

Plaquette 4 - Loi des grands nombres et échantillons

Concentration, loi des grands nombres — Terminale spé maths

Corrections détaillées

Boîte à outils

Échantillon : $X_1, \dots, X_n$ indépendantes, de même loi (espérance μ , variance V).

Somme $S_n = X_1 + \dots + X_n$ et moyenne $M_n = \frac{S_n}{n}$:

$$E(M_n) = \mu \quad V(M_n) = \frac{V}{n} \quad \sigma(M_n) = \frac{\sigma}{\sqrt{n}}$$

Inégalité de concentration (Bienaymé-Tchebychev appliquée à M_n) : pour tout $\delta > 0$,

$$P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$$

Loi (faible) des grands nombres : pour tout $\delta > 0$,

$$\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$$

La moyenne d'un échantillon **se rapproche** de l'espérance quand sa taille augmente.

Taille d'échantillon : pour un risque au plus α et une précision δ , il suffit que $n \geq \frac{V}{\alpha\delta^2}$.

Cas particulier : pour des variables de Bernoulli de paramètre p , M_n est la **fréquence** des succès, et l'on retrouve la concentration des fréquences.

Correction de l'exercice 1 - Moyenne d'un échantillon

a) **Linéarité** : $E(S_n) = E(X_1) + \dots + E(X_n) = n\mu$, donc $E(M_n) = \frac{1}{n} \times n\mu = \mu$.

Indépendance : $V(S_n) = V(X_1) + \dots + V(X_n) = nV$, donc

$$V(M_n) = \frac{1}{n^2} V(S_n) = \frac{nV}{n^2} = \frac{V}{n}$$

b) On applique **Bienaymé-Tchebychev** à M_n :

$$- P(|M_n - \mu| \geq \delta) \leq \frac{V(M_n)}{\delta^2} = \frac{V}{n\delta^2}.$$

— **À retenir** : l'indépendance est indispensable pour additionner les variances ; sans elle, la moyenne ne se concentre pas forcément.

Réponse. $E(M_n) = \mu$, $V(M_n) = \frac{V}{n}$, puis Bienaymé-Tchebychev.

Correction de l'exercice 2 - Démontrer la loi des grands nombres

Pour tout $n \geq 1$, une probabilité est positive et l'inégalité de concentration la majore :

$$0 \leq P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$$

— δ et V sont fixés, donc $\frac{V}{n\delta^2} \rightarrow 0$ quand $n \rightarrow +\infty$.

— Par le **théorème des gendarmes**, la probabilité tend vers 0.

- **Interprétation** : aussi petite que soit la précision δ exigée, la probabilité que la moyenne s'en écarte devient aussi petite qu'on veut en prenant un échantillon assez grand.
- **Nuance** : le résultat porte sur une **probabilité** ; il n'affirme pas que chaque suite de résultats converge « à coup sûr » (c'est la loi *forte*, hors programme).

Réponse. Encadrement entre 0 et $\frac{V}{n\delta^2} \rightarrow 0$: gendarmes.

Correction de l'exercice 3 - Durée de vie d'ampoules

a) $E(M) = 1\,000$ h et

$$\sigma(M) = \frac{100}{\sqrt{50}} \approx 14,1 \text{ h}$$

b) **Inégalité de concentration** avec $V = 10\,000$, $n = 50$, $\delta = 30$:

$$- P(|M - 1\,000| \geq 30) \leq \frac{10\,000}{50 \times 900} \approx 0,222.$$

- **Interprétation** : dans au moins 78 % des tests, la durée moyenne mesurée sur 50 ampoules est à moins de 30 h de la vraie valeur, alors qu'une ampoule seule s'en écarte typiquement de 100 h.

Réponse. a) 1 000 h ; $\approx 14,1$ h · b) $\leq 0,222$.

Correction de l'exercice 4 - Estimer π au hasard

a) Le point est uniforme dans le carré d'aire 1 ; la probabilité d'être dans le quart de disque est le rapport des aires, $\frac{\pi \times 1^2/4}{1} = \frac{\pi}{4}$.

b) Les X_i sont des Bernoulli de paramètre $p = \frac{\pi}{4} \approx 0,785$:

$$E(4M_n) = 4p = \pi \quad V(4M_n) = 16 \times \frac{p(1-p)}{n} \approx 16 \times \frac{0,785 \times 0,215}{10\,000} \approx 0,000\,27$$

c) **Bienaymé-Tchebychev** :

$$- P(|4M_n - \pi| \geq 0,05) \leq \frac{0,000\,27}{0,002\,5} \approx 0,108.$$

- **Interprétation** : avec 10 000 points, l'estimation de π est à $\pm 0,05$ près dans au moins 89 % des cas. C'est la **méthode de Monte-Carlo**.

Réponse. b) $E = \pi$, $V \approx 0,000\,27$ · c) $\leq 0,108$.

Correction de l'exercice 5 - Le poids moyen de pommes

a) **Concentration** avec $V = 400$, $n = 100$, $\delta = 5$:

$$P(|M - \mu| \geq 5) \leq \frac{400}{100 \times 25} = 0,16$$

b) On veut $\frac{400}{n \times 25} \leq 0,05 \Leftrightarrow n \geq \frac{400}{25 \times 0,05} = 320$.

- Il faut peser au moins 320 **pommes**.

- **Remarque** : on n'a pas besoin de connaître μ (c'est justement ce qu'on cherche) ; seule la **variance** intervient dans la garantie.

Réponse. a) $\leq 0,16$ · b) $n \geq 320$.

Correction de l'exercice 6 - Taille minimale d'un échantillon

On veut $P(|M_n - \mu| \geq 3) \leq 0,01$. Par l'inégalité de concentration, il suffit que

$$\frac{V}{n\delta^2} = \frac{225}{9n} = \frac{25}{n} \leq 0,01 \Leftrightarrow n \geq 2\,500$$

- Un échantillon de 2 500 individus suffit.
- **Lecture de la formule** $n \geq \frac{V}{\alpha\delta^2}$: plus la grandeur est dispersée (V grand), plus la précision exigée est fine (δ petit) et plus le risque toléré est faible (α petit), plus il faut de données.

Réponse. $n \geq 2\,500$.

Correction de l'exercice 7 - Somme contre moyenne

a) Formules pour n variables indépendantes :

$$\sigma(S_n) = 2\sqrt{n} \quad \sigma(M_n) = \frac{2}{\sqrt{n}}$$

- $n = 100$: $\sigma(S) = 20$ et $\sigma(M) = 0,2$.
- $n = 10\,000$: $\sigma(S) = 200$ et $\sigma(M) = 0,02$.

b) L'écart typique entre S_n et $10n$ **augmente** (de 20 à 200) : la somme ne se « rapproche » pas de $10n$ en valeur absolue.

L'écart typique entre M_n et 10 **diminue** : c'est la **moyenne** qui se concentre, conformément à la loi des grands nombres.

- **Piège** : croire que la somme des gains d'un joueur finit par valoir « exactement » n fois l'espérance.

Réponse. a) 20 et 0,2 ; 200 et 0,02 · b) la moyenne se concentre, pas la somme.

Correction de l'exercice 8 - Le casino gagne à la longue

a) Être gagnant signifie $S_n \geq 0$, soit $M_n \geq 0$, donc $M_n - (-0,1) \geq 0,1$, et a fortiori $|M_n + 0,1| \geq 0,1$.

b) **Concentration** avec $V = 4$ et $\delta = 0,1$:

$$P(\text{gagnant}) \leq P(|M_n + 0,1| \geq 0,1) \leq \frac{4}{n \times 0,01} = \frac{400}{n}$$

- Pour $n = 10\,000$: au plus 4 %.
- **Interprétation** : sur une partie, un joueur gagne souvent ; sur des milliers de parties, il est presque sûr de perdre. Pour le casino, qui joue des millions de parties, le gain moyen 0,10 € par partie est **quasi certain**.

Réponse. b) $\leq \frac{400}{n}$, soit $\leq 0,04$ pour $n = 10\,000$.

Correction de l'exercice 9 - Mutualiser les risques

La mutuelle perd de l'argent si le remboursement moyen M dépasse 820 €, ce qui implique $|M - 800| \geq 20$.

a) $n = 10\,000$:

$$P(M \geq 820) \leq \frac{1\,000^2}{10\,000 \times 20^2} = \frac{1\,000\,000}{4\,000\,000} = 0,25$$

b) $n = 100\,000$: le majorant est divisé par 10, soit 0,025.

- **Interprétation** : un adhérent seul a des dépenses très imprévisibles ($\sigma = 1\,000$ €), mais la **moyenne** sur un grand nombre d'adhérents est très stable. C'est le principe de la **mutualisation**, fondement de l'assurance.

Réponse. a) $\leq 0,25$ · b) $\leq 0,025$.

Correction de l'exercice 10 - Fréquence et moyenne

a) X_i suit une loi de Bernoulli de paramètre $\frac{1}{6}$. La moyenne $M_n = \frac{X_1 + \dots + X_n}{n}$ est la **fréquence** des 6 parmi les n lancers.

b) $V(X_i) = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36}$. **Concentration** :

$$P\left(M_n - \frac{1}{6} \geq 0,05\right) \leq \frac{5/36}{n \times 0,0025} = \frac{55,6}{n}$$

- $n = 60$: majorant $\approx 0,93$, presque inutile ;
- $n = 6\,000$: majorant $\approx 0,009$.
- **Lien** : la loi des grands nombres appliquée à des variables 0 / 1 explique pourquoi la fréquence **tend vers** la probabilité — l'intuition qui fonde la notion même de probabilité.

Réponse. a) Bernoulli ; M_n est la fréquence des 6 · b) $\approx 0,93$ puis $\approx 0,009$.

Correction de l'exercice 11 - Moyenner des mesures

La moyenne M_n des mesures a pour espérance la vraie longueur ℓ (erreur d'espérance nulle) et pour variance $\frac{0,64}{n}$.

a) $n = 25$:

$$P(|M_{25} - \ell| \geq 0,4) \leq \frac{0,64}{25 \times 0,16} = 0,16$$

b) $n = 100$: $\frac{0,64}{100 \times 0,16} = 0,04$.

- **Interprétation** : en physique, on répète les mesures et on en fait la moyenne précisément pour **réduire l'incertitude** ; la loi des grands nombres le justifie.
- **Limite** : cela ne corrige pas une erreur **systématique** (un appareil mal étalonné), qui déplace l'espérance elle-même.

Réponse. a) $\leq 0,16$ · b) $\leq 0,04$.

Correction de l'exercice 12 - Pas de compensation

a) Les tours étant **indépendants**, la probabilité de « rouge » reste $\frac{18}{37}$, quels que soient les résultats passés.

b) La loi des grands nombres affirme que la **fréquence** des « rouge » se rapproche de $\frac{18}{37}$ quand le nombre de tours augmente :

$$\lim_{n \rightarrow +\infty} P\left(F_n - \frac{18}{37} \geq \delta\right) = 0$$

- Elle ne dit **rien** sur le prochain tour. La série de 10 noirs ne sera pas « compensée » : elle sera **diluée** dans les milliers de tours suivants, où rouge et noir sortiront dans les proportions habituelles.

— **Erreur du joueur** : attribuer à la roulette une « mémoire » qu'elle n'a pas.

Réponse. a) $\frac{18}{37}$ · b) les écarts sont dilués, pas compensés.

Correction de l'exercice 13 - Deux instituts de mesure

On compare les variances des moyennes, $\frac{V}{n}$:

$$V(M_A) = \frac{4}{40} = 0,1 \quad V(M_B) = \frac{1}{10} = 0,1$$

— Les deux moyennes ont **la même variance** : même précision.

— **Majorants** : $\frac{0,1}{0,25} = 0,4$ dans les deux cas.

— **Interprétation** : un appareil deux fois plus précis (σ divisé par 2) équivaut à **quatre fois plus** de mesures avec l'appareil moins précis.

— **Choix pratique** : cela dépend du coût d'une mesure et du coût d'un meilleur appareil.

Réponse. Même précision : $V(M_A) = V(M_B) = 0,1$, majorants 0,4.

Correction de l'exercice 14 - Moyenne de lancers et seuil

a) **Concentration** :

$$P(|M_n - 3,5| \geq 0,2) \leq \frac{35/12}{n \times 0,04} = \frac{35}{0,48n} \approx \frac{72,9}{n}$$

b) $\frac{72,9}{n} \leq 0,01 \Leftrightarrow n \geq 7\,292$ environ (plus précisément $n \geq \frac{35}{0,0048} \approx 7\,291,7$).

— Il faut au moins 7 292 **lancers**.

— **Interprétation** : avec 7 292 lancers, dans au moins 99 % des cas, la moyenne est comprise strictement entre 3,3 et 3,7.

Réponse. a) $\leq \frac{35}{0,48n}$ · b) $n \geq 7\,292$.

Correction de l'exercice 15 - Pourquoi l'indépendance compte

a) Toutes les valeurs sont égales, donc $M_n = X_1$: la moyenne est simplement le résultat du dé.

$$V(M_n) = V(X_1) = \frac{35}{12}$$

qui ne diminue **pas** quand n augmente.

b) **Non** : $P(|M_n - 3,5| \geq 2) = P(X_1 \in \{1; 6\}) = \frac{1}{3}$ pour tout n , elle ne tend pas vers 0.

— L'hypothèse manquante est l'**indépendance** : les X_i sont identiques, pas indépendantes ; on n'a qu'une seule information répétée.

— **Moralité** : un grand échantillon n'est utile que si ses observations apportent des informations **indépendantes** (sondage sans doublons, mesures répétées réellement).

Réponse. a) $M_n = X_1$, $V = \frac{35}{12}$ · b) non, faute d'indépendance.