Plaquette 4 – Simulations en Python et lecture de résultats

Échantillonnage — Première spé maths

Corrections détaillées

Boîte à outils

Listes.

Écriture	Effet
L = [] puis *L.append(x)*	liste vide, puis ajout de *x* à la fin
[f() for i in range(n)]	liste de *n* appels à *f()* (en compréhension)
[x for x in L if condition]	éléments de *L* qui vérifient la condition
len(L), sum(L), min(L), max(L)	longueur, somme, plus petit, plus grand
L[0], L[-1]	premier et dernier élément

Statistiques d'une liste L de longueur n :

$$m = \frac{\text{sum}(L)}{n} \qquad s = \sqrt{\frac{1}{n}\sum (x-m)^2}$$

Monte-Carlo. Pour estimer une aire A contenue dans le carré unité, on tire des points au hasard dans le carré : la fréquence des points qui tombent dans la région est, sauf exception, proche de A quand le nombre de points est grand.

Bonnes pratiques. Distinguer n (taille d'un échantillon) et N (nombre d'échantillons) ; vérifier une fonction avec un grand n avant de l'utiliser ; ne jamais confondre la valeur affichée (qui fluctue) avec l'espérance.

Correction de l'exercice 1 – Lire un script de simulation

a) La somme des huit valeurs vaut $6+2+5+1+3+6+4+5 = 32$, donc $m = \frac{32}{8} = 4$. La liste *grands* garde les valeurs au moins égales à 5 : $6, 5, 6, 5$, soit 4 éléments, et le second nombre vaut $\frac{4}{8} = 0{,}5$.

b) m est la moyenne d'un échantillon de 8 lancers, à comparer à l'espérance 3,5. Le second nombre est la fréquence des résultats « 5 ou 6 », à comparer à la probabilité

$$P(X \geq 5) = \frac{2}{6} = \frac{1}{3}$$

Sur 8 lancers seulement, les écarts (0,5 pour la moyenne, environ 0,17 pour la fréquence) sont normaux.

Réponse. a) 4 et 0,5 ; b) moyenne (espérance 3,5) et fréquence de « 5 ou 6 » (probabilité $\frac{1}{3}$).

Correction de l'exercice 2 – Corriger un script

Première erreur : *randint(0, 6)* renvoie un entier entre 0 et 6 inclus, soit 7 valeurs ; un dé se simule par *randint(1, 6)*.

Deuxième erreur : la moyenne d'un échantillon de taille n se calcule en divisant par n (ou *len(ech)*), pas par N, qui est le nombre d'échantillons.

Troisième erreur : *L.append(m)* est placé hors de la boucle ; il n'est exécuté qu'une fois,

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).

et la liste ne contient que la dernière moyenne. Il faut l'indenter dans la boucle.

```
from random import randint

def moyennes(N, n):
    L = []
    for k in range(N):
        ech = [randint(1, 6) for i in range(n)]
        L.append(sum(ech) / n)
    return L
```

Contrôle : *len(moyennes(100, 25))* doit valoir 100, et la moyenne de la liste doit être proche de 3,5.

Réponse. *randint(1, 6)* ; diviser par *n* ; *L.append* dans la boucle.

Correction de l'exercice 3 – Une fonction écart type

a) La moyenne vaut $m = \frac{16}{4} = 4$. La liste *carres* contient $(2-4)^2 = 4$, 0, 0 et $(6-4)^2 = 4$. Leur somme vaut 8, donc la fonction renvoie

$$\sqrt{\frac{8}{4}} = \sqrt{2} \approx 1{,}414$$

b) La moyenne des carrés vaut $\frac{4+16+16+36}{4} = 18$, et $18 - 4^2 = 2$: on retrouve une variance de 2 et un écart type de $\sqrt{2}$.

c) Toutes les valeurs sont égales à la moyenne 5 : les carrés des écarts sont nuls et la fonction renvoie 0. Une série sans dispersion a un écart type nul.

Réponse. a) $\sqrt{2} \approx 1{,}414$; b) $18 - 16 = 2$; c) 0.

Correction de l'exercice 4 – Simuler une variable de loi donnée

a) Les entiers de 1 à 10 sont équiprobables, chacun de probabilité 0,1. On réserve 5 valeurs à -1, 3 valeurs à 0 et 2 valeurs à 3 : les seuils sont *k <= 5*, puis *k <= 8*.

b) $E(X) = -1 \times 0{,}5 + 0 + 3 \times 0{,}2 = 0{,}1$. Ensuite $E(X^2) = 1 \times 0{,}5 + 9 \times 0{,}2 = 2{,}3$, d'où

$$V(X) = 2{,}3 - 0{,}1^2 = 2{,}29 \qquad \sigma(X) \approx 1{,}51$$

c) Par exemple *print(sum([X() for i in range(10000)]) / 10000)*. On attend une valeur proche de 0,1 ; plus précisément, dans environ 95 % des cas, à moins de $\frac{2 \times 1{,}51}{100} \approx 0{,}03$ de 0,1.

Réponse. a) *k <= 5* puis *k <= 8* ; b) $E(X) = 0{,}1$, $\sigma \approx 1{,}51$; c) environ 0,1 (à 0,03 près dans 95 % des cas).

Correction de l'exercice 5 – L'aire sous la parabole

a) Le point $(x\,;y)$ est tiré au hasard dans le carré $[0\,;1]^2$, d'aire 1 : la probabilité qu'il tombe sous la parabole, c'est-à-dire que $y \leq x^2$, est l'aire A. La fonction renvoie la fréquence de cet événement sur n points ; d'après la loi des grands nombres, elle est proche de A quand n est grand.

b) Les estimations se stabilisent autour de 0,335 : on conjecture $A = \frac{1}{3}$.

c) La variable « le point est sous la parabole » est de Bernoulli de paramètre $\frac{1}{3}$: $\sigma = \sqrt{\frac{1}{3} \times \frac{2}{3}} = \frac{\sqrt{2}}{3} \approx 0{,}471$. Donc

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).

$$\frac{2\sigma}{\sqrt{100\,000}} \approx \frac{0{,}943}{316{,}2} \approx 0{,}003$$

L'estimation 0,33504 est à 0,0017 de $\frac{1}{3}$, bien dans l'écart habituel.

Réponse. a) Fréquence des points sous la courbe, proche de l'aire ; b) $\frac{1}{3}$; c) $\approx 0{,}003$.

Correction de l'exercice 6 – Compter les échantillons extrêmes

a) On filtre la liste et on mesure sa longueur : *len([m for m in L if m >= 4])*.

b) Une moyenne est hors de l'intervalle si elle est inférieure à 2,817 ou supérieure à 4,183 :

```
hors = [m for m in L if m < 2.817 or m > 4.183]
print(len(hors) / len(L))
```

On pouvait aussi écrire la condition *abs(m - 3.5) > 0.683*.

c) L'intervalle est exactement $[3{,}5 - 0{,}683\,;\,3{,}5 + 0{,}683]$: les simulations prévoient qu'environ

$$5\,\%$$

des moyennes en sortent. La valeur 4,8 %, soit 48 échantillons sur 1 000, est conforme à cette observation.

Réponse. a) *len([m for m in L if m >= 4])* ; b) proportion des *m* avec *abs(m - 3.5) > 0.683* ; c) conforme aux 5 % attendus.

Correction de l'exercice 7 – Une marche aléatoire

a) $E(D) = 1 \times \frac{1}{2} - 1 \times \frac{1}{2} = 0$ et $E(D^2) = 1$, donc $V(D) = 1$. Par linéarité, l'espérance de la position finale est $100 \times 0 = 0$.

b) Le premier nombre est la moyenne des 1 000 positions finales : proche de 0, conformément à l'espérance. Le second est la moyenne des carrés des positions finales, environ 95 : la position finale s'écarte typiquement de 0 d'environ

$$\sqrt{95} \approx 9{,}7 \approx \sqrt{100}$$

pas, même si elle vaut 0 en moyenne.

c) Si le pion fait a pas vers la droite et $100 - a$ vers la gauche, il arrive en $a - (100 - a) = 2a - 100 = 2(a - 50)$, nombre pair.

Réponse. a) $E(D) = 0$, $V(D) = 1$, espérance finale 0 ; b) moyenne ≈ 0, moyenne des carrés ≈ 100 (écart typique ≈ 10) ; c) position $= 2(a - 50)$.

Correction de l'exercice 8 – Un tirage sans remise

a) *choice(urne)* choisit un élément au hasard dans la liste. *urne.remove(b1)* retire de la liste la première occurrence de la boule tirée : la seconde boule est tirée parmi les 4 boules restantes, ce qui simule le tirage sans remise.

b) La première boule est rouge avec la probabilité $\frac{3}{5}$; sachant cela, il reste 2 rouges sur 4 boules :

$$P(RR) = \frac{3}{5} \times \frac{2}{4} = \frac{6}{20} = 0{,}3$$

c) On compte les résultats égaux à "*RR*" :

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 frequence_RR(n):
    return len([t for t in [deux_boules() for i in range(n)] if t == "RR"]) / n
```

Pour n grand, la valeur renvoyée est, sauf exception, proche de 0,3.

Réponse. a) Retirer la boule tirée de l'urne ; b) 0,3 ; c) proportion des "*RR*" dans la liste des n tirages.

Correction de l'exercice 9 – La fréquence des lettres d'un texte

a) *lettres* contient les caractères du texte, espaces exclus. La fonction compte les occurrences de la lettre demandée et renvoie leur fréquence parmi les lettres du texte.

b) « le petit chat » compte 11 lettres hors espaces (l, e, p, e, t, i, t, c, h, a, t), dont 3 « t ». La fonction renvoie

$$\frac{3}{11} \approx 0{,}273$$

c) Non. Un texte de 20 lettres est un petit échantillon : la fréquence d'une lettre y fluctue beaucoup. Avec 0,15 comme valeur de référence, la variable « la lettre est un e » a pour écart type $\sqrt{0{,}15 \times 0{,}85} \approx 0{,}357$, et $\dfrac{2 \times 0{,}357}{\sqrt{20}} \approx 0{,}16$: une fréquence de 0,3 reste dans l'ordre des écarts habituels, à la limite.

Réponse. a) Les lettres du texte ; la fréquence de la lettre ; b) $\frac{3}{11} \approx 0{,}273$; c) non, 20 lettres, c'est peu.

Correction de l'exercice 10 – Repérer une erreur de simulation

a) On obtient 6 soit au premier test (probabilité 0,3), soit lorsque ce test échoue puis que *randint(1, 6)* renvoie 6 (probabilité $0{,}7 \times \frac{1}{6}$). Au total,

$$0{,}3 + \frac{0{,}7}{6} \approx 0{,}417 \neq 0{,}3$$

b) En cas d'échec du premier test, il faut tirer uniformément parmi les faces 1 à 5 : *return randint(1, 5)*. Chacune a alors la probabilité $0{,}7 \times \frac{1}{5} = 0{,}14$, comme voulu.

c) $E = 0{,}14 \times (1 + 2 + 3 + 4 + 5) + 0{,}3 \times 6 = 0{,}14 \times 15 + 1{,}8 = 2{,}1 + 1{,}8 = 3{,}9$.

Réponse. a) $\approx 0{,}417$; b) *return randint(1, 5)* ; c) $E = 3{,}9$.

Correction de l'exercice 11 – Nombre d'échantillons et taille d'échantillon

a) Chacun a simulé $10 \times 1\,000 = 1\,000 \times 10 = 10\,000$ lancers.

b) Léo : ses échantillons sont grands, donc ses 10 moyennes sont proches de 3,5 (écart de référence $\dfrac{2 \times 1{,}708}{\sqrt{1\,000}} \approx 0{,}108$). Mais avec 10 moyennes seulement, il ne peut pas estimer sérieusement une proportion de l'ordre de 95 % : chaque moyenne pèse 10 points. Zoé : ses moyennes fluctuent beaucoup (écart de référence $\approx 1{,}08$), mais elle en a 1 000, ce qui lui permet d'estimer cette proportion de façon fiable.

c) La taille n des échantillons règle la **précision** de chaque moyenne : plus n est grand, plus m est proche de μ. Le nombre N d'échantillons règle la **qualité de l'observation** de la fluctuation : plus N est grand, mieux on connaît la répartition des moyennes.

Réponse. a) 10 000 chacun ; b) Léo pour la précision des moyennes, Zoé pour la proportion ; c) n : précision ; N : observation de la fluctuation.

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 – Construire une liste par ajouts successifs

a) La liste contient 4 éléments : *len(L)* vaut 4, et *L[-1]*, dernier élément, vaut 6.

b) Oui : la boucle s’arrête dès que *d* vaut 6, et ce 6 a été ajouté juste avant. Elle ne peut pas être vide : comme *d* vaut 0 au départ, la boucle est exécutée au moins une fois.

c) Par exemple *sum([len(jusqu_a_six()) for k in range(10000)]) / 10000*. La longueur de la liste est le nombre de lancers nécessaires pour obtenir un 6. Des simulations donnent une moyenne proche de

$$6$$

ce qui est cohérent avec une probabilité de $\frac{1}{6}$ à chaque lancer.

Réponse. a) 4 et 6 ; b) toujours terminée par 6, jamais vide ; c) moyenne des longueurs, environ 6.

Correction de l’exercice 13 – Indépendance dans une simulation

a) Les deux conditions portent sur le **même** nombre u : elles sont vraies ou fausses en même temps. La fonction renvoie *True* avec la probabilité 0,6, comme si le second lancer recopiait le premier. Les deux lancers ne sont pas indépendants.

b) Il faut un nouveau tirage pour chaque lancer :

```
def deux_faces():
    return random() < 0.6 and random() < 0.6
```

Les deux lancers étant indépendants, la probabilité d’obtenir deux « face » est le produit

$$0{,}6 \times 0{,}6 = 0{,}36$$

Une simulation avec un grand nombre d’appels doit donner une fréquence proche de 0,36.

Réponse. a) 0,6 : le même tirage sert deux fois ; b) deux appels à *random()* ; probabilité 0,36.

Correction de l’exercice 14 – Programmer une étude complète

a) X suit une loi de Bernoulli de paramètre 0,25 : $\mu = 0{,}25$ et $\sigma = \sqrt{0{,}25 \times 0{,}75} \approx 0{,}433$. Pour $n = 400$,

$$\frac{2\sigma}{\sqrt{400}} \approx \frac{0{,}866}{20} \approx 0{,}043$$

b) On construit la liste des moyennes, puis les quatre indicateurs :

```
from random import random
from math import sqrt

def etude(N, n):
    mu = 0.25
    e = 2 * sqrt(0.25 * 0.75) / sqrt(n)
    L = [sum([1 for i in range(n) if random() < 0.25]) / n for k in range(N)]
    proches = [m for m in L if abs(m - mu) <= e]
    return sum(L) / N, min(L), max(L), len(proches) / N
```

c) La moyenne des moyennes doit être très proche de 0,25 ; la plupart des moyennes sont entre 0,207 et 0,293, les extrêmes s’en écartant un peu ; la proportion renvoyée doit être proche de 0,95.

Réponse. a) $\mu = 0{,}25$, $\sigma \approx 0{,}433$, écart $\approx 0{,}043$; b) liste des moyennes puis indicateurs ;

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) $\approx 0{,}25$, extrêmes autour de $[0{,}2\,;0{,}3]$, proportion $\approx 0{,}95$.

Correction de l'exercice 15 – Synthèse : le protocole complet

a) Un double a la probabilité $\frac{6}{36}=\frac{1}{6}$. Donc G vaut 4 avec la probabilité $\frac{1}{6}$ et -1 avec la probabilité $\frac{5}{6}$. On obtient $E(G)=\frac{4}{6}-\frac{5}{6}=-\frac{1}{6}\approx -0{,}17$ €, puis $E(G^2)=\frac{16}{6}+\frac{5}{6}=\frac{21}{6}=3{,}5$, d'où

$$V(G)=3{,}5-\frac{1}{36}\approx 3{,}472 \qquad \sigma(G)\approx 1{,}86$$

b) On teste l'égalité des deux dés :

```
from random import randint

def gain():
    if randint(1, 6) == randint(1, 6):
        return 4
    return -1
```

c) Une liste en compréhension de moyennes :

```
def moyennes(N, n):
    return [sum([gain() for i in range(n)]) / n for k in range(N)]
```

d) $\frac{2\times 1{,}86}{\sqrt{400}}=\frac{3{,}72}{20}\approx 0{,}19$. Environ 5 % des 1 000 gains moyens, soit une cinquantaine, devraient être hors de $[-0{,}17-0{,}19\,;-0{,}17+0{,}19]=[-0{,}36\,;0{,}02]$.

e) Non : l'espérance est négative, le joueur perd en moyenne environ 0,17 € par partie. Si le joueur perd une mise x au lieu de 1 € quand il n'obtient pas de double, l'espérance vaut $\frac{4}{6}-\frac{5x}{6}$, nulle pour $x=0{,}8$ € : il faudrait que le joueur ne perde que 0,80 €.

Réponse. a) 4 avec $\frac{1}{6}$, -1 avec $\frac{5}{6}$; $E(G)\approx -0{,}17$, $\sigma\approx 1{,}86$; b) et c) voir corrigé ; d) $\approx 0{,}19$; une cinquantaine ; e) non ; perte de 0,80 € au lieu de 1 €.

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).