

LOI DES GRANDS NOMBRES

I Inégalité de Bienaymé-Tchebychev

On sait que l'écart-type σ permet de déterminer la **dispersion** des valeurs d'une série par rapport à la **moyenne** de cette série.

De manière plus pragmatique, on peut imaginer que les écarts entre les valeurs d'une série et la moyenne de cette série sont de l'ordre de σ . Autrement dit, la « majorité » des valeurs d'une série seraient comprises dans l'intervalle $[\bar{x} - \sigma ; \bar{x} + \sigma]$.

C'est cette idée que précise l'inégalité de Bienaymé-Tchebychev, énoncée en 1853 par le mathématicien français Bienaymé et démontrée en 1867 par le mathématicien russe Tchebychev.

Inégalité de Bienaymé-Tchebychev

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

$$p(|X - \mu| \geq \delta) \leq \frac{V}{\delta^2}$$

Autrement dit, la probabilité que l'écart entre les valeurs de X et leur moyenne μ dépasse δ est inférieure à $\frac{V}{\delta^2}$. Ainsi, s'éloigner de la moyenne se produit rarement...

Par exemple, avec $\delta = 2\sigma$, on obtient $p(|X - \mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2}$

$$p(|X - \mu| \geq 2\sigma) \leq \frac{1}{4}$$

⇒ La probabilité que l'écart entre X et $E(X)$ dépasse 2σ est inférieure à 25%

⇒ Dans 75% des cas, l'écart entre une valeur de la série et sa moyenne est inférieur à 2σ

Ex : Dans une population, la taille d'un adulte pris au hasard est donnée par une variable aléatoire :

- F d'espérance 165 et de variance 25 pour une femme

- H d'espérance 180 et de variance 36 pour un homme

1. a. Majorer la probabilité $p(|F - 165| \geq 8)$ à 10^{-4} près.

b. Minorer la probabilité que la taille d'une femme soit strictement comprise entre 155 et 175 cm.

2. Appliquer l'inégalité de Bienaymé-Tchebychev à l'événement $|H - 180| \geq 10$ puis en donner l'interprétation.

1. a. Appliquons l'inégalité de Bienaymé-Tchebychev : $p(|F - 165| \geq 8) \leq \frac{25}{8^2}$

$$p(|F - 165| \geq 8) \leq 0,3907$$

b. On veut $155 < F < 175$ c'est-à-dire $155 - 165 < F - 165 < 175 - 165$

$$-10 < F - 165 < 10$$

$$|F - 165| < 10$$

$$\text{Or, } p(|F - 165| \geq 10) \leq \frac{25}{10^2}$$

$$-p(|F - 165| \geq 10) \geq -0,25$$

$$1 - p(|F - 165| \geq 10) \geq 1 - 0,25$$

$$p(|F - 165| < 10) \geq 0,75$$

2. D'après l'inégalité de Bienaymé-Tchebychev, $p(|H - 180| \geq 10) \leq \frac{36}{10^2}$

$$p(|H - 180| \geq 10) \leq 0,36$$

La probabilité qu'un homme mesure plus de 1m90 ou moins de 1m70 est inférieure à 36%.

Autrement dit, il y a plus de 64% de chances qu'un homme mesure entre 1m70 et 1m90.

Remarque : dans les faits, $p(|X - E(X)| \geq 2\sigma)$ est généralement très largement inférieure à 0,25...

Prenons un exemple : supposons que la probabilité de gagner à un jeu soit de 20%. Un joueur décide de jouer 50 parties. On note X le nombre de parties gagnées.

$$X \hookrightarrow \mathcal{B}(50; 0,20)$$

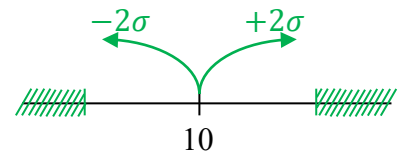
$$E(X) = 50 \times 0,20 = 10 \text{ et } \sigma(X) = \sqrt{50 \times 0,20 \times (1 - 0,20)} \approx 2,83$$

D'après Bienaymé-Tchebychev, $p(|X - 10| \geq 2\sigma) \leq 0,25$

$$\text{Or, } p(|X - 10| \geq 2\sigma) = p(X \geq 10 + 2\sigma) + p(X \leq 10 - 2\sigma)$$

$$= p(X \geq 16) + p(X \leq 4)$$

$$\approx 0,0493$$



Ainsi, l'inégalité de Bienaymé-Tchebychev donne une majoration *toujours vraie* mais qui est loin d'être optimale !

II Inégalité de concentration

1° Variable aléatoire moyenne

Soit X une variable aléatoire et $(X_1, X_2, \dots, X_n)$ un échantillon de taille n de variables aléatoires indépendantes suivant la même loi que X .

La variable aléatoire moyenne M_n de l'échantillon est donnée par :

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

On a alors : $E(M_n) = E(X)$

$$V(M_n) = \frac{1}{n}V(X) \quad \left(\text{et donc } \sigma(M_n) = \frac{1}{\sqrt{n}}\sigma(X) \right)$$

Preuve :

$$\begin{aligned}
 E(M_n) &= E\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) & V(M_n) &= V\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) \\
 &= \frac{1}{n}E(X_1 + X_2 + \dots + X_n) & &= \left(\frac{1}{n}\right)^2 V(X_1 + X_2 + \dots + X_n) \\
 &= \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) & &= \frac{1}{n^2}(V(X_1) + V(X_2) + \dots + V(X_n)) \\
 &= \frac{1}{n} \times nE(X) \quad \text{car } E(X_i) = E(X) & &= \frac{1}{n^2} \times nV(X) \quad \text{car } V(X_i) = V(X) \\
 &= E(X) & &= \frac{1}{n}V(X)
 \end{aligned}$$

2° Inégalité de concentration

Soit M_n la variable aléatoire moyenne d'un échantillon de taille n de variables aléatoires indépendantes suivant la même loi que X .

Pour tout réel $\delta > 0$, on a $p(|M_n - E(X)| \geq \delta) \leq \frac{V(X)}{n\delta^2}$

Preuve : c'est l'inégalité de Bienaymé-Tchebychev appliquée à la variable aléatoire M_n .

$$\begin{aligned}
 p(|M_n - E(M_n)| \geq \delta) &\leq \frac{V(M_n)}{\delta^2} \\
 p(|M_n - E(X)| \geq \delta) &\leq \frac{\frac{1}{n}V(X)}{\delta^2} \quad \text{c'est-à-dire } p(|M_n - E(X)| \geq \delta) \leq \frac{V(X)}{n\delta^2}
 \end{aligned}$$

Ex : On lance un dé équilibré à 6 faces n fois. Soit X_i la variable aléatoire donnant le résultat du $i^{\text{ème}}$ lancer.

Loi de X_i :

x_i	1	2	3	4	5	6
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X_i) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + \dots = 3,5$$

$$V(X_i) = \frac{1}{6}(1 - 3,5)^2 + \frac{1}{6}(2 - 3,5)^2 + \frac{1}{6}(3 - 3,5)^2 + \dots = \frac{35}{12}$$

Appliquons l'inégalité de concentration à $M_{100} = \frac{1}{100}(X_1 + X_2 + \dots + X_{100})$ avec $\delta = 0,5$:

$$p(|M_{100} - 3,5| \geq 0,5) \leq \frac{\frac{35}{12}}{100 \times 0,5^2}$$

$$p(|M_{100} - 3,5| \geq 0,5) \leq 0,1167$$

Ainsi, la probabilité que « l'écart entre la moyenne des 100 premiers lancers et 3,5 soit supérieur à 0,5 » est inférieure à 11,66%.

3° Détermination de la taille d'un échantillon

Reprenons notre exemple :

On lance un dé équilibré à 6 faces n fois. Soit X_i la variable aléatoire donnant le résultat du $i^{\text{ème}}$ lancer. $(X_1, X_2, \dots, X_n)$ est un échantillon de taille n de variables aléatoires indépendantes d'espérance 3,5 et de variance $\frac{35}{12}$.

On note $M_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. On a $E(M_n) = 3,5$ et $V(M_n) = \frac{1}{n} \times \frac{35}{12}$.

Déterminer le plus petit entier n tel que la probabilité que M_n appartienne à $]3 ; 4[$ soit supérieure à 95%.

On cherche donc à déterminer n tel que $p(3 < M_n < 4) \geq 0,95$

Pour utiliser l'inégalité de concentration, on introduit l'espérance :

$$p(3 - 3,5 < M_n - 3,5 < 4 - 3,5) \geq 0,95$$

$$p(-0,5 < M_n - 3,5 < 0,5) \geq 0,95$$

$$p(|M_n - 3,5| < 0,5) \geq 0,95$$

$$1 - p(|M_n - 3,5| \geq 0,5) \leq 1 - 0,95$$

$$p(|M_n - 3,5| \geq 0,5) \leq 0,05$$

Or, d'après l'inégalité de concentration, $p(|M_n - 3,5| \geq 0,5) \leq \frac{\frac{35}{12}}{n \times 0,5^2}$

Il faut donc que $\frac{\frac{35}{12}}{n \times 0,5^2} \leq 0,05$

$$\frac{\frac{35}{12}}{0,05 \times 0,5^2} \leq n$$

$$233,33... \leq n$$

C'est donc à partir du 234^{ème} lancer que $p(|M_n - 3,5| \geq 0,5) \leq 0,05$.

Autrement dit, à partir du 234^{ème} lancer, il y a plus de 95% de chances que la valeur moyenne des

lancers de dés soit comprise entre 3 et 4.

4° Loi faible des grands nombres

Soit M_n la variable aléatoire d'un échantillon de taille n de la variable aléatoire X .

Pour tout réel $\delta > 0$, on a :

$$\lim_{n \rightarrow +\infty} P(|M_n - E(X)| \geq \delta) = 0$$

Cette loi traduit le fait que, plus la taille de l'échantillon d'une variable aléatoire X est grande, plus l'écart entre la moyenne de cet échantillon et l'espérance de la variable aléatoire X est faible.