Plaquette 2 – Simulation et stabilisation des fréquences

Échantillonnage et fluctuation — Seconde

Corrections détaillées

Boîte à outils

Les briques Python de la simulation.

Instruction	Rôle
from random import random, randint	importe les générateurs
random()	un décimal au hasard dans [0 ; 1[
random() < p	condition vraie avec la probabilité p
randint(a, b)	un entier au hasard entre a et b, **inclus**
for i in range(n)	répète n fois
while condition	répète tant que la condition est vraie

Structure type. Initialiser un compteur à 0, répéter n fois l'expérience, ajouter 1 au compteur quand l'événement se produit, puis renvoyer la fréquence :

$$f = \frac{\text{compteur}}{n}$$

Tableur. *=ALEA()* joue le rôle de *random()*, *=ALEA.ENTRE.BORNES(1;6)* celui de *randint(1, 6)*, et *=NB.SI(plage;6)* compte les cellules égales à 6. La touche F9 relance tous les tirages.

Ce que l'on observe. Deux simulations donnent des résultats différents (fluctuation) ; quand n grandit, la fréquence se rapproche, sauf exception, de la probabilité (loi des grands nombres).

Correction de l'exercice 1 – Simuler un événement de probabilité donnée

a) Le nombre renvoyé par *random()* est réparti uniformément dans [0 ; 1[: la probabilité qu'il tombe dans [0 ; 0,3[est la longueur de cet intervalle. La condition *random() < 0.3* est donc vraie avec la probabilité 0,3.

b) Pour simuler un événement de probabilité 0,65, on teste si le nombre tiré tombe dans un intervalle de longueur 0,65 : la condition *random() < 0.65* convient, « vrai » signifiant « lancer réussi ».

c) *random()* renvoie toujours un nombre strictement inférieur à 1 : la condition *random() < 1* est toujours vraie. Elle simule un événement certain, de probabilité 1, où le hasard ne joue aucun rôle.

Réponse. a) 0,3 ; b) *random() < 0.65* ; c) elle est toujours vraie (événement certain).

Correction de l'exercice 2 – Simuler un dé et l'événement « pair »

a) *randint(1, 6)* renvoie un entier entre 1 et 6, bornes incluses. Le générateur étant supposé parfait, ces six valeurs sont équiprobables, chacune de probabilité $\frac{1}{6}$.

b) La condition est vraie pour les trois valeurs 2, 4 et 6. Les issues étant équiprobables :

$$P(\text{« pair »}) = \frac{3}{6} = \frac{1}{2}$$

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) On remplace la condition par *if d == 5 or d == 6:* et l'affichage par *print("gagné").* Deux issues sur six réalisent l'événement, donc la probabilité d'afficher « gagné » vaut $\frac{2}{6} = \frac{1}{3}$.

Réponse. a) 1 à 6, chacune avec la probabilité $\frac{1}{6}$; b) $\frac{1}{2}$; c) *if d == 5 or d == 6:*, probabilité $\frac{1}{3}$.

Correction de l'exercice 3 – Exécuter une simulation à la main

a) À chaque passage, on compare le nombre tiré à 0,4 et l'on ajoute 1 à *succes* seulement s'il est strictement inférieur :

Passage	**1**	**2**	**3**	**4**	**5**
random()	0,12	0,75	0,43	0,91	0,38
succes	1	1	1	1	2

Le nombre 0,43 n'est pas inférieur à 0,4 : il ne compte pas.

b) Après la boucle, la fonction renvoie

$$\frac{\text{succes}}{n} = \frac{2}{5} = 0{,}4$$

C'est la fréquence de succès observée sur ces 5 répétitions d'une expérience de probabilité 0,4. Ici elle tombe exactement sur la probabilité, mais c'est une coïncidence : avec d'autres tirages, on aurait pu obtenir 0,2 ou 0,6.

Réponse. a) 1, 1, 1, 1, 2 ; b) 0,4, la fréquence de succès sur 5 répétitions.

Correction de l'exercice 4 – Corriger un programme

a) L'instruction *compteur = 0* est placée **dans** la boucle : à chaque lancer, le compteur est remis à zéro, puis augmenté de 1 seulement si ce lancer donne 6. À la sortie de la boucle, *compteur* ne retient que le dernier lancer : il vaut 1 si le dernier lancer est un 6, et 0 sinon.

b) La fonction renvoie *compteur*, c'est-à-dire un nombre de « 6 », alors qu'on veut une fréquence : il faut diviser par *n*. Voici la version corrigée :

```
from random import randint

def frequence_six(n):
    compteur = 0
    for i in range(n):
        if randint(1, 6) == 6:
            compteur = compteur + 1
    return compteur / n
```

L'initialisation se fait une seule fois, avant la boucle, et la fonction renvoie une fréquence, qui devrait être proche de $\frac{1}{6} \approx 0{,}167$ pour $n = 1\,000$.

Réponse. a) Le compteur est remis à 0 à chaque tour ; b) la fonction renvoie un effectif au lieu d'une fréquence : initialiser avant la boucle et renvoyer *compteur / n*.

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

a) Les couples (premier dé, second dé) donnant 7 sont $(1,6)$, $(2,5)$, $(3,4)$, $(4,3)$, $(5,2)$ et $(6,1)$, soit 6 issues sur 36 :

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).

$$P(S = 7) = \frac{6}{36} = \frac{1}{6} \approx 0{,}167$$

b) Chaque dé est simulé par un appel à *randint(1, 6)* :

```
from random import randint

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

c) Les fréquences sont $\frac{16}{100} = 0{,}16$, puis $\frac{165}{1\,000} = 0{,}165$ et $\frac{1\,647}{10\,000} = 0{,}1647$. Elles se rapprochent de 0,167 à mesure que le nombre de lancers augmente, comme l'annonce la loi des grands nombres.

Réponse. a) $\frac{1}{6}$; b) *return randint(1, 6) + randint(1, 6)* ; c) 0,16 ; 0,165 ; 0,1647, de plus en plus proches de $\frac{1}{6}$.

Correction de l'exercice 6 – Attendre le premier 6

a) On suit l'exécution. Le premier lancer donne 3, différent de 6 : *lancers* passe à 2. Le deuxième donne 5 : *lancers* passe à 3. Le troisième donne 1 : *lancers* passe à 4. Le quatrième donne 6 : la condition *randint(1, 6) != 6* est fausse, la boucle s'arrête et la fonction renvoie 4.

b) La valeur renvoyée est le numéro du lancer où apparaît le premier 6, autrement dit le nombre de lancers qu'il a fallu effectuer pour obtenir un 6.

c) On ne sait pas à l'avance combien de lancers seront nécessaires : cela peut être 1, 4 ou 20. Une boucle *for* répète un nombre fixé de fois ; la boucle *while* répète **tant que** le 6 n'est pas sorti, ce qui correspond exactement à l'expérience.

Réponse. a) 4 ; b) le rang du premier 6 ; c) le nombre de répétitions n'est pas connu à l'avance.

Correction de l'exercice 7 – Une simulation au tableur

a) Chaque cellule de A1 à A500 contient un entier tiré au hasard entre 1 et 6 : la colonne simule 500 lancers d'un dé équilibré.

b) *NB.SI(A1:A500;6)* compte les cellules égales à 6 ; divisé par 500, cela donne la fréquence des « 6 » sur les 500 lancers. On s'attend à une valeur proche de la probabilité

$$p = \frac{1}{6} \approx 0{,}167$$

c) La touche F9 relance tous les tirages : on obtient un nouvel échantillon de 500 lancers, donc une nouvelle fréquence. C'est la fluctuation d'échantillonnage. Pour qu'elle soit plus faible, il faut des échantillons plus grands : en recopiant la formule jusqu'à A5000 et en divisant par 5 000, les fréquences seront nettement plus groupées autour de 0,167.

Réponse. a) 500 lancers simulés ; b) la fréquence des « 6 », proche de $\frac{1}{6}$; c) nouveaux tirages (fluctuation) ; augmenter le nombre de lancers.

Correction de l'exercice 8 – Estimer π par simulation

a) Le carré a une aire de 1 ; le quart de disque de rayon 1 a une aire égale au quart de $\pi \times 1^2$:

$$P = \frac{\pi}{4} \approx 0{,}785$$

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).

Le point M est dans le quart de disque lorsque sa distance à O est au plus 1, c'est-à-dire $\sqrt{x^2+y^2} \leq 1$, ou encore $x^2+y^2 \leq 1$.

b) Sur la figure, seuls 2 points sont au-delà de l'arc de cercle : 28 points sur 30 sont dans le quart de disque. La fréquence $\frac{28}{30}$ estime $\frac{\pi}{4}$, donc $\pi \approx 4 \times \frac{28}{30} \approx 3{,}73$. L'estimation est grossière : 30 points, c'est trop peu.

c) La fréquence vaut $\frac{78\,360}{100\,000} = 0{,}7836$, d'où $\pi \approx 4 \times 0{,}7836 = 3{,}1344$. Avec beaucoup plus de points, la fréquence est, sauf exception, proche de $\frac{\pi}{4}$, et l'on retrouve les deux premières décimales de $\pi \approx 3{,}14$.

Réponse. a) $\frac{\pi}{4}$; $x^2+y^2 \leq 1$; b) 28 points, $\pi \approx 3{,}73$; c) $\pi \approx 3{,}1344$.

Correction de l'exercice 9 – Une marche aléatoire

a) Un nombre inférieur à 0,5 fait avancer le pion, un nombre supérieur le fait reculer :

Étape	1	2	3	4	5	6
random()	0,7	0,2	0,4	0,9	0,1	0,3
position	−1	0	1	0	1	2

La fonction renvoie 2.

b) Si le pion avance a fois et recule r fois, alors $a + r = 6$ et sa position finale vaut

$$a - r = a - (6 - a) = 2a - 6 = 2(a - 3)$$

C'est un nombre pair, quels que soient les tirages. Les positions possibles après 6 étapes sont donc -6, -4, -2, 0, 2, 4 et 6.

Réponse. a) −1 ; 0 ; 1 ; 0 ; 1 ; 2, la fonction renvoie 2 ; b) la position vaut $2(a-3)$.

Correction de l'exercice 10 – Deux simulations, deux résultats

a) Anaïs obtient $f_A = \frac{7}{20} = 0{,}35$ et Bruno $f_B = \frac{326}{2\,000} = 0{,}163$.

b) Pas nécessairement. Sur 20 lancers seulement, la fréquence fluctue beaucoup : obtenir 7 « 6 » au lieu des 3 ou 4 attendus est inhabituel, mais possible avec un dé équilibré. Un écart sur un si petit échantillon ne permet pas de conclure que le programme est faux ; il faudrait le relancer avec beaucoup plus de lancers.

c) La fréquence de Bruno est la plus fiable. D'après la loi des grands nombres, quand n est grand, la fréquence est, sauf exception, proche de la probabilité ; or 2 000 lancers, c'est cent fois plus que 20. Elle est d'ailleurs très proche de

$$\frac{1}{6} \approx 0{,}167$$

Réponse. a) 0,35 et 0,163 ; b) non, la fluctuation est forte sur 20 lancers ; c) celle de Bruno.

Correction de l'exercice 11 – Probabilité contre fréquence

a) En notant P pour « pile » et F pour « face », les quatre issues équiprobables sont PP, PF, FP et FF. Une seule réalise l'événement :

$$P(\text{deux « face »}) = \frac{1}{4} = 0{,}25$$

b) Chaque variable *piece1*, *piece2* contient un booléen, vrai pour « face ». On veut que les

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).

deux soient vraies : la fonction renvoie *piece1 and piece2*.

c) La fréquence observée est $\frac{255}{1\,000} = 0{,}255$. Elle s'écarte de 0,005 seulement de la probabilité 0,25 : sur 1 000 répétitions, l'écart est faible, comme le prévoit la loi des grands nombres, sans que la fréquence soit exactement égale à la probabilité.

Réponse. a) $\frac{1}{4}$; b) *return piece1 and piece2* ; c) $f = 0{,}255$, très proche de 0,25.

Correction de l'exercice 12 – Simuler un sondage

a) Chaque appel simule l'interrogation de n habitants au hasard ; la condition *random() < 0.4* traduit qu'une personne interrogée est favorable avec la probabilité 0,4. Le nombre renvoyé est la fréquence de « oui » dans cet échantillon.

b) Pour $n = 100$, les valeurs vont de 0,35 à 0,48 ; pour $n = 1\,000$, de 0,38 à 0,415. Les étendues sont donc

$$0{,}48 - 0{,}35 = 0{,}13 \qquad \text{et} \qquad 0{,}415 - 0{,}38 = 0{,}035$$

Les sondages de 1 000 personnes donnent des résultats nettement plus groupés autour de la proportion réelle 0,4 : plus l'échantillon est grand, plus la fluctuation est faible. Avec 100 personnes, un sondage pourrait annoncer à tort près d'un habitant sur deux favorable.

Réponse. a) La fréquence de « oui » dans un échantillon ; b) étendues 0,13 et 0,035 : la fluctuation diminue quand n augmente.

Correction de l'exercice 13 – Deux façons de simuler une probabilité de un tiers

a) Les trois jetons ont la même chance d'être tirés et un seul est rouge : la probabilité vaut $\frac{1}{3}$.

b) La condition A est vraie lorsque le nombre tiré dans $[0\,;1[$ tombe dans $\left[0\,;\frac{1}{3}\right[$, intervalle de longueur $\frac{1}{3}$: elle est vraie avec la probabilité $\frac{1}{3}$. La condition B tire un entier parmi 1, 2, 3, de façon équiprobable, et n'est vraie que pour l'un d'eux : probabilité $\frac{1}{3}$ aussi. On peut imaginer que l'entier 1 représente le jeton rouge et 2, 3 les jetons bleus.

c) *randint(0, 3)* tire un entier parmi 0, 1, 2, 3 : il y a **quatre** valeurs possibles, les bornes étant incluses. La condition est vraie pour une seule d'entre elles, donc elle simule un événement de probabilité

$$\frac{1}{4} \neq \frac{1}{3}$$

Réponse. a) $\frac{1}{3}$; b) chacune est vraie avec la probabilité $\frac{1}{3}$; c) quatre valeurs possibles : probabilité $\frac{1}{4}$.

Correction de l'exercice 14 – Une fonction à deux arguments

a) On reprend la structure habituelle en remplaçant la probabilité écrite en dur par le paramètre p :

```
from random import random

def frequence(n, p):
    succes = 0
    for i in range(n):
```

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).

```
        if random() < p:
            succes = succes + 1
    return succes / n
```

b) Pour la pièce : *frequence(500, 0.5)*. Pour l'urne : *frequence(2000, 0.15)*, le succès étant « la boule est rouge ».

c) Non. Sur 10 répétitions, obtenir 8 succès alors que la probabilité vaut 0,5 est peu fréquent, mais possible : la fluctuation d'échantillonnage est forte sur un si petit nombre d'essais. Pour tester la fonction, il vaut mieux l'appeler avec un grand n, par exemple *frequence(100000, 0.5)*, qui doit renvoyer, sauf exception, un nombre très proche de 0,5.

Réponse. a) Boucle de n tours avec la condition *random()* < p, puis *return succes / n* ; b) *frequence(500, 0.5)* et *frequence(2000, 0.15)* ; c) non, simple fluctuation sur 10 essais.

Correction de l'exercice 15 – Synthèse : le problème du duc de Toscane

a) On ajoute trois tirages entiers entre 1 et 6 :

```
from random import randint

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

b) Pour 100 000 lancers, les fréquences sont $\frac{11\,713}{100\,000} \approx 0{,}117$ pour la somme 9 et $\frac{12\,502}{100\,000} = 0{,}125$ pour la somme 10.

c) Les issues étant équiprobables :

$$P(S = 9) = \frac{25}{216} \approx 0{,}116 \qquad P(S = 10) = \frac{27}{216} = 0{,}125$$

Les fréquences obtenues sur 100 000 lancers sont très proches de ces probabilités. L'erreur du raisonnement « six façons chacune » est de compter $\{3\,;3\,;3\}$ comme une façon, alors que ce tirage ne correspond qu'à une issue, tandis que $\{1\,;2\,;6\}$ en regroupe six.

d) Les deux probabilités ne diffèrent que de $\frac{2}{216} \approx 0{,}009$. Sur 1 000 lancers, les fréquences (0,110 et 0,115) fluctuent autant que cet écart : la différence est noyée dans la fluctuation. Il faut un très grand nombre de parties pour que, selon la loi des grands nombres, les fréquences soient assez proches des probabilités pour révéler l'écart.

Réponse. a) Somme de trois *randint(1, 6)* ; b) ≈ 0,117 et 0,125 ; c) $\frac{25}{216} \approx 0{,}116$ et $\frac{27}{216} = 0{,}125$; d) l'écart, ≈ 0,009, est plus petit que la fluctuation sur peu de parties.

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).