Plaquette 1 – Simuler une variable aléatoire, moyenne d'un échantillon

Échantillonnage — Première spé maths

Corrections détaillées

Boîte à outils

Espérance, variance, écart type d'une variable X prenant les valeurs $x_1, \dots, x_k$ avec les probabilités $p_1, \dots, p_k$:

$$\mu = E(X) = \sum_{i=1}^{k} p_i x_i \qquad V(X) = \sum_{i=1}^{k} p_i (x_i - \mu)^2 = E(X^2) - \mu^2 \qquad \sigma = \sqrt{V(X)}$$

Linéarité. $E(aX + b) = aE(X) + b$ et $E(X + Y) = E(X) + E(Y)$.

Variable de Bernoulli de paramètre p (valeur 1 en cas de succès, 0 sinon) : $E(X) = p$, $V(X) = p(1 - p)$.

Moyenne d'un échantillon $(x_1, \dots, x_n)$: $m = \dfrac{x_1 + \dots + x_n}{n}$. Pour une variable de Bernoulli, m est la **fréquence** des succès.

Simulation en Python. *random()* renvoie un décimal de $[0\,;1[$; la condition *random() < p* est vraie avec la probabilité p ; *randint(a, b)* renvoie un entier compris entre a et b, bornes incluses. Une liste en compréhension *[X() for i in range(n)]* fournit un échantillon de taille n, et *sum(ech) / len(ech)* sa moyenne.

Ce que l'on observe. Les moyennes de plusieurs échantillons de même taille diffèrent (fluctuation d'échantillonnage) ; plus n est grand, plus elles sont proches de μ.

Correction de l'exercice 1 – Espérance et écart type d'un dé

a) Les six faces sont équiprobables : $P(X = k) = \dfrac{1}{6}$ pour $k \in \{1, \dots, 6\}$. Donc

$$\mu = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3{,}5$$

b) $E(X^2) = \dfrac{1 + 4 + 9 + 16 + 25 + 36}{6} = \dfrac{91}{6}$. Par la formule de König-Huygens,

$$V(X) = E(X^2) - \mu^2 = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12}$$

et $\sigma = \sqrt{\dfrac{35}{12}} \approx 1{,}708$.

c) Sur un grand nombre de lancers, la moyenne des résultats obtenus est proche de 3,5 : c'est le résultat « moyen » d'un lancer, même si 3,5 n'est jamais obtenu.

Réponse. a) Loi uniforme sur $\{1, \dots, 6\}$, $\mu = 3{,}5$; b) $E(X^2) = \dfrac{91}{6}$, $V = \dfrac{35}{12}$, $\sigma \approx 1{,}708$; c) moyenne des résultats sur un grand nombre de lancers.

Correction de l'exercice 2 – Variable de Bernoulli

a) $P(X = 1) = 0{,}7$ et $P(X = 0) = 0{,}3$. Donc $E(X) = 0 \times 0{,}3 + 1 \times 0{,}7 = 0{,}7$ et, comme $X^2 = X$, $E(X^2) = 0{,}7$, d'où $V(X) = 0{,}7 - 0{,}7^2 = 0{,}21$.

b) Dans le cas général, $E(X) = 0 \times (1 - p) + 1 \times p = p$. Comme X ne prend que les valeurs 0 et 1, $X^2 = X$, donc $E(X^2) = p$ et

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).

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

c) La somme des valeurs vaut 7, donc $m = \dfrac{7}{10} = 0{,}7$. La moyenne d'un échantillon de variables de Bernoulli est la **fréquence** des succès : ici, 7 lancers réussis sur 10. Elle tombe exactement sur p, mais c'est une coïncidence : un autre échantillon donnerait 0,6 ou 0,8.

Réponse. a) $E(X) = 0{,}7$, $V(X) = 0{,}21$; b) $X^2 = X$ donne $V = p - p^2$; c) $m = 0{,}7$, la fréquence de réussite.

Correction de l'exercice 3 – Le gain d'un jeu

a) X prend les valeurs 10, 2 et -2 avec les probabilités 0,1, 0,3 et $1 - 0{,}1 - 0{,}3 = 0{,}6$. Donc

$$E(X) = 10 \times 0{,}1 + 2 \times 0{,}3 - 2 \times 0{,}6 = 1 + 0{,}6 - 1{,}2 = 0{,}4$$

b) $E(X^2) = 100 \times 0{,}1 + 4 \times 0{,}3 + 4 \times 0{,}6 = 10 + 1{,}2 + 2{,}4 = 13{,}6$, d'où $V(X) = 13{,}6 - 0{,}4^2 = 13{,}44$ et $\sigma = \sqrt{13{,}44} \approx 3{,}67$.

c) L'espérance est positive : sur un grand nombre de parties, le joueur gagne en moyenne 0,40 € par partie, le jeu lui est favorable. L'écart type, bien plus grand que l'espérance, montre que les gains d'une partie à l'autre sont très dispersés : sur quelques parties, le joueur peut très bien être en perte.

Réponse. a) $E(X) = 0{,}4$; b) $V(X) = 13{,}44$, $\sigma \approx 3{,}67$; c) favorable en moyenne, mais gains très dispersés.

Correction de l'exercice 4 – Simuler le gain du jeu

a) La fonction renvoie 2 lorsque u n'est pas inférieur à 0,1 mais est inférieur à 0,4, c'est-à-dire lorsque $u \in [0{,}1\,;0{,}4[$. Cet intervalle a pour longueur 0,3, donc la probabilité vaut 0,3. De même, $u \in [0\,;0{,}1[$ donne 10 (probabilité 0,1) et $u \in [0{,}4\,;1[$ donne -2 (probabilité 0,6).

b) Pour $u = 0{,}05$, on a $u < 0{,}1$: la fonction renvoie 10. Pour $u = 0{,}4$, ni $u < 0{,}1$ ni $u < 0{,}4$: elle renvoie -2. Pour $u = 0{,}73$, elle renvoie aussi -2.

c) Le nouveau test fait un second tirage, indépendant du premier. Le gain 2 est alors obtenu lorsque le premier tirage est au moins 0,1 **et** le second inférieur à 0,3, avec la probabilité

$$0{,}9 \times 0{,}3 = 0{,}27 \neq 0{,}3$$

On découpe toujours **un seul** tirage en intervalles de longueurs égales aux probabilités.

Réponse. a) $u \in [0{,}1\,;0{,}4[$, de longueur 0,3 ; b) 10 ; -2 ; -2 ; c) deux tirages : probabilité 0,27.

Correction de l'exercice 5 – La somme de deux dés

a) Par linéarité de l'espérance,

$$E(S) = E(X_1) + E(X_2) = 3{,}5 + 3{,}5 = 7$$

b) On simule chaque dé par *randint(1, 6)* et on construit la liste en compréhension :

```
from random import randint

def somme():
    return randint(1, 6) + randint(1, 6)
```

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).

```
def echantillon(n):
    return [somme() for i in range(n)]
```

c) La moyenne vaut $m = \frac{356}{50} = 7{,}12$. Elle est proche de $E(S) = 7$ sans lui être égale : l'écart de 0,12 est dû à la fluctuation d'échantillonnage, et un autre échantillon de 50 lancers donnerait une autre moyenne.

Réponse. a) $E(S) = 7$; b) *somme()* et *[somme() for i in range(n)]* ; c) $m = 7{,}12$, proche de 7.

Correction de l'exercice 6 – Calculer la moyenne d'un échantillon

a) La somme des valeurs vaut $5 + 1 + 4 + 4 + 5 + 1 + 2 + 4 + 4 + 3 = 33$, donc

$$m = \frac{33}{10} = 3{,}3 \qquad |m - \mu| = |3{,}3 - 3{,}5| = 0{,}2$$

b) La face 4 apparaît 4 fois sur 10 : fréquence 0,4, contre une probabilité de $\frac{1}{6} \approx 0{,}167$. Sur 10 lancers, une fréquence peut s'écarter fortement de la probabilité.

c) Pas nécessairement. Sur un échantillon de taille 10, la moyenne fluctue beaucoup : une simulation de dix échantillons de 10 lancers d'un dé équilibré a donné des moyennes allant de 2,9 à 4,6. Pour juger le dé, il faudrait un échantillon bien plus grand.

Réponse. a) $m = 3{,}3$, écart 0,2 ; b) 0,4 contre $\frac{1}{6}$; c) non, simple fluctuation sur 10 lancers.

Correction de l'exercice 7 – Une fonction qui renvoie la moyenne d'un échantillon

a) *ech* est une liste de n entiers, chacun obtenu par un appel à *X()* : c'est un échantillon de taille n de la variable « résultat d'un dé ». Sa longueur *len(ech)* vaut n.

b) *sum(ech)* additionne les n valeurs ; la fonction renvoie donc la moyenne m de l'échantillon. Pour $n = 10\,000$, la moyenne est, sauf exception, très proche de l'espérance

$$\mu = 3{,}5$$

c) On remplace la liste par un accumulateur :

```
def moyenne_echantillon(n):
    s = 0
    for i in range(n):
        s = s + X()
    return s / n
```

Les deux versions renvoient la même chose ; la liste a l'avantage de garder les valeurs, par exemple pour calculer ensuite leur écart type.

Réponse. a) n valeurs simulées, *len(ech)* = n ; b) la moyenne de l'échantillon, proche de 3,5 ; c) accumuler dans s puis renvoyer s / n.

Correction de l'exercice 8 – Dix échantillons de taille 10

a) La plus petite moyenne est 2,9 et la plus grande 4,6 ; l'étendue vaut $4{,}6 - 2{,}9 = 1{,}7$.

b) Une moyenne m convient si $3 \leq m \leq 4$. C'est le cas de 3,4 ; 3,2 ; 3,6 ; 3,5 ; 3,5 ; 3,9 ; 3,5, soit 7 moyennes sur 10. Les moyennes 2,9, 4,6 et 4,5 s'écartent de plus de 0,5.

c) La somme des dix moyennes vaut 36,6, d'où une moyenne de

$$\frac{36{,}6}{10} = 3{,}66$$

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).

Les dix échantillons ayant la même taille, c'est la moyenne des 100 lancers réunis. Elle s'écarte de 0,16 de μ : réunir les échantillons donne une estimation plus précise que chacun d'eux, mais 100 lancers fluctuent encore.

Réponse. a) 1,7 ; b) 7 moyennes ; c) 3,66, la moyenne des 100 lancers.

Correction de l'exercice 9 – Quand la taille de l'échantillon augmente

a) Les étendues valent $4{,}6 - 2{,}9 = 1{,}7$ pour $n = 10$, puis $4{,}01 - 3{,}13 = 0{,}88$ pour $n = 100$ et

$$3{,}594 - 3{,}446 = 0{,}148$$

pour $n = 1\,000$.

b) Plus les échantillons sont grands, moins leurs moyennes sont dispersées, et toutes se resserrent autour de 3,5, l'espérance de X. La moyenne d'un échantillon est une **estimation** de l'espérance, d'autant meilleure que l'échantillon est grand. On remarque aussi que multiplier n par 100 (de 10 à 1 000) a divisé l'étendue par environ 11, et non par 100 : la précision progresse bien moins vite que la taille.

Réponse. a) 1,7 ; 0,88 ; 0,148 ; b) les moyennes se resserrent autour de $\mu = 3{,}5$ quand n augmente.

Correction de l'exercice 10 – La fréquence est une moyenne

a) X suit une loi de Bernoulli de paramètre $p = 0{,}3$: $E(X) = 0{,}3$ et $V(X) = 0{,}3 \times 0{,}7 = 0{,}21$, d'où

$$\sigma = \sqrt{0{,}21} \approx 0{,}458$$

b) La somme des 20 valeurs vaut 5 (chaque « 1 » compte pour 1, chaque « 0 » pour rien), donc $m = \dfrac{5}{20} = 0{,}25$. C'est exactement la proportion de « 1 » dans l'échantillon, c'est-à-dire la fréquence des utilisateurs de vélo.

c) Une fréquence observée est la moyenne d'un échantillon d'une variable de Bernoulli, d'espérance p. Les propriétés observées pour les moyennes (fluctuation d'un échantillon à l'autre, rapprochement de l'espérance quand n grandit) valent donc pour les fréquences, qui se rapprochent de la proportion p.

Réponse. a) $E(X) = 0{,}3$, $\sigma \approx 0{,}458$; b) $m = 0{,}25$, proportion de « 1 » ; c) une fréquence est la moyenne d'un échantillon de Bernoulli.

Correction de l'exercice 11 – Gain moyen observé et espérance

a) Les écarts valent $|0{,}4 - 0{,}4| = 0$, puis $|-0{,}2 - 0{,}4| = 0{,}6$, puis 0,104 et enfin $|0{,}3952 - 0{,}4| = 0{,}0048$. Pour $n = 10$, l'écart nul est une coïncidence ; ensuite, il diminue nettement avec n.

b) Non. L'espérance décrit la moyenne sur un **grand** nombre de parties. L'écart type du gain d'une partie, environ 3,67 €, est grand devant l'espérance : sur 100 parties, la moyenne fluctue assez pour devenir négative.

c) Sur 10 000 parties, on peut espérer environ

$$10\,000 \times 0{,}4 = 4\,000 \text{ €}$$

de gain. Dans la simulation, le gain total vaut $10\,000 \times 0{,}3952 = 3\,952$ €, très proche.

Réponse. a) 0 ; 0,6 ; 0,104 ; 0,0048 ; b) non, fluctuation sur 100 parties ; c) environ 4 000 € ; 3 952 € obtenus.

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 – Transformer une variable

a) Par linéarité de l'espérance, $E(Y) = 2E(X) + 1 = 2 \times 3{,}5 + 1 = 8$.

b) Chaque lancer donne $y_i = 2x_i + 1$ points. En additionnant les 20 égalités puis en divisant par 20 :

$$m_Y = \frac{\sum y_i}{20} = 2 \times \frac{\sum x_i}{20} + 1 = 2m_X + 1 = 8{,}6$$

La moyenne d'un échantillon se transforme comme l'espérance.

c) $E(Z) = 3{,}5 + 3{,}5 = 7$ et $E(T) = 2 \times 3{,}5 = 7$: les espérances sont égales. Mais les lois diffèrent : T ne prend que des valeurs paires $(2, 4, \ldots, 12)$, chacune avec la probabilité $\frac{1}{6}$, alors que Z prend toutes les valeurs de 2 à 12, et vaut 7 avec la probabilité $\frac{6}{36}$. Deux variables de même espérance peuvent être très différentes.

Réponse. a) 8 ; b) 8,6 ; c) même espérance 7, lois différentes.

Correction de l'exercice 13 – Estimer l'espérance d'un dé déformé

a) Sur un grand nombre de lancers, la moyenne de l'échantillon est, sauf exception, proche de l'espérance. Avec 2 000 lancers, on estime $\mu \approx 3{,}83$.

b) On applique la définition de l'espérance :

$$\mu = 0{,}12 + 2 \times 0{,}14 + 3 \times 0{,}16 + 4 \times 0{,}16 + 5 \times 0{,}18 + 6 \times 0{,}24$$

soit $\mu = 0{,}12 + 0{,}28 + 0{,}48 + 0{,}64 + 0{,}9 + 1{,}44 = 3{,}86$.

c) L'estimation 3,8255 s'écarte de 0,0345 seulement de la valeur exacte : elle était bonne. Les deux valeurs dépassent nettement 3,5, l'espérance d'un dé équilibré : le dé favorise les grands nombres, en particulier le 6.

Réponse. a) $\mu \approx 3{,}83$ (moyenne d'un grand échantillon) ; b) $\mu = 3{,}86$; c) oui, écart 0,0345 ; le dé favorise les grands nombres.

Correction de l'exercice 14 – Moyenne d'échantillon ou espérance ?

a) **Faux.** L'espérance est un nombre attaché à la loi de X : elle ne dépend d'aucun échantillon. C'est la moyenne de l'échantillon qui change.

b) **Faux.** La moyenne est proche de 3,5, mais pas exactement égale : une simulation de dix échantillons de 1 000 lancers a donné des moyennes entre 3,446 et 3,594.

c) **Vrai.** Chaque nouvel échantillon donne une nouvelle moyenne : la moyenne d'un échantillon de taille n peut être vue comme une variable aléatoire, définie par la répétition de n expériences.

d) **Vrai.** C'est ce qu'on observe sur toutes les simulations : les moyennes se resserrent autour de μ quand n augmente, même si un échantillon particulier peut s'en écarter.

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

Correction de l'exercice 15 – Synthèse : le stand de la kermesse

a) Le jeu compte 4 as et 12 figures, donc 16 autres cartes. Le joueur gagne $10 - 2 = 8$ € avec la probabilité $\frac{4}{32} = \frac{1}{8}$, $4 - 2 = 2$ € avec la probabilité $\frac{12}{32} = \frac{3}{8}$, et perd 2 € avec la probabilité $\frac{16}{32} = \frac{1}{2}$.

b) $E(G) = 8 \times \frac{1}{8} + 2 \times \frac{3}{8} - 2 \times \frac{1}{2} = 1 + 0{,}75 - 1 = 0{,}75$. Puis

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).

$E(G^2) = 64 \times \frac{1}{8} + 4 \times \frac{3}{8} + 4 \times \frac{1}{2} = 8 + 1{,}5 + 2 = 11{,}5$, d'où

$$V(G) = 11{,}5 - 0{,}75^2 = 10{,}9375 \qquad \sigma(G) \approx 3{,}31$$

L'espérance est **positive** : en moyenne, le joueur gagne 0,75 € par partie, donc les organisateurs **perdent** 0,75 € par partie. Le jeu est mal conçu pour eux.

c) On découpe un tirage *randint(1, 32)* selon les effectifs :

```
from random import randint

def partie():
    c = randint(1, 32)
    if c <= 4:
        return 8
    elif c <= 16:
        return 2
    return -2

def moyenne(n):
    return sum([partie() for i in range(n)]) / n
```

d) Sur 400 parties, le gain moyen des joueurs est, sauf exception, proche de 0,75 € : les organisateurs peuvent s'attendre à **perdre** environ $400 \times 0{,}75 = 300$ €, et non à faire un bénéfice. Le résultat réel fluctuera autour de cette valeur. Pour que le stand rapporte, il faut baisser les gains ou augmenter la mise.

Réponse. a) 8 ; 2 ; -2 avec $\frac{1}{8}$; $\frac{3}{8}$; $\frac{1}{2}$; b) $E(G) = 0{,}75$, $\sigma \approx 3{,}31$: défavorable aux organisateurs ; c) *randint(1, 32)* découpé en 4, 12, 16 ; d) perte d'environ 300 € pour les organisateurs.

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).