2\times 6}{6} = 2$ g.

b) L’écart avec 250 g vaut 2 g. Pour la machine B, il est égal à l’écart de référence : la moyenne est à la limite de l’intervalle $[248\,;252]$, elle reste dans les 95 % de cas habituels. Pour la machine A, 2 g représente trois fois l’écart de référence : un tel écart serait très inhabituel et devrait alerter.

c) On veut $\dfrac{12}{\sqrt{n}} = \dfrac{4}{6}$, soit $\sqrt{n} = 18$, donc

$$n = 18^2 = 324$$

sachets : l’écart type étant 3 fois plus grand, il faut un échantillon $3^2 = 9$ fois plus grand.

Réponse. a) ≈ 0,67 g et 2 g ; b) habituel pour B (à la limite), très inhabituel pour A ; c) 324 sachets.

Correction de l’exercice 14 – Une expérience en classe

a) $E(S) = 7$ et $\sigma(S) = \sqrt{\dfrac{35}{6}} \approx 2{,}415$, d’où $\dfrac{2\times 2{,}415}{\sqrt{50}} \approx 0{,}683$.

b) Chaque moyenne a environ 5 % de chances d’être hors de l’intervalle : on peut s’attendre à environ $35 \times 0{,}05 = 1{,}75$, soit une ou deux moyennes hors de l’intervalle, ce qui n’a rien d’anormal.

c) Pour $n = 1\,750$, $\sqrt{n} \approx 41{,}83$ et

$$\frac{2\sigma}{\sqrt{1\,750}} \approx \frac{4{,}83}{41{,}83} \approx 0{,}115$$

La moyenne de la classe sera donc, dans environ 95 % des cas, entre 6,885 et 7,115 : regrouper les lancers rend l’estimation de $E(S)$ beaucoup plus précise.

Réponse. a) 7 ; ≈ 2,415 ; ≈ 0,683 ; b) environ 1 ou 2 ; c) ≈ 0,115.

Correction de l’exercice 15 – Synthèse : le contrôle d’un lot de piles

a) $\dfrac{2\times 6}{\sqrt{25}} = \dfrac{12}{5} = 2{,}4$ heures. Si l’annonce est exacte, dans environ 95 % des contrôles, $37{,}6 \leq m \leq 42{,}4$.

b) $m = 37{,}1$ est en dehors de cet intervalle : l’écart $|37{,}1 - 40| = 2{,}9$ dépasse l’écart de référence 2,4. C’est un résultat inhabituel (il n’arrive que dans environ 5 % des contrôles si l’annonce est exacte), qui invite à approfondir le test.

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) Avec la probabilité $\frac{1}{2}$ chacune :

```
from random import random

def duree():
    if random() < 0.5:
        return 34
    return 46
```

$E = \frac{34 + 46}{2} = 40$, et les écarts à 40 valent ± 6, d'où $V = 36$ et $\sigma = 6$.

d) On reprend la structure des simulations précédentes :

```
from math import sqrt

def test(N, n):
    e = 2 * 6 / sqrt(n)
    L = [sum([duree() for i in range(n)]) / n for k in range(N)]
    return len([m for m in L if abs(m - 40) <= e]) / N
```

e) Pour $n = 100$, l'écart de référence vaut $\frac{12}{10} = 1{,}2$ heure. L'écart observé, $|39{,}2 - 40| = 0{,}8$, est inférieur : ce résultat est tout à fait habituel si l'annonce est exacte. Le premier contrôle faisait sans doute partie des 5 % de cas inhabituels ; le second, fondé sur un échantillon plus grand, ne donne aucune raison de douter du fabricant.

Réponse. a) 2,4 h, $[37{,}6\,;42{,}4]$; b) inhabituel ($2{,}9 > 2{,}4$) ; c) 34 ou 46 avec $\frac{1}{2}$: $E = 40$, $\sigma = 6$; d) proportion des $|m - 40| \leq \frac{12}{\sqrt{n}}$; e) écart $0{,}8 < 1{,}2$: compatible avec l'annonce.

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).