Chapitre 10

Loi des grands nombres

10.1 Introduction historique

> *On ne peut donc pas douter que la loi des grands nombres ne convienne aux choses morales qui dépendent de la volonté de l'homme, de ses intérêts, de ses lumières et de ses passions.*
>
> Siméon Denis Poisson - *Recherches sur la probabilité des jugements* (1837)

Chacun sait qu'en lançant un dé un grand nombre de fois, la fréquence du 6 finit par s'approcher de $\frac{1}{6}$. Cette conviction est si ancrée qu'elle paraît évidente — et pourtant, elle demande une démonstration. Pourquoi le hasard, imprévisible à chaque lancer, devient-il presque certain lorsqu'on le répète ? Ce chapitre répond à cette question, et surtout la rend *quantitative* : il ne dira pas seulement que la moyenne se rapproche, il dira à quelle vitesse.

10.1.1 Le théorème d'or de Bernoulli

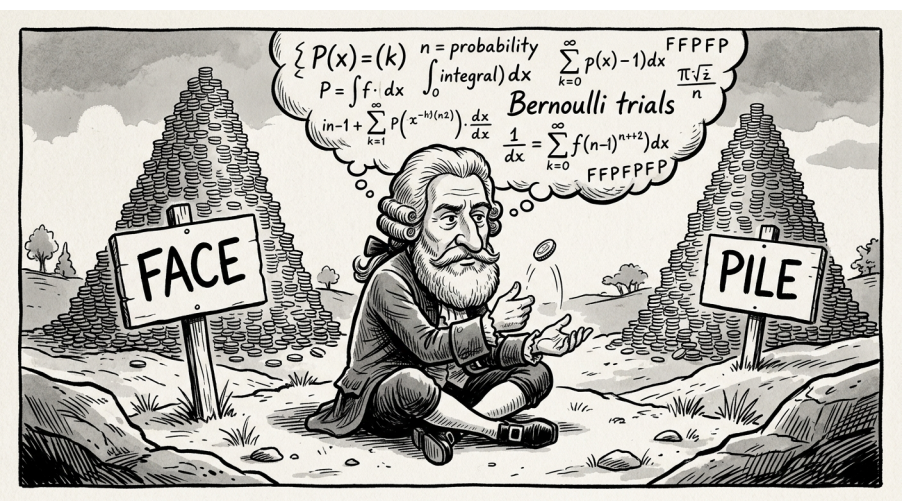
C'est **Jacques Bernoulli** qui, dans l'*Ars conjectandi* publié en 1713, démontre le premier résultat de ce type, qu'il appelait lui-même son « théorème d'or » et auquel il consacra, dit-on, une vingtaine d'années. Il y établit que la fréquence observée d'un événement se rapproche de sa probabilité lorsque le nombre d'épreuves augmente.

L'expression *loi des grands nombres* n'apparaît toutefois qu'un siècle et demi plus tard, sous la plume de **Siméon Denis Poisson**, qui l'emploie dans ses travaux des années 1835 à 1837. Il y défend, comme le montre la citation ci-dessus, l'idée que cette loi ne gouverne pas seulement les jeux de hasard, mais aussi les phénomènes sociaux.

10.1.2 L'école russe et les inégalités

Restait à mesurer l'écart entre la moyenne observée et la valeur théorique. C'est l'apport décisif de l'école russe de probabilités : **Pafnouti Tchebychev** et son élève **Andreï Markov** établissent au XIX^{e} siècle des inégalités remarquablement simples, qui majorent la probabilité de s'écarter beaucoup de la moyenne. Le mathématicien français **Irénée-Jules Bienaymé** avait obtenu indépendamment l'une d'elles, d'où le nom qu'on lui donne aujourd'hui.

Ces inégalités sont l'ossature de ce chapitre : nous les établirons l'une après l'autre, chacune découlant de la précédente, jusqu'à la loi des grands nombres qui en sera le couronnement.



10.2 Échantillon d'une variable aléatoire

Définition : Soit X une variable aléatoire et n un entier naturel non nul. On appelle **échantillon de taille n** de la loi de X toute liste $(X_1 ; X_2 ; \dots ; X_n)$ de variables aléatoires :

- **indépendantes** les unes des autres ;
- suivant toutes la **même loi** que X.

On lui associe la **somme** $S_n = X_1 + X_2 + \cdots + X_n$ et la **moyenne** $M_n = \dfrac{S_n}{n}$.

Remarque : Un échantillon modélise donc la *répétition* d'une même expérience dans les mêmes conditions : n lancers d'un même dé, n personnes interrogées dans un sondage, n pièces prélevées sur une chaîne de production. La variable M_n est le résultat moyen observé, et c'est son comportement qui nous intéresse.

Proposition 1 : Soit $(X_1 ; \dots ; X_n)$ un échantillon de taille n de la loi de X, d'espérance μ et de variance σ^2. Alors :

$$E(S_n) = n\mu \quad ; \quad V(S_n) = n\sigma^2$$

$$E(M_n) = \mu \quad ; \quad V(M_n) = \frac{\sigma^2}{n}$$

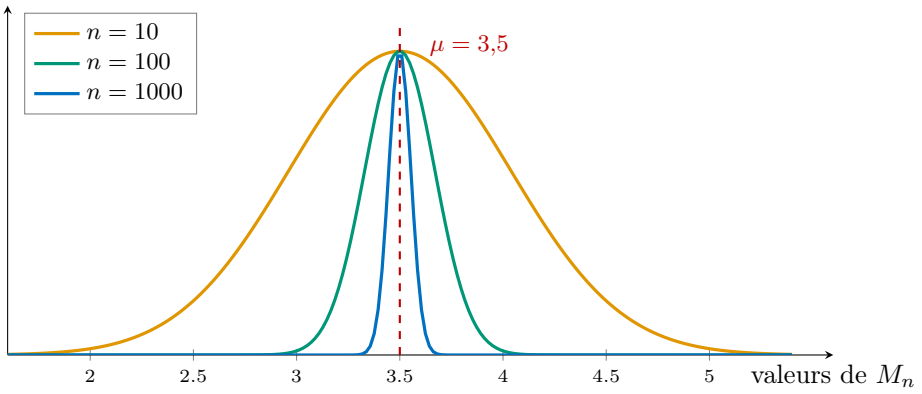
∘ Démonstration : L'espérance étant linéaire, $E(S_n) = E(X_1) + \cdots + E(X_n) = n\mu$. La variance d'une somme de variables *indépendantes* étant la somme des variances — résultat admis —, on a de même $V(S_n) = n\sigma^2$.

Pour la moyenne, on utilise les propriétés vues au chapitre 2 : $E(aX) = aE(X)$ et $V(aX) = a^2V(X)$, avec $a = \frac{1}{n}$:

$$E(M_n) = \frac{1}{n} \times n\mu = \mu \qquad \text{et} \qquad V(M_n) = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}$$

□

Remarque : Ces deux dernières formules contiennent toute la suite du chapitre. La moyenne d'un échantillon a *la même espérance* que X : elle vise juste. Mais sa variance est *divisée par* n : elle est d'autant plus resserrée autour de μ que l'échantillon est grand. Autrement dit, augmenter la taille de l'échantillon ne déplace pas la cible, cela réduit la dispersion des tirs.



Ces trois courbes représentent la répartition de la moyenne M_n pour des lancers d'un dé équilibré, dont l'espérance vaut 3,5. Le sommet reste au même endroit, mais la courbe se resserre à mesure que n grandit : c'est exactement ce que traduit la formule $V(M_n) = \frac{\sigma^2}{n}$.

10.3 Inégalité de Markov

Théorème 1 : Inégalité de Markov :
Soit X une variable aléatoire ne prenant que des valeurs *positives*, et a un réel strictement positif. Alors :

$$P(X \geqslant a) \leqslant \frac{E(X)}{a}$$

◦ Démonstration : Notons x_1, …, x_k les valeurs prises par X, toutes positives, et p_1, …, p_k leurs probabilités. Par définition :

$$E(X) = \sum_{i=1}^{k} p_i x_i$$

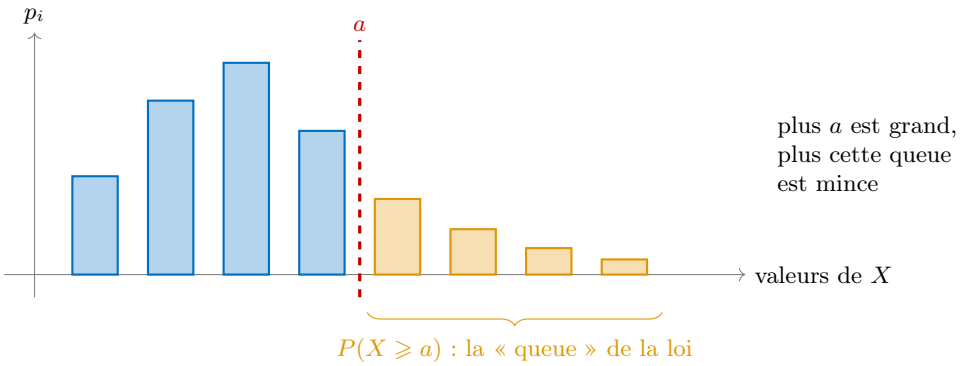
Tous les termes de cette somme étant positifs, elle est supérieure à la somme partielle obtenue en ne conservant que les valeurs supérieures ou égales à a :

$$E(X) \geqslant \sum_{i\,:\,x_i \geqslant a} p_i x_i$$

Or, dans cette dernière somme, chaque x_i est supérieur ou égal à a, donc :

$$\sum_{i\,:\,x_i \geqslant a} p_i x_i \geqslant \sum_{i\,:\,x_i \geqslant a} p_i a = a \times P(X \geqslant a)$$

On obtient donc $E(X) \geqslant a\,P(X \geqslant a)$, puis le résultat en divisant par a, qui est strictement positif. □



L'inégalité de Markov majore donc le poids de cette queue à l'aide de la seule espérance : une variable dont la moyenne est petite ne peut pas prendre souvent de très grandes valeurs.

Exemple : Dans une file d'attente, le temps d'attente T, exprimé en minutes, est une variable aléatoire positive d'espérance 4 minutes. Quelle est la probabilité d'attendre au moins 20 minutes ?

L'inégalité de Markov donne directement :

$$P(T \geqslant 20) \leqslant \frac{E(T)}{20} = \frac{4}{20} = 0{,}2$$

Au plus 20 % des clients attendent donc un quart d'heure de plus que la moyenne.

Remarque : La majoration est souvent grossière, et c'est le prix de sa généralité : elle ne suppose *rien* sur la loi de X, hormis sa positivité. On ne l'utilise donc presque jamais telle quelle, mais elle sert de tremplin à l'inégalité de la section suivante, bien plus fine.

10.4 Inégalité de Bienaymé-Tchebychev

Théorème 2 : Inégalité de Bienaymé-Tchebychev :
Soit X une variable aléatoire d'espérance μ et de variance σ^2, et soit δ un réel strictement positif. Alors :

$$P\big(|X - \mu| \geqslant \delta\big) \leqslant \frac{\sigma^2}{\delta^2}$$

◦ Démonstration : Posons $Y = (X - \mu)^2$. Cette variable aléatoire ne prend que des valeurs positives, et son espérance vaut, par définition de la variance :

$$E(Y) = E\big((X - \mu)^2\big) = V(X) = \sigma^2$$

Appliquons-lui l'inégalité de Markov avec le seuil $a = \delta^2$:

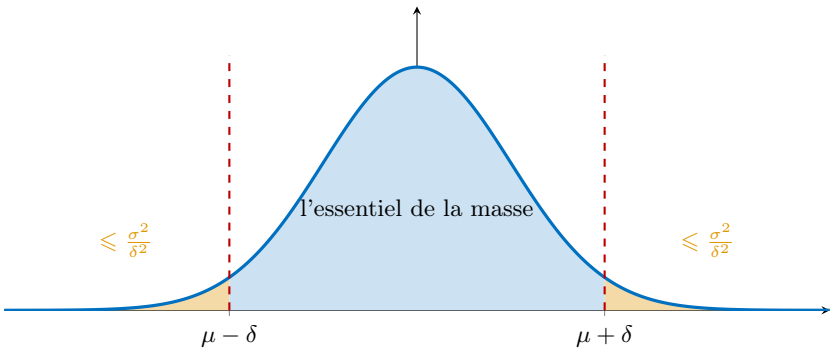
$$P\big(Y \geqslant \delta^2\big) \leqslant \frac{\sigma^2}{\delta^2}$$

Or, la fonction carré étant croissante sur les positifs :

$$Y \geqslant \delta^2 \iff (X-\mu)^2 \geqslant \delta^2 \iff |X-\mu| \geqslant \delta$$

Les deux événements sont donc identiques, et ont par conséquent la même probabilité. □

Remarque : L'événement $|X-\mu| \geqslant \delta$ se lit : « X s'écarte de son espérance d'au moins δ ». L'inégalité affirme donc que la probabilité d'un tel écart est d'autant plus faible que la variance est petite — ce qui est bien le sens de la variance, mesure de dispersion.



Les deux zones orangées représentent l'événement « s'écarter d'au moins δ » : c'est leur poids total que l'inégalité majore par $\frac{\sigma^2}{\delta^2}$.

Exemple : Une machine produit des pièces dont la masse X, en grammes, a pour espérance 100 et pour variance 25. Majorons la probabilité qu'une pièce s'écarte de plus de 10 grammes de la masse moyenne :

$$P\big(|X-100| \geqslant 10\big) \leqslant \frac{25}{10^2} = \frac{25}{100} = 0{,}25$$

Au plus un quart des pièces sortent donc de l'intervalle $[90\,;110]$.

10.5 Inégalité de concentration

Théorème 3 : Inégalité de concentration :
Soit $(X_1\,;\ldots\,;X_n)$ un échantillon de taille n d'une variable aléatoire d'espérance μ et de variance σ^2, et M_n sa moyenne. Alors, pour tout réel $\delta > 0$:

$$P\big(|M_n-\mu| \geqslant \delta\big) \leqslant \frac{\sigma^2}{n\,\delta^2}$$

∘ Démonstration : Il suffit d'appliquer l'inégalité de Bienaymé-Tchebychev à la variable aléatoire M_n, dont la proposition 1 donne l'espérance et la variance :

$$E(M_n) = \mu \qquad \text{et} \qquad V(M_n) = \frac{\sigma^2}{n}$$

d'où :

$$P\big(|M_n-\mu| \geqslant \delta\big) \leqslant \frac{V(M_n)}{\delta^2} = \frac{\sigma^2}{n\delta^2}$$

□

Remarque : C'est le résultat central du chapitre, et le seul qui fasse intervenir la taille de l'échantillon. Le majorant comporte n au *dénominateur* : doubler la taille de l'échantillon divise donc par deux la borne obtenue. C'est ce qui rend la répétition efficace.

♡ **Méthode :** Cette inégalité sert le plus souvent à **déterminer une taille d'échantillon suffisante**. Pour garantir que la moyenne s'écarte de moins de δ avec une probabilité d'au moins $1-\alpha$:

1. On traduit l'exigence par l'inégalité $\dfrac{\sigma^2}{n\delta^2} \leqslant \alpha$.
2. On isole n, ce qui donne $n \geqslant \dfrac{\sigma^2}{\alpha\,\delta^2}$.
3. On conclut en prenant le plus petit entier convenable.

Exemple : On lance un dé équilibré n fois. Son espérance vaut $\mu = 3{,}5$ et sa variance $\sigma^2 = \frac{35}{12} \approx 2{,}92$. Combien de lancers faut-il pour que la moyenne obtenue s'écarte de moins de 0,1 de 3,5, avec une probabilité d'au moins 95 % ?

On cherche n tel que la probabilité de l'écart soit inférieure à $\alpha = 0{,}05$:

$$\frac{\sigma^2}{n\delta^2} \leqslant 0{,}05 \quad\Longleftrightarrow\quad n \geqslant \frac{35/12}{0{,}05 \times 0{,}1^2} = \frac{35/12}{0{,}0005} \approx 5833{,}3$$

Il faut donc au moins 5834 lancers. On mesure au passage le prix de la précision : diviser δ par 10 multiplierait ce nombre par 100.

10.6 Loi faible des grands nombres

Théorème 4 : Loi faible des grands nombres :
Soit $(X_1 ; \dots ; X_n)$ un échantillon d'une variable aléatoire d'espérance μ et de variance σ^2, et M_n sa moyenne. Alors, pour tout réel $\delta > 0$:

$$\lim_{n\to+\infty} P\big(|M_n - \mu| \geqslant \delta\big) = 0$$

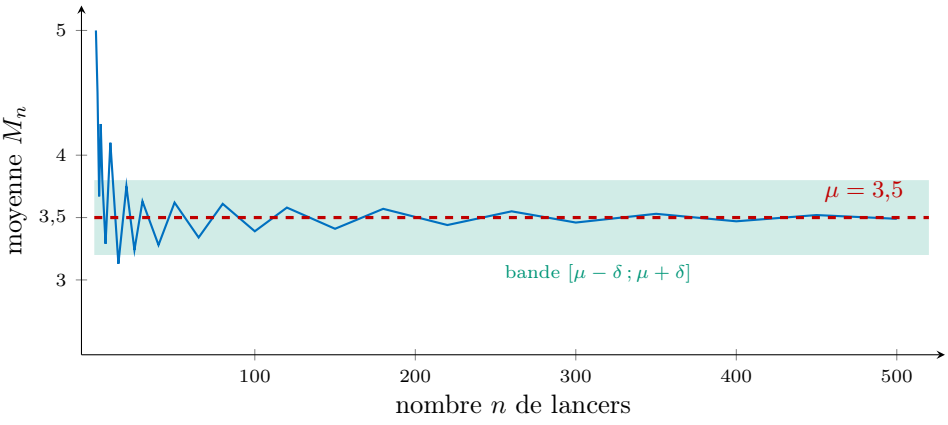
∘ Démonstration : D'après le théorème 3, pour tout entier $n \geqslant 1$:

$$0 \;\leqslant\; P\big(|M_n - \mu| \geqslant \delta\big) \;\leqslant\; \frac{\sigma^2}{n\delta^2}$$

Le majorant est de la forme $\frac{C}{n}$ avec $C = \frac{\sigma^2}{\delta^2}$ constante, et tend donc vers 0 lorsque n tend vers $+\infty$. Le théorème des gendarmes, établi au chapitre 3, permet de conclure. □

Remarque : L'énoncé mérite d'être lu attentivement. Il ne dit pas que la moyenne observée *atteint* μ, ni même qu'elle s'en rapproche à coup sûr : il dit que la probabilité de s'en écarter notablement devient aussi petite que l'on veut. Le seuil δ est fixé d'avance, aussi petit que l'on souhaite, et c'est n qui s'adapte.

On l'appelle loi *faible* car il en existe une version dite forte, hors programme, qui affirme davantage.



Cette simulation illustre le théorème : la moyenne des lancers oscille beaucoup au début, puis se stabilise et finit par ne plus quitter la bande de largeur δ autour de 3,5. La loi des grands nombres affirme précisément que la probabilité d'en sortir tend vers zéro.

Remarque : C'est ce théorème qui fonde les usages les plus courants des probabilités :

- en **sondage**, il justifie qu'un échantillon de quelques milliers de personnes renseigne sur une population de millions ;
- en **assurance**, il explique qu'une compagnie, incapable de prévoir un sinistre individuel, prévoie très bien le coût total de milliers de contrats ;
- en **simulation**, il garantit qu'une quantité inconnue peut s'estimer en répétant un grand nombre de tirages aléatoires.