

Fiche 4 : loi des grands nombres et échantillonnage

Sept exercices pour maîtriser la loi des grands nombres :

- une pièce suspecte : que peut-on conclure ?
- la moyenne d'un dé et le rang n à partir duquel elle est proche de 3,5 ;
- estimer une probabilité par simulation : combien de simulations ?
- une loi binomiale comparée à l'inégalité de Bienaymé-Tchebychev ;
- un risque d'assurance ;
- un script Python de simulation ;
- vrai/faux.

Toutes les corrections sont rédigées comme on l'attend au bac :

- μ , V et δ écrits ;
- l'inégalité de concentration citée ;
- la conclusion est **nuancée** : une probabilité, pas une certitude.

Le modèle de rédaction d'une conclusion statistique

« Si la pièce est équilibrée, F_n est la fréquence de « pile » sur $n = 10\,000$ lancers, d'espérance 0,5 et de variance $\frac{0,25}{10\,000}$. D'après l'inégalité de concentration avec $\delta = 0,04$:

$$P(|F_n - 0,5| \geq 0,04) \leq \frac{0,25}{10\,000 \times 0,0016} \approx 0,0156.$$

Une fréquence de 0,54 (écart de 0,04) ne se produirait donc qu'avec une probabilité d'au plus 1,6 % pour une pièce équilibrée : il est raisonnable de **douter** de l'équilibre de la pièce. »

- **Formule** de la borne, **valeur numérique**, **conclusion nuancée**.
- **Ne dis jamais** « la pièce est truquée » : dis « il est raisonnable de douter ».

Les pièges classiques

- Croire que la loi des grands nombres dit que la moyenne **est égale** à l'espérance : elle dit que la **probabilité d'un écart** tend vers 0.
- Conclure avec certitude à partir d'une **majoration**.
- Oublier que la précision ne progresse qu'en $\frac{1}{\sqrt{n}}$: quatre fois plus de données pour deux fois plus de précision.
- Oublier la **majoration** $p(1-p) \leq \frac{1}{4}$ quand p est inconnu.

« Après dix pile de suite, face est bien due. » Non.

La pièce n'a pas de mémoire.

La loi des grands nombres ne rembourse pas les dettes : elle dilue les écarts.

Exercice 1 (★★) : une pièce suspecte

On lance n fois une pièce et on note F_n la fréquence de « pile ».

1. Pour $n = 1\,000$, on observe $F_n = 0,54$. Si la pièce est équilibrée, majorer $P(|F_n - 0,5| \geq 0,04)$. Peut-on douter de l'équilibre de la pièce ?
2. Même question pour $n = 10\,000$ et la même fréquence observée 0,54.

Correction de l'exercice 1

Si la pièce est équilibrée, $E(F_n) = 0,5$ et $V(F_n) = \frac{0,25}{n}$. D'après l'inégalité de concentration avec $\delta = 0,04$:

$$P(|F_n - 0,5| \geq 0,04) \leq \frac{0,25}{n \times 0,0016}$$

Question 1

Pour $n = 1\,000$: $\frac{0,25}{1\,000 \times 0,0016} = 0,15625$.

$$P(|F_{1000} - 0,5| \geq 0,04) \leq 0,156$$

Un écart de 0,04 peut se produire avec une probabilité allant **jusqu'à 15,6 %** : ce n'est pas assez petit pour douter.

On ne peut pas douter de l'équilibre de la pièce.

Question 2

Pour $n = 10\,000$: $\frac{0,25}{10\,000 \times 0,0016} = 0,015625$.

$$P(|F_{10000} - 0,5| \geq 0,04) \leq 0,0156$$

Un écart de 0,04 aurait au plus 1,6 % de chances de se produire pour une pièce équilibrée.

Il est raisonnable de douter de l'équilibre de la pièce.

Exercice 2 (★★) : la moyenne d'un dé converge

On lance n fois un dé équilibré et on note M_n la moyenne des résultats.

1. Montrer que $P(|M_n - 3,5| \geq 0,1) \leq \frac{875}{3n}$.
2. En déduire la limite de $P(|M_n - 3,5| \geq 0,1)$ quand $n \rightarrow +\infty$.
3. À partir de quelle valeur de n cette probabilité est-elle inférieure ou égale à 0,05 ?

Correction de l'exercice 2

Question 1

Pour un lancer : $\mu = \frac{7}{2}$ et $V = \frac{35}{12}$. D'après l'inégalité de concentration avec $\delta = 0,1$:

$$P(|M_n - 3,5| \geq 0,1) \leq \frac{35/12}{n \times 0,01} = \frac{35}{0,12n} = \frac{875}{3n}$$

$$P(|M_n - 3,5| \geq 0,1) \leq \frac{875}{3n} \approx \frac{291,7}{n}$$

Question 2

Pour tout $n \geq 1$: $0 \leq P(|M_n - 3,5| \geq 0,1) \leq \frac{875}{3n}$. Comme $\frac{875}{3n} \rightarrow 0$, d'après le théorème des gendarmes :

$$\lim_{n \rightarrow +\infty} P(|M_n - 3,5| \geq 0,1) = 0$$

C'est la loi faible des grands nombres.

Question 3

$$\frac{875}{3n} \leq 0,05 \iff n \geq \frac{875}{0,15} \approx 5\,833,3$$

Il suffit de $n = 5\,834$ lancers.

Exercice 3 (★★★) : estimer une probabilité par simulation

On veut estimer par simulation une probabilité inconnue p . On répète n fois une expérience indépendante et on note F_n la fréquence de réalisations de l'événement.

Déterminer une valeur de n garantissant, avec une probabilité d'au moins 0,95, que F_n soit à moins de 0,01 de p , quelle que soit la valeur de p .

Correction de l'exercice 3

Le nombre de réalisations suit $\mathcal{B}(n, p)$, donc $E(F_n) = p$ et $V(F_n) = \frac{p(1-p)}{n} \leq \frac{1}{4n}$.

D'après l'inégalité de concentration avec $\delta = 0,01$:

$$P(|F_n - p| \geq 0,01) \leq \frac{1}{4n \times 0,0001} = \frac{2\,500}{n}$$

$$\text{On veut } \frac{2\,500}{n} \leq 0,05 \iff n \geq \frac{2\,500}{0,05} = 50\,000.$$

Il suffit de $n = 50\,000$ simulations.

Remarque : pour une précision dix fois meilleure ($\delta = 0,001$), il faudrait cent fois plus de simulations (5×10^6) : la précision ne progresse qu'en $\frac{1}{\sqrt{n}}$.

Exercice 4 (★★) : une loi binomiale

Soit X une variable de loi $\mathcal{B}(1\,000; 0,3)$ et $F = \frac{X}{1\,000}$.

1. Calculer $E(F)$ et $V(F)$.
2. Minorer, avec l'inégalité de Bienaymé-Tchebychev, $P(0,25 < F < 0,35)$.
3. La valeur exacte est d'environ 0,9994. Commenter.

Correction de l'exercice 4

Question 1

$$E(X) = 300 \text{ et } V(X) = 1\,000 \times 0,3 \times 0,7 = 210.$$

$$E(F) = 0,3$$

$$V(F) = \frac{210}{10^6} = 0,00021$$

Question 2

D'après l'inégalité avec $\delta = 0,05$:

$$P(|F - 0,3| \geq 0,05) \leq \frac{0,00021}{0,0025} = 0,084$$

Par événement contraire :

$$P(0,25 < F < 0,35) \geq 1 - 0,084 = 0,916$$

Question 3

La valeur exacte **0,9994** est bien supérieure au minorant **0,916** : l'inégalité est **vraie mais prudente**. L'inégalité de Bienaymé-Tchebychev ne demande que E et V , pas la loi.

Exercice 5 (★★) : un risque d'assurance

Une compagnie assure **1 000** contrats indépendants. Chaque contrat déclare un sinistre dans l'année avec la probabilité **0,02**. Soit N le nombre de sinistres.

1. Donner la loi de N , son espérance et sa variance.
2. Majorer la probabilité qu'il y ait au moins **40** sinistres.
3. Commenter.

Correction de l'exercice 5

Question 1

On répète **1 000** fois, de façon identique et indépendante, une épreuve de Bernoulli de succès « le contrat déclare un sinistre », de probabilité **0,02**.

N suit $\mathcal{B}(1\,000; 0,02)$.

$$E(N) = 1\,000 \times 0,02 = 20$$

$$V(N) = 1\,000 \times 0,02 \times 0,98 = 19,6$$

Question 2

$N \geq 40$ entraîne $N - 20 \geq 20$, donc $|N - 20| \geq 20$. D'après l'inégalité de Bienaymé-Tchebychev avec $\delta = 20$:

$$P(N \geq 40) \leq P(|N - 20| \geq 20) \leq \frac{19,6}{400}$$

$$P(N \geq 40) \leq 0,049$$

Question 3

La compagnie a au plus **4,9%** de risque d'avoir au moins deux fois plus de sinistres que prévu. La valeur exacte est de l'ordre de 4×10^{-5} : la majoration est prudente, mais garantie sans calcul de loi binomiale.

Exercice 6 (★★) : la simulation en Python

On considère :

```
import random

def frequence_pile(n):
    compteur = 0
    for _ in range(n):
        if random.random() < 0.5:
            compteur = compteur + 1
    return compteur / n
```

1. Que représente `frequence_pile(10000)` ?
2. Avec quelle probabilité minimale le résultat est-il strictement compris entre 0,48 et 0,52 ?

Correction de l'exercice 6

Question 1

La fonction simule n lancers d'une pièce équilibrée (avec `random.random() < 0.5`) et renvoie la fréquence de « pile ». Pour $n = 10\,000$, c'est une réalisation de F_{10000} .

Question 2

$E(F_n) = 0,5$ et $V(F_n) = \frac{0,25}{10\,000}$. D'après l'inégalité de concentration avec $\delta = 0,02$:

$$P(|F_n - 0,5| \geq 0,02) \leq \frac{0,25}{10\,000 \times 0,0004} = 0,0625$$

$$P(0,48 < F_n < 0,52) \geq 1 - 0,0625 = 0,9375$$

Conclusion : le résultat du script est, avec une probabilité d'au moins **93,75%**, strictement compris entre 0,48 et 0,52.

Exercice 7 (★★) : vrai ou faux**Justifier chaque réponse :**

1. La loi des grands nombres affirme que la moyenne M_n est égale à $E(X)$ pour n assez grand.
2. Après dix « pile » de suite, la probabilité d'obtenir « face » est supérieure à $\frac{1}{2}$.
3. L'écart-type de la moyenne d'un échantillon est proportionnel à $\frac{1}{\sqrt{n}}$.
4. Pour estimer une proportion à 0,01 près avec une probabilité d'au moins 0,95, 50 000 observations suffisent (avec la majoration de Bienaymé-Tchebychev).

Correction de l'exercice 7

Affirmation 1

La loi des grands nombres dit que la **probabilité** d'un écart supérieur à δ tend vers 0 : M_n reste aléatoire, seulement de plus en plus concentrée autour de $E(X)$.

Faux.

Affirmation 2

Les lancers sont indépendants : la pièce n'a pas de mémoire. La probabilité de « face » reste $\frac{1}{2}$.

Faux.

Affirmation 3

$$\sigma(M_n) = \frac{\sigma}{\sqrt{n}}.$$

Vrai.

Affirmation 4

C'est le résultat de l'exercice 3.

Vrai.

Ce qui peut tomber au bac

- Une **conclusion statistique** justifiée par l'inégalité de concentration (exercice 1).
- La **loi des grands nombres** avec limite et gendarmes (exercice 2).
- La **taille d'un échantillon** ou d'une simulation (exercices 2, 3).
- L'**inégalité** appliquée à une loi binomiale (exercices 4 et 5).
- Un **script de simulation** (exercice 6).

QCM : une seule bonne réponse par question

1. Pour la moyenne M_n de n variables indépendantes de même loi (écart-type σ), $\sigma(M_n)$ vaut :

- A. σ
- B. $\sigma\sqrt{n}$
- C. $\frac{\sigma}{n}$
- D. $\frac{\sigma}{\sqrt{n}}$ (bonne réponse)

Justification : $V(M_n) = \frac{V(X)}{n}$, donc l'écart-type est divisé par $\sqrt{n}$.

2. Pour diviser par 2 l'écart-type de la moyenne d'un échantillon, il faut multiplier la taille n par :

- A. 4 (bonne réponse)
- B. 2
- C. 8
- D. 16

Justification : $\frac{\sigma}{\sqrt{4n}} = \frac{1}{2} \times \frac{\sigma}{\sqrt{n}}$.

3. Pour estimer une proportion à 0,03 près avec une probabilité d'au moins 0,95, l'inégalité de concentration garantit que $n = 5\,556$ suffit. Comment obtient-on ce nombre ?

- A. $\frac{1}{4n \times 0,03^2} \leq 0,05$ (bonne réponse)
- B. $n \times 0,03 \leq 0,95$
- C. $\frac{1}{n \times 0,03} \leq 0,05$
- D. $n = \frac{0,95}{0,03}$

Justification : On utilise $p(1-p) \leq \frac{1}{4} : n \geq \frac{1}{4 \times 0,0009 \times 0,05} \approx 5\,555,6$.

4. La loi faible des grands nombres affirme que, pour tout $\delta > 0$:

- A. $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$ (bonne réponse)
- B. $M_n = \mu$ pour n assez grand

C. $P(|M_n - \mu| \geq \delta) = 1$

D. M_n tend vers 0

Justification : C'est la probabilité d'un écart qui tend vers 0 : M_n reste aléatoire.

5. On lance une pièce équilibrée et on obtient dix « pile » de suite. Quelle est la probabilité d'obtenir « face » au lancer suivant ?

A. $\frac{1}{2}$ (**bonne réponse**)

B. plus de $\frac{1}{2}$

C. moins de $\frac{1}{2}$

D. 1

Justification : Les lancers sont indépendants : la pièce n'a pas de mémoire.

6. Dans l'exercice 1, pour $n = 10\,000$ et une fréquence observée $F_n = 0,54$, quelle majoration obtient-on pour $P(|F_n - 0,5| \geq 0,04)$ si la pièce est équilibrée ?

A. 0,04

B. **environ 0,0156 (bonne réponse)**

C. 0,5

D. environ 0,156

Justification : $\frac{0,25}{10\,000 \times 0,0016} \approx 0,0156$: un écart aussi grand serait rare pour une pièce équilibrée.

7. Dans l'exercice 2, quelle est la limite de $P(|M_n - 3,5| \geq 0,1)$ quand $n \rightarrow +\infty$?

A. **0 (bonne réponse)**

B. Elle n'a pas de limite

C. 1

D. 0,5

Justification : Par encadrement (gendarmes), avec le majorant $\frac{875}{3n} \rightarrow 0$, la probabilité tend vers 0.

8. Dans l'exercice 3, combien de simulations faut-il pour estimer une probabilité à 0,01 près avec une confiance de 95 %, quelle que soit la vraie probabilité ?

A. **$n = 50\,000$ (bonne réponse)**

B. $n = 500\,000$

C. $n = 10\,000$

D. $n = 5\,000$

Justification : On utilise $p(1-p) \leq \frac{1}{4} : n \geq \frac{1}{4 \times 0,05 \times 0,0001} = 50\,000$.

9. Dans l'exercice 5 (assurance, 1 000 contrats, $p = 0,02$), quelle majoration obtient-on pour $P(N \geq 40)$?

- A. 0,02
- B. 0,2
- C. 0,049 (bonne réponse)
- D. 0,5

Justification : $N \geq 40$ entraîne $|N - 20| \geq 20$; l'inégalité donne $P(|N - 20| \geq 20) \leq \frac{19,6}{400} \approx 0,049$.

10. Dans l'exercice 7, l'affirmation « après dix "pile" de suite, la probabilité d'obtenir "face" est supérieure à $\frac{1}{2}$ » est-elle vraie ?

- A. Oui, c'est la loi des grands nombres
- B. **Non : les lancers sont indépendants, la probabilité reste $\frac{1}{2}$ (bonne réponse)**
- C. Cela dépend du nombre de lancers précédents
- D. Oui, car la pièce doit se rééquilibrer

Justification : C'est l'illusion classique du « joueur » : une pièce n'a pas de mémoire, chaque lancer reste indépendant des précédents.