— 18 —

Concentration et loi des grands nombres

I. Échantillon de variables aléatoires

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
$$\mathbb{E}(M_n) = E[X_1]$$
, $V(M_n) = \frac{1}{n}V(X_1)$ et $\sigma(M_n) = \frac{1}{\sqrt{n}}\sigma(X_1)$

Preuve. Pour tout n > 0, on a:

$$E[M_n] = E\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right] = \frac{1}{n}(E[X_1] + E[X_2] + \dots + E[X_n]) = \frac{1}{n} \times nE[X_1] = E[X_1]$$

Par ailleurs, les variables aléatoires étant indépendantes,

$$V(M_n) = V\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) = \frac{1}{n^2}(V(X_1) + V(X_2) + \dots + V(X_n))$$

et donc

$$V(M_n) = \frac{1}{n^2} \times nV(X_1) = \frac{V(X_1)}{n}$$

Année 2024/2025 Page 1/4

Exemple :

On considère une variable aléatoire X qui suit une loi binomiale de paramètre 3 et $\frac{1}{3}$.

On rappelle que
$$E[X] = 3 \times \frac{1}{3} = 1$$
 et $V(X) = 3 \times \frac{1}{3} \times \left(1 - \frac{1}{3}\right) = \frac{2}{3}$.

On considère un échantillon (X_1, \ldots, X_{100}) de variables aléatoires indépendantes de même loi que X et on note $M_{100} = \frac{1}{100}(X_1 + X_2 + \ldots + X_{100})$.

On a alors
$$E[M_{100}] = E[X] = 1$$
 et $V(M_n) = \frac{V(X)}{100} = \frac{2}{300}$.

II. Inégalité de Bienaymé-Tchebychev

Propriété 2 : Inégalité de Bienaymé-Tchebychev (Admise)

Soit X une variable aléatoire réelle. Pour tout réel $\delta > 0$, on a :

$$\mathbb{P}(|X - E[X]| \ge \delta) \le \frac{\mathbb{V}(X)}{\delta^2}$$

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

// Exemple :

Soit X une variable aléatoire d'espérance 10 et de variance 1.

D'après l'inégalité de Bienaymé-Tchebychev, appliquée à $\delta = 4$, on a $\mathbb{P}(|X-10| \ge 4) \le \frac{1}{4^2}$. Par ailleurs, les événements $|X-10| \ge 4$ et |X-10| < 4 étant contraires, on a donc

$$\mathbb{P}(|X - 10| < 4) = 1 - \mathbb{P}(|X - 10| \ge 4) \ge 1 - \frac{1}{4^2}.$$

Or, |X - 10| < 4 est équivalent à $X \in]10 - 4; 10 + 4[$, c'est-à-dire $X \in]6; 14[$. Finalement, $\mathbb{P}(6 < X < 14) \geqslant \frac{15}{16}$.

Exemple :

Soit une variable aléatoire X qui suit la loi binomiale de paramètres n=20 et p=0,1. Appliquons l'inégalité de Bienaymé-Tchebytchev avec $\delta=2\sigma(X)$:

$$E(X) = 20 \times 0, 1 = 2$$
 $V(X) = 20 \times 0, 1 \times 0, 9 = 1, 8$ $\sigma(X) = \sqrt{1, 8}$

$$P(|X - E(X)| \ge 2\sigma(X)) \le \frac{V(X)}{(2\sigma(X))^2}$$
 donc $P(|X - 2| \ge 2\sqrt{1.8}) \le 0.25$

La probabilité que l'écart de X à E(X) soit supérieur à $2\sigma(X)$ est majorée par 0, 25. Faisons de même avec $\delta = 3\sigma(X)$ et $\delta = 4\sigma(X)$:

$$P(|X - E(X)| \ge 3\sigma(X)) \le \frac{V(X)}{(3\sigma(X))^2}$$
 donc $P(|X - 2| \ge 3\sqrt{1.8}) \le \frac{1}{9}$

Année 2024/2025 Page 2/4

$$P(|X - E(X)| \ge 4\sigma(X)) \le \frac{V(X)}{(4\sigma(X))^2}$$
 donc $P(|X - 2| \ge 4\sqrt{1,8}) \le 0,0625$

On peut en déduire que les écarts de X à E(X) de quelques σ deviennent peu probables.

⚠ Remarque :

L'inégalité de Bienaymé-Tchebychev est très peu optimale mais elle a l'avantage d'être accessible et tout de même utile pour prouver des résultats importants.

III. Applications de l'inégalité de Bienaymé-Tchebychev

1. 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 $\delta > 0$, on a :

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

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

$/\!\!/$ Exemple :

Soit X une variable aléatoire d'espérance 3 et de variance 100.

On considère un échantillon (X_1, \ldots, X_n) 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 $\mathbb{P}(|M_n - E(X_1)| \ge \delta) \le \frac{V(X_1)}{n\delta^2}$,

c'est-à-dire
$$\mathbb{P}(|M_n - 3| \ge \delta) \le \frac{100}{n\delta^2}$$
.

En particulier, pour n = 100000 et $\delta = 0, 1$, on a $\mathbb{P}(|M_n - 3| \ge 0, 1) \le \frac{100}{100000 \times 0, 1^2}$ Ainsi, $\mathbb{P}(|M_n - 3| \ge 0, 1) \le 0, 1$.

En passant au complémentaire, on obtient alors que $\mathbb{P}(|M_n-3|<0,1)=1-\mathbb{P}(|M_n-3|\geqslant 0,1)\geqslant 0,9.$

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.

Année 2024/2025 Page 3/4

2. Loi des grands nombres

Théorème 1 : 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,

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

Preuve. On applique l'inégalité de concentration à cet échantillon :

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

Or,
$$\lim_{n\to+\infty}\frac{V(X_1)}{n\delta^2}=0$$
. De plus, $\mathbb{P}(|M_n-E(X_1)|\geqslant\delta)\geqslant0$. D'après le théorème d'encadrement, on a donc $\lim_{n\to+\infty}\mathbb{P}(|M_n-E(X_1)|\geqslant\delta)=0$.

Année 2024/2025 Page 4/4