Plaquette 2 – Distance entre la moyenne d'un échantillon et l'espérance

Échantillonnage — Première spé maths

Exercices

Méthode

On mesure la **distance** $|m - \mu|$ entre la moyenne m d'un échantillon simulé et l'espérance μ de la variable. En simulant de nombreux échantillons de tailles différentes, on étudie comment cette distance se comporte : elle diminue quand n augmente, à peu près comme $\frac{1}{\sqrt{n}}$, et elle est d'autant plus grande que la variable est dispersée (grand écart type σ).

Exercice 1 – Dix distances à l'espérance

On a simulé dix échantillons de 100 lancers d'un dé équilibré. Leurs moyennes sont :

3,56 ; 3,43 ; 3,16 ; 3,13 ; 3,68 ; 3,65 ; 4,01 ; 3,71 ; 3,33 ; 3,43.

a) Calculer, pour chaque échantillon, la distance $|m - 3{,}5|$.

b) Quelle est la plus grande distance ? Calculer la moyenne des dix distances.

c) Quelle proportion des échantillons a une moyenne à moins de 0,2 de l'espérance ?

Exercice 2 – Programmer le calcul des distances

On dispose de la fonction *moyenne_echantillon(n)*, qui renvoie la moyenne de n lancers simulés d'un dé équilibré.

```
def distances(N, n):
    return [abs(moyenne_echantillon(n) - 3.5) for k in range(N)]

def distance_moyenne(N, n):
    d = distances(N, n)
    return sum(d) / N
```

a) Que contient la liste renvoyée par *distances(10, 100)* ? Combien de lancers sont simulés ?

b) Que renvoie *distance_moyenne(N, n)* ?

c) Écrire une fonction *distance_max(N, n)* qui renvoie la plus grande des N distances.

Exercice 3 – Quand n est multiplié par 4

Pour 500 échantillons de chaque taille n (lancers d'un dé équilibré), on a calculé la distance moyenne d entre la moyenne de l'échantillon et l'espérance 3,5 :

Taille n	**25**	**100**	**400**
Distance moyenne d	0,2739	0,1345	0,0689

a) Par combien la distance moyenne est-elle divisée quand n est multiplié par 4 ?

b) Calculer $d \times \sqrt{n}$ pour chaque taille. Quelle conjecture peut-on formuler ?

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) Prévoir la distance moyenne pour $n = 10\,000$.

Exercice 4 – Écart type des moyennes et σ/√n

Avec les mêmes 500 échantillons par taille, on a calculé l'écart type des moyennes obtenues :

Taille n	**25**	**100**	**400**
Moyenne des moyennes	3,524	3,496	3,511
Écart type des moyennes	0,3436	0,1693	0,0862

a) Que remarque-t-on sur la moyenne des moyennes ?

b) Calculer $\frac{\sigma}{\sqrt{n}}$ pour chaque taille, avec $\sigma = \sqrt{\frac{35}{12}}$ (arrondir au dix-millième). Comparer à l'écart type des moyennes.

Exercice 5 – Deux jeux de même espérance

On compare deux jeux. Au jeu A, on gagne 2 € avec la probabilité 0,5, rien sinon. Au jeu B, on gagne 10 € avec la probabilité 0,1, rien sinon.

a) Calculer l'espérance et l'écart type du gain à chaque jeu.

b) On a simulé 500 échantillons de 50 parties pour chaque jeu. Les gains moyens par partie ont un écart type de 0,143 au jeu A et de 0,455 au jeu B. Comparer ces valeurs à $\frac{\sigma}{\sqrt{50}}$.

c) Sur 50 parties, à quel jeu le gain moyen est-il le plus imprévisible ?

Exercice 6 – Taille nécessaire pour une précision donnée

On a constaté par simulation que les moyennes d'échantillons de taille n d'une variable d'écart type σ ont un écart type d'environ $\frac{\sigma}{\sqrt{n}}$.

a) Pour un dé équilibré ($\sigma \approx 1{,}708$), quelle taille d'échantillon faut-il pour que cet écart type soit au plus 0,05 ?

b) Pour une variable de Bernoulli de paramètre 0,5, quelle taille faut-il pour que l'écart type des fréquences soit au plus 0,01 ?

Exercice 7 – Fréquences d'échantillons

Dans une population, 30 % des personnes sont abonnées à un service de musique. On a simulé 1 000 sondages de 100 personnes ; les fréquences d'abonnés obtenues allaient de 0,15 à 0,45.

a) Quelle est l'espérance et l'écart type de la variable de Bernoulli « la personne est abonnée » ?

b) Calculer $\frac{\sigma}{\sqrt{100}}$. Que représente cette valeur pour les fréquences simulées ?

c) Les fréquences extrêmes 0,15 et 0,45 sont-elles fréquentes ? Combien d'écarts types les séparent de 0,3 ?

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).

Exercice 8 – Vrai ou faux sur la distance à l'espérance

Pour chaque affirmation, dire si elle est vraie ou fausse et justifier.

a) Un échantillon de taille 400 a toujours une moyenne plus proche de μ qu'un échantillon de taille 100.

b) En multipliant la taille des échantillons par 9, on divise environ par 3 la distance typique entre m et μ.

c) Si $\sigma = 0$, toutes les moyennes d'échantillons sont égales à μ.

d) La distance typique entre m et μ ne dépend pas de la variable étudiée.

Exercice 9 – Estimer l'écart type à partir d'un échantillon

Voici un échantillon de 10 lancers d'un dé équilibré : $5, 1, 4, 4, 5, 1, 2, 4, 4, 3$.

a) Calculer sa moyenne m et la moyenne des carrés de ses valeurs.

b) En déduire l'écart type s de l'échantillon (arrondi au millième) et le comparer à l'écart type $\sigma \approx 1{,}708$ du dé.

c) Écrire une fonction Python *ecart_type(liste)* qui renvoie l'écart type d'une liste de nombres.

Exercice 10 – Changer d'unité

Une variable X donne la durée d'un trajet en heures ; elle a pour espérance $\mu = 0{,}75$ et pour écart type $\sigma = 0{,}2$. On note $Y = 60X$ la même durée en minutes.

a) Calculer l'espérance et l'écart type de Y.

b) Pour des échantillons de 16 trajets, la distance typique entre la moyenne et l'espérance est d'environ $\frac{\sigma}{\sqrt{16}}$. Calculer cette distance pour X, puis pour Y.

c) Commenter.

Exercice 11 – Moyenne de plusieurs échantillons

Cinq élèves ont chacun simulé 200 lancers d'un dé équilibré et obtenu les moyennes 3,42 ; 3,61 ; 3,55 ; 3,38 ; 3,49.

a) Calculer la moyenne de ces cinq moyennes. À quelle taille d'échantillon correspond-elle ?

b) La distance typique pour un échantillon de taille 200 est d'environ $\frac{\sigma}{\sqrt{200}}$. Calculer cette valeur et celle qui correspond à l'échantillon regroupé ($\sigma \approx 1{,}708$).

c) La moyenne regroupée est-elle proche de 3,5 ?

Exercice 12 – Lire un programme de simulation complet

On considère le programme suivant.

```
from random import random

def X():
    if random() < 0.2:
        return 5
    return 0
```

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

L = moyennes(200, 100)
print(sum(L) / len(L))
print(min(L), max(L))
```

a) Déterminer la loi, l'espérance et l'écart type de la variable simulée par *X()*.

b) Que contient la liste L ? Combien de valeurs de X sont simulées ?

c) Autour de quelle valeur s'attend-on à ce que soit le premier nombre affiché ? Quelle distance typique entre *min(L)* ou *max(L)* et cette valeur peut-on prévoir, en nombre d'écarts types $\frac{\sigma}{\sqrt{n}}$?

Exercice 13 – Le temps d'attente au guichet

Le temps d'attente T (en minutes) d'un client à un guichet suit la loi suivante :

t	**0**	**5**	**10**	**20**
$P(T = t)$	0,4	0,3	0,2	0,1

a) Calculer $E(T)$ et $\sigma(T)$ (arrondir au centième).

b) Un jour, on relève le temps d'attente de 36 clients ; leur moyenne vaut 6,9 minutes. Calculer la distance à l'espérance et la comparer à $\frac{\sigma}{\sqrt{36}}$.

c) Le responsable estime que l'attente « s'est allongée ce jour-là ». Les données le justifient-elles ?

Exercice 14 – Estimer une proportion inconnue

Pour estimer la proportion p de foyers équipés de panneaux solaires dans une région, on interroge n foyers tirés au hasard. La réponse d'un foyer est une variable de Bernoulli de paramètre p.

a) Exprimer l'écart type σ de cette variable en fonction de p. Pour quelle valeur de p est-il le plus grand ? Que vaut-il alors ?

b) En déduire que, quelle que soit la valeur de p, la distance typique entre la fréquence observée et p, d'environ $\frac{\sigma}{\sqrt{n}}$, est au plus $\frac{1}{2\sqrt{n}}$.

c) Quelle taille d'échantillon garantit une distance typique d'au plus 0,01 ?

Exercice 15 – Synthèse : la masse des pommes

Un producteur trie ses pommes par calibre. La masse M (en grammes) d'une pomme du calibre « moyen » prend, de façon simplifiée, les valeurs 150, 160, 170 et 180 avec les probabilités 0,2 ; 0,3 ; 0,3 ; 0,2.

a) Calculer $E(M)$ et $\sigma(M)$ (arrondir au centième).

b) Les pommes sont vendues en sachets de 25. Exprimer la masse moyenne d'une pomme d'un sachet comme la moyenne d'un échantillon. Quelle est la distance typique entre cette masse moyenne et $E(M)$?

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) Écrire une fonction Python *sachet()* qui simule la masse totale d'un sachet.

d) La masse totale d'un sachet est 25 fois la masse moyenne de ses pommes. En déduire la masse totale « attendue » d'un sachet et la distance typique à cette valeur.

e) Le producteur affiche « 4 kg » sur ses sachets. Est-ce raisonnable ?

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).