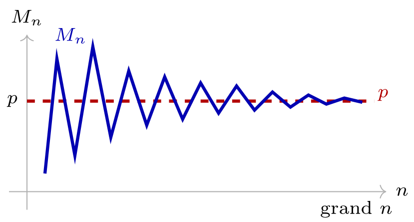
Concentration & loi des grands nombres — fiche révision

À retenir



INÉGALITÉ DE MARKOV Pour X variable aléatoire **positive** et $a > 0$: $P(X \geqslant a) \leqslant \dfrac{E(X)}{a}$.

INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV Pour X d'espérance μ, de variance $V(X)$, et tout $\delta > 0$:
$P(|X-\mu| \geqslant \delta) \leqslant \dfrac{V(X)}{\delta^2}$.

X est d'autant plus **concentrée** autour de μ que $V(X)$ est petite.

INÉGALITÉ DE CONCENTRATION (MOYENNE EMPIRIQUE)
Soient $X_1, \dots, X_n$ indépendantes, **de même loi** (espérance μ, variance $V(X)$). Avec $M_n = \dfrac{X_1 + \cdots + X_n}{n}$: $E(M_n) = \mu$, $V(M_n) = \dfrac{V(X)}{n}$, et pour tout $\delta > 0$:
$P(|M_n - \mu| \geqslant \delta) \leqslant \dfrac{V(X)}{n\,\delta^2}$.

LOI DES GRANDS NOMBRES Conséquence : $P(|M_n - \mu| \geqslant \delta) \longrightarrow 0$ quand $n \to +\infty$.
La **moyenne empirique** M_n se concentre autour de μ.
Pour $X \sim \mathcal{B}(p)$: la **fréquence des succès** tend vers p.

Exemple type

Markov : majoration simple
Une variable X positive vérifie $E(X) = 4$. Majorer $P(X \geqslant 20)$.

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

Bienaymé-Tchebychev
X d'espérance $\mu = 10$ et de variance $V(X) = 4$. Majorer $P(|X-10| \geqslant 5)$.

$$P(|X-10| \geqslant 5) \leqslant \frac{4}{5^2} = \frac{4}{25} = 0{,}16.$$

Concentration d'une fréquence binomiale
On lance n fois une pièce équilibrée ($p = 0{,}5$). M_n = fréquence de Pile. Combien de lancers garantir $P(|M_n - 0{,}5| \geqslant 0{,}05) \leqslant 0{,}05$?

Ici $V(X) = p(1-p) = 0{,}25$. L'inégalité de concentration donne $P(|M_n - 0{,}5| \geqslant 0{,}05) \leqslant \dfrac{0{,}25}{n \times 0{,}05^2} = \dfrac{100}{n}$.
On veut $\dfrac{100}{n} \leqslant 0{,}05$, soit $n \geqslant 2000$.

Pièges classiques

Markov : $X \geqslant 0$ obligatoire : l'inégalité $P(X \geqslant a) \leqslant E(X)/a$ ne vaut que pour une variable **positive**. Sur une variable signée, elle est fausse.
Bornes grossières : Markov et Bienaymé-Tchebychev donnent des bornes **universelles** mais souvent très lâches. Si la loi de X est connue (ex. $\mathcal{B}(n\,;\,p)$), le calcul direct est bien plus précis.
LGN $\neq$ « compensation » : la moyenne M_n s'approche de μ quand n grandit, mais les tirages restent **indépendants** : après une série de Piles, Face n'est **pas** plus probable. C'est l'amplitude relative de l'écart qui décroît, pas un rééquilibrage.