Loi des grands nombres

1 Échantillon de variables indépendantes

Définition 1 : Un échantillon est un ensemble de variables aléatoires réelles $(X_1, ..., X_n)$ indépendantes et de même loi.

La variable aléatoire moyenne de cette échantillon est la variable aléatoire notée M_n ou $\overline{X}$, définie par

$$M_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

Propriété 1 : On a alors $\mathbb{E}(M_n) = \mathbb{E}(X_1)$, $Var(M_n) = \frac{1}{n}Var(X_1)$ et $\sigma(M_n) = \frac{1}{\sqrt{n}}\sigma(X_1)$

■ **Exemple 1 :** On considère une variable aléatoire X qui suit une loi binomiale de paramètre 3 et $\frac{1}{3}$. On rappelle que $\mathbb{E}(X) = 3 \times \frac{1}{3} = 1$ et $Var(X) = 3 \times \frac{1}{3} \times \left(1 - \frac{1}{3}\right) = \frac{2}{3}$. On considère un échantillon $(X_1, \ldots, X_{100})$ de variables aléatoires indépendantes de même loi que X et on note $M_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.

On a alors $\mathbb{E}(M_n) = \mathbb{E}(X) = 1$ et $Var(M_n) = \frac{2}{300}$. ■

R On remarque que lorsque n tend vers $+\infty$, la variance de M_n tend vers 0 alors que l'espérance ne change pas. Cela signifie intuitivement que la variable aléatoire M_n se rapproche d'une variable aléatoire "constante". Ce comportement sera précisé plus en détails dans les parties qui suivent.

2 Concentration et loi des grands nombres

2.1 Inégalité de Bienaymé-Tchebychev

Propriété 2 — Inégalité de Bienaymé-Tchebychev. : Soit X une variable aléatoire réelle. Pour tout réel δ strictement positif

$$\mathbb{P}(|X - E(x)| \geqslant \delta) \leqslant \frac{Var(x)}{\delta^2}$$

R Cette inégalité traduit le fait que la variance permet de mesurer l'écart d'une variable aléatoire par rapport à son espérance.

■ **Exemple 2 :** On lance 180 fois un dé équilibré à 6 faces, numérotées de 1 à 6. On appelle X la variable aléatoire qui donne le nombre de 1 obtenus. X suit une loi binomiale de paramètres 180 et $\frac{1}{6}$. Ainsi, $\mathbb{E}(X) = 180 \times \frac{1}{6} = 30$ et $Var(X) = 180 \times \frac{1}{6} \times \frac{5}{6} = 25$. On souhaite minorer la probabilité que X soit compris entre 21 et 39, c'est-à-dire $\mathbb{P}(|X - \mathbb{E}(X)| < 10)$.

D'après l'inégalité de Bienaymé-Tchebychev,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geqslant 10) \leqslant \frac{Var(X)}{10^2} = \frac{25}{100} = \frac{1}{4}$$

Ainsi, puisque $\mathbb{P}(|X - \mathbb{E}(X)| < 10) + \mathbb{P}(|X - \mathbb{E}(X)| \geqslant 10) = 1$, on a que $\mathbb{P}(|X - \mathbb{E}(X)| < 10) = 1 - \mathbb{P}(|X - \mathbb{E}(X)| \geqslant 10)$.

Ainsi, $\mathbb{P}(|X - \mathbb{E}(X)| < 10) \geqslant \frac{3}{4}$. ■

R Cette borne n'est pas toujours optimale. En l'occurence, en faisant les calculs précisément, on s'aperçoit que $\mathbb{P}(|X - \mathbb{E}(X)| < 10) \simeq 0.9434$.

2.2 Inégalité de concentration

Propriété 3 — Inégalité de concentration. : Soit $(X_1, \ldots, X_n)$ un échantillon de n variables aléatoires indépendantes, et M_n la variable aléatoire moyenne de cette échantillon. Alors, pour tout réel δ strictement positif,

$$\mathbb{P}(|M_n - \mathbb{E}(X_1)| \geqslant \delta) \leqslant \frac{Var(X_1)}{n\delta^2}$$

Démonstration 2.1 : On applique simplement l'inégalité de Bienaymé-Tchebychev à la variable aléatoire M_n. Son espérance vaut $\mathbb{E}(X_1)$ et sa variance $\frac{Var(X_1)}{n}$. □

■ **Exemple 3 :** Soit X une variable aléatoire d'espérance 3 et de variance 100. On considère un On considère un échantillon $(X_1, \ldots, X_{100})$ de variables aléatoires indépendantes de même loi que X et on note $M_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.

Pour tout entier naturel non nul n et tout réel δ strictement positif, on a alors

$$\mathbb{P}(|M_n - \mathbb{E}(X_1)| \geqslant \delta) \leqslant \frac{Var(X_1)}{n\delta^2}$$

C'est-à-dire,

$$\mathbb{P}(|M_n - 3| \geqslant \delta) \leqslant \frac{100}{na^2}$$

En particulier, pour $n = 100000$ et $a = 0.1$

$$\mathbb{P}(|M_n - 3| \geqslant 0.1) \leqslant \frac{100}{100000 * 0.1^2}$$

c'est-à-dire

$$\mathbb{P}(|M_n - 3| \geqslant 0.1) \leqslant 0.1$$

Bien que la variable aléatoire X ait une grande variance, si l'on répète un grand nombre de fois l'expérience aléatoire, la moyenne des résultats est très proche de l'espérance de X : avec probabilité 0.9, la moyenne est entre 2.9 et 3.1. ■

2.3 Loi des grands nombres

Théorème 2.2 — Loi faible des grands nombres. : Soit $(X_1, ..., X_n)$ un échantillon de n variables aléatoires indépendantes et M_n la variable aléatoire moyenne de cet échantillon. Pour tout réel δ strictement positif,

$$\lim_{n \to +\infty} \mathbb{P}(|M_n - \mathbb{E}(X_1)| \geqslant \delta) = 0$$

Démonstration 2.3 : On applique l'inégalité de concentration à cet échantillon :

$$\mathbb{P}(|M_n - \mathbb{E}(X_1)| \geqslant \delta) \leqslant \frac{Var(X_1)}{n\delta^2}$$

Or, $\lim_{n \to +\infty} \frac{Var(X_1)}{n\delta^2} = 0$. De plus, $\mathbb{P}(|M_n - \mathbb{E}(X_1)| \geqslant \delta) \geqslant 0$. Par théorème d'encadrement, on a donc $\lim_{n \to +\infty} \mathbb{P}(|M_n - \mathbb{E}(X_1)| \geqslant \delta) = 0$ □

R Tout comme de nombreux résultats de probabilités de terminale, la loi faible des grands nombres est énoncée dans l'*Ars Conjectandi* de Jacques Bernoulli, paru en 1713. Bernoulli est si fier de ce théorème qu'il l'appelle son "théorème d'or". Le nom de loi des grands nombres ne viendra qu'au XIXe siècle.
Bernouli y aborde, comme nous l'avons déjà vu précédemment, la loi binomiale, les combinaisons, les permutations, mais également de la dérivation ou de calcul littéral en tout genre. Si vous avez déjà consulté le module sur les suites, vous avez peut-être rencontré une fameuse inégalité, l'inégalité de Bernoulli, qui dit que pour tout réel $x > 0$ et pour tout entier naturel n, $(1+x)^n \geqslant 1 + nx$. Cette inégalité est également tirée de cet ouvrage fascinant.
Et pour enfoncer le clou, sachez que le premier programme informatique, conçu par Ada Lovelace en 1843, avait justement pour but de calculer des nombres particuliers : les nombres de Bernoulli, encore mentionnés dans ce livre. C'est d'ailleurs également l'occasion de rappeler que le premier informaticien était en fait une informaticienne...