16

LOI DES GRANDS NOMBRES

Résumé

Nous étudions ici la répartition et le comportement d'un échantillon de variables aléatoires. Un résultat majeur est énoncé : la loi faible des grands nombres.

1 Échantillon d'une loi de probabilité

Définition 1

Soit X une variable aléatoire définie sur l'ensemble Ω des issues d'une expérience aléatoire.
Un **échantillon** de taille n de la loi de X est un n-uplet $(X_1; X_2; \ldots; X_n)$ de variables aléatoires indépendantes et identiquement distribuées suivant cette loi.

Remarque 2 Une suite $x_1, x_2, \ldots, x_n$ de valeurs prises par les X_i est une **réalisation** de cet échantillon.

Exemple 3 On considère X la variable aléatoire qui, à chaque paquet de cartes Pokémon , associe sa valeur une fois ouvert.
On note X_i la variable aléatoire qui, dans un lot de 3 paquets, associe la valeur du i^e paquet.

Les variables X_1, X_2 et X_3 sont indépendantes et suivent la même loi : celle de X. $(X_1; X_2; X_3)$ est un échantillon de la loi de X de taille 3.

Définitions 4

Soit $(X_1; X_2; \ldots; X_n)$ un échantillon de taille n de la loi de X.

- La **variable aléatoire somme** de l'échantillon est la variable aléatoire :

$$S_n = X_1 + X_2 + \cdots + X_n.$$

- La **variable aléatoire moyenne** de l'échantillon est la variable aléatoire :

$$\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Exemple 5 Dans l'exemple précédent, $S_3 = X_1 + X_2 + X_3$ est la valeur totale des trois paquets et $\overline{X_3} = \dfrac{X_1 + X_2 + X_3}{3}$ est la valeur moyenne du lot.

Propriétés 6 | Indicateurs de S_n

Soit $(X_1; X_2; \ldots; X_n)$ un échantillon de taille n de la loi de X.

- $\mathbb{E}[S_n] = n\mathbb{E}[X]$
- $\mathrm{Var}(S_n) = n\mathrm{Var}(X)$
- $\sigma(S_n) = \sqrt{n}\sigma(X)$

Démonstration.
- $\mathbb{E}[S_n] = \mathbb{E}[X_1 + \cdots + X_n] = \exp[X_1] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X]$ par linéarité de l'espérance.
- $\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\mathrm{Var}(X)$ par indépendance des X_i.
- $\sigma(S_n) = \sqrt{\mathrm{Var}(S_n)} = \sqrt{n}\sqrt{\mathrm{Var}(X)} = \sqrt{n}\sigma(X)$

□

Propriétés 7 | Indicateurs de $\overline{X_n}$

Soit $(X_1; X_2; \ldots; X_n)$ un échantillon de taille n de la loi de X.

- $\mathbb{E}\left[\overline{X_n}\right] = \mathbb{E}[X]$
- $\mathrm{Var}\left(\overline{X_n}\right) = \dfrac{1}{n}\mathrm{Var}(X)$
- $\sigma\left(\overline{X_n}\right) = \dfrac{1}{\sqrt{n}}\sigma(X)$

Démonstration. $\overline{X_n} = \dfrac{1}{n}S_n$ donc on utilise les propriétés de l'espérance et de la variance. □

Remarque 8 Plus n est grand, plus $\mathbb{E}\left[\overline{X_n}\right]$ est précis pour estimer $\mathbb{E}[X]$.
En effet, la **fluctuation d'échantillonnage** $\mathrm{Var}\left(\overline{X_n}\right)$ diminue : $\lim\limits_{n\to+\infty} \mathrm{Var}\left(\overline{X_n}\right) = 0$.

Théorème 9 | Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire.
Pour tout réel $\delta > 0$,
$$\mathbb{P}\Big(|X - \mathbb{E}[X]| \geqslant \delta\Big) \leqslant \frac{\mathrm{Var}(X)}{\delta^2}.$$

Démonstration. Soit $\delta > 0$ fixé. Notons $x_1, x_2, \ldots, x_n$ les images possibles de X rangées dans l'ordre strictement croissant. Il existe un ensemble A formé des x_i tels que $|x_i - \mathbb{E}[X]| \geqslant \delta$.

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{n} (x_i - \mathbb{E}[X])^2 \times \mathbb{P}(X = x_i) \\
&= \sum_{\substack{i=1 \\ x_i \in A}}^{n} (x_i - \mathbb{E}[X])^2 \times \mathbb{P}(X = x_i) + \sum_{\substack{i=1 \\ x_i \notin A}}^{n} (x_i - \mathbb{E}[X])^2 \times \mathbb{P}(X = x_i) \\
&\geqslant \sum_{\substack{i=1 \\ x_i \in A}}^{n} (x_i - \mathbb{E}[X])^2 \times \mathbb{P}(X = x_i) \\
&\geqslant \sum_{\substack{i=1 \\ x_i \in A}}^{n} \delta^2 \times \mathbb{P}(X = x_i) \\
&\geqslant \delta^2 \sum_{x_i \in A} \mathbb{P}(X = x_i) \\
&\geqslant \delta^2 \mathbb{P}\Big(|X - \mathbb{E}[X]| \geqslant \delta\Big)
\end{aligned}$$

□

Exemple 10 On lance 3 600 fois une pièce de monnaie équilibrée. Soit X la variable aléatoire qui associe le nombre de *Pile* obtenus.
On remarque que $X \sim \mathscr{B}(3\,600; 0{,}5)$.
Ainsi, $\mathbb{E}[X] = 3\,600 \times 0{,}5 = 1\,800$ et $\mathrm{Var}(X) = 3\,600 \times 0{,}5 \times 0{,}5 = 900$.

L'inégalité de Bienaymé-Tchebychev donne que, pour tout $\delta > 0$:

$$\mathbb{P}\Big(|X - \mathbb{E}[X]| \geqslant \delta\Big) \leqslant \frac{\mathrm{Var}(X)}{\delta^2} \Leftrightarrow \mathbb{P}\Big(|X - 1\,800| \geqslant \delta\Big) \leqslant \frac{900}{\delta^2}.$$

Si on souhaite estimer la probabilité que X soit strictement compris entre 1 600 et 2 000, on peut poser $\delta = 200$ et on a :

$$\begin{aligned}
\mathbb{P}\Big(1\,600 < X < 2\,000\Big) &= \mathbb{P}\Big(1\,800 - \delta < X < 1\,800 + \delta\Big) \\
&= \mathbb{P}\Big(|X - 1\,800| < \delta\Big) \\
&= 1 - \mathbb{P}\Big(|X - 1\,800| \geqslant \delta\Big) \\
&\geqslant 1 - \frac{900}{\delta^2} \\
&\geqslant 1 - \frac{900}{200^2} \\
&\geqslant 0{,}977\,5.
\end{aligned}$$

Corollaire 11 | Inégalité de concentration

Soit X une variable aléatoire et un échantillon $(X_1; X_2; \ldots; X_n)$ de X.
Pour tout réel $\delta > 0$,
$$\mathbb{P}\Big(\Big|\overline{X_n} - \mathbb{E}[X]\Big| \geqslant \delta\Big) \leqslant \frac{\mathrm{Var}(X)}{n\delta^2}$$

Démonstration. On applique Bienaymé-Tchebychev à $\overline{X_n}$. □

Exemple 12 On effectue des tirages successifs, avec remise, d'une boule dans une urne contenant deux boules rouges et trois boules noires. On note X la variable aléatoire qui, à un tirage donné, associe 1 si la boule est rouge et 0 sinon.
On considère un échantillon $(X_1; X_2; \ldots; X_n)$ de X correspondant à n tirages.

Par construction, $X \sim \mathscr{B}(0{,}4)$ donc $\mathbb{E}[X] = 0{,}4$ et $\mathrm{Var}(X) = 0{,}4 \times 0{,}8 = 0{,}32$.
L'inégalité de concentration donne que, pour tout $\delta > 0$:

$$\mathbb{P}\Big(\Big|\overline{X_n} - \mathbb{E}[X]\Big| \geqslant \delta\Big) \leqslant \frac{\mathrm{Var}(X)}{n\delta^2} \Leftrightarrow \mathbb{P}\Big(\Big|\overline{X_n} - 0{,}4\Big| \geqslant \delta\Big) \leqslant \frac{0{,}32}{n\delta^2}.$$

Nous souhaitons maintenant déterminer à partir de combien de tirages, l'indicateur $\overline{X_n}$ est compris, avec une probabilité supérieure ou égale à 0,95, entre 0,35 et 0,45.

$$\begin{aligned}
\mathbb{P}\Big(0{,}35 < \overline{X_n} < 0{,}45\Big) &= \mathbb{P}\Big(\Big|\overline{X_n} - 0{,}4\Big| < 0{,}05\Big) \\
&= 1 - \mathbb{P}\Big(\Big|\overline{X_n} - 0{,}4\Big| \geqslant 0{,}05\Big) \\
&\geqslant 1 - \frac{0{,}32}{n \times 0{,}05^2}.
\end{aligned}$$

Donc, pour que $\mathbb{P}\left(0{,}35 < \overline{X_n} < 0{,}45\right) \geqslant 0{,}95$, regardons n tel que :

$$\begin{aligned}
1 - \frac{0{,}32}{n \times 0{,}05^2} \geqslant 0{,}95 \\
\Leftrightarrow 0{,}05 \geqslant \frac{0{,}32}{n \times 0{,}05^2} \\
\Leftrightarrow n \geqslant \frac{0{,}32}{0{,}05^3} \\
\Leftrightarrow n \geqslant 2\,650.
\end{aligned}$$

Théorème 13 | Loi des grands nombres

Soit X une variable aléatoire et un échantillon $(X_1; X_2; \ldots; X_n)$ de X.
Pour tout $\epsilon > 0$ fixé,

$$\lim_{n \to +\infty} \mathbb{P}\left(\left|\overline{X_n} - \mathbb{E}[X]\right| \geqslant \epsilon\right) = 0.$$

Démonstration. On applique l'inégalité de concentration à $\overline{X_n}$:

$$0 \leqslant \mathbb{P}\left(\left|\overline{X_n} - \mathbb{E}[X]\right| \geqslant \delta\right) \leqslant \frac{\mathrm{Var}(X)}{n\epsilon^2}.$$

On conclut par le théorème des gendarmes.

□

Remarque 14 On dira que la suite de variables aléatoires $(\overline{X_n})$ converge en probabilité vers $\mathbb{E}[X]$.