Concentration et loi des grands nombres

Prélude - Historique et contexte

En 1713, dans son *Ars Conjectandi*, Jacques Bernoulli démontre que la fréquence d'un événement se rapproche de sa probabilité quand on répète l'expérience : c'est la première loi des grands nombres. Au XIX^{e} siècle, Bienaymé et Tchebychev donnent une inégalité simple qui permet de quantifier cette concentration. Ces résultats justifient les sondages, les contrôles qualité et les simulations informatiques.

I Inégalité de Bienaymé-Tchebychev

Propriété - Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance μ et de variance V. Pour tout réel $\delta > 0$:

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

La probabilité que X s'écarte de son espérance d'au moins δ est majorée par $\frac{V}{\delta^2}$.

Remarque - Interprétation avec l'écart type

Avec $\delta = k\sigma$: $P\big(|X-\mu| \geqslant k\sigma\big) \leqslant \dfrac{1}{k^2}$. Ainsi, pour toute loi, il y a au moins 75 % de chances que X soit dans $]\mu - 2\sigma\ ;\ \mu + 2\sigma[$, et au moins 89 % pour $]\mu - 3\sigma\ ;\ \mu + 3\sigma[$.
Cette majoration est valable pour toutes les lois ; elle est souvent assez grossière.

Exemple

$X \sim \mathscr{B}(100\ ;\ 0{,}5)$: $\mu = 50$, $V = 25$. $P\big(|X-50| \geqslant 10\big) \leqslant \frac{25}{100} = 0{,}25$. (La valeur exacte est environ $0{,}057$.)
Donc $P(40 < X < 60) \geqslant 0{,}75$.

II Inégalité de concentration

Propriété

Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi d'espérance μ et de variance V, et $M_n = \dfrac{X_1 + \cdots + X_n}{n}$. Pour tout $\delta > 0$:

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

Démonstration : on applique Bienaymé-Tchebychev à M_n, avec $E(M_n) = \mu$ et $V(M_n) = \frac{V}{n}$.

Méthode - Déterminer une taille d'échantillon

On lance une pièce équilibrée n fois; M_n est la fréquence de Pile ($\mu = 0,5$, $V = 0,25$). Pour que $P\big(|M_n - 0,5| \geqslant 0,05\big) \leqslant 0,1$, il suffit que

$$\frac{0,25}{n \times 0,05^2} \leqslant 0,1 \iff n \geqslant \frac{0,25}{0,1 \times 0,0025} = 1\,000.$$

III Loi des grands nombres

Propriété - Loi (faible) des grands nombres

Pour tout $\delta > 0$:

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

Quand la taille de l'échantillon augmente, la moyenne M_n se concentre autour de l'espérance μ.

Remarque - Interprétation

- Pour un événement de probabilité p, la fréquence observée sur n répétitions se rapproche de p : cela justifie l'approche fréquentiste des probabilités vue en Seconde.
- L'espérance d'un gain est la moyenne des gains « à long terme ».

Méthode - Simulation

```
from random import random
def moyenne(n):
    s = 0
    for i in range(n):
        if random() < 0.3:
            s = s + 1
    return s / n
```

Les valeurs renvoyées par `moyenne(10000)` sont très proches de 0,3.