La loi des grands nombres

Cours

Sommaire

I L'inégalité de Bienaymé-Tchebychev

L'écart-type σ d'une variable aléatoire X est une mesure de dispersion correspondant à un écart moyen entre les valeurs prises par X et son espérance μ. On peut se questionner sur la probabilité que la variable X prenne des valeurs plus ou moins éloignées de son espérance et mesurer cet éloignement avec l'écart-type.

THÉORÈME

Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire réelle d'espérance μ et de variance V.

Pour tout réel $\delta > 0$, on a :

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

EXEMPLE

On joue 100 fois à « pile ou face » avec une pièce équilibrée.

Soit X la variable aléatoire comptant le nombre de « pile » obtenus durant ces 100 lancers.

On sait que X suit la loi binomiale $\mathcal{B}(100;0{,}5)$.
Son espérance est $\mu = 100 \times 0{,}5 = 50$.
Sa variance est $V = 100 \times 0{,}5 \times 0{,}5 = 25$.

Ainsi, pour tout réel $\delta > 0$, on a :

$$P(|X-50| \geq \delta) \leq \frac{25}{\delta^2}$$

En particulier :

$P(|X-50|\geq 10)\leq \dfrac{25}{100}$, soit $P(|X-50|\geq 10)\leq 0{,}25$.

REMARQUE

Soit X une variable aléatoire réelle d'espérance μ , de variance V et d'écart-type σ .

Soit n un entier naturel non nul.

En choisissant $\delta = n\sigma$ dans l'inégalité de Bienaymé-Tchebychev, on obtient :

$$P(|X-\mu|\geq n\sigma)\leq \frac{1}{n^2}$$

Ainsi on obtient une information sur la probabilité que l'écart entre X et son espérance dépasse un certain de nombre de fois l'écart-type.

EXEMPLE

On joue 100 fois à « pile ou face » avec une pièce équilibrée.

Soit X la variable aléatoire comptant le nombre de « pile » obtenus durant ces 100 lancers.

On sait que X suit la loi binomiale $\mathcal{B}(100;0{,}5)$.
Son espérance est $\mu = 100\times 0{,}5 = 50$.
Sa variance est $V = 100\times 0{,}5\times 0{,}5 = 25$.
Son écart-type est $\sigma = \sqrt{25} = 5$.

Ainsi, pour tout entier naturel non nul n , on a :

$$P(|X-50|\geq 5n)\leq \frac{1}{n^2}$$

REMARQUE

Soit X une variable aléatoire réelle d'espérance μ , de variance V et d'écart-type σ .

- L'inégalité de Bienaymé-Tchebychev a un caractère universel.
- En contrepartie, elle est loin d'être optimale. En effet, elle montre qu'un écart à μ supérieur à 2σ est de probabilité inférieure ou égale à $\dfrac{1}{4}$, alors que par simulation on obtient que cette probabilité est souvent majorée par 0,05.

II L'inégalité de concentration

On peut ensuite majorer l'écart entre une variable aléatoire et son espérance dans le cas où l'on répète plusieurs fois la même épreuve et où l'on calcule la moyenne des résultats obtenus.

PROPRIÉTÉ

Soit n un entier naturel non nul.

Soient une loi de probabilité et $(X_1; X_2; ...; X_n)$ un échantillon de cette loi constitué de variables aléatoires d'espérance μ et de variance V .

Soit $M_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$ la variable aléatoire moyenne.

Alors pour tout réel $\delta > 0$, on a :

$$P\left(|M_n - \mu| \geq \delta\right) \leq \frac{V}{n\delta^2}$$

EXEMPLE

Jeanne joue tous les jours d'une semaine donnée 100 fois à « pile ou face » avec une pièce équilibrée.

Soit X_i la variable aléatoire comptant le nombre de « pile » obtenus durant ces 100 lancers le i ème jour.

On sait que quel que soit l'entier i compris entre 1 et 7 :

- X_i suit la loi binomiale $\mathcal{B}(100; 0{,}5)$;
- son espérance est $E(X_i) = \mu = 100 \times 0{,}5 = 50$;
- sa variance est $V(X_i) = V = 100 \times 0{,}5 \times 0{,}5 = 25$.

Soit $M = \dfrac{X_1 + X_2 + ... + X_7}{7}$ la variable aléatoire moyenne.

Ainsi, pour tout réel $\delta > 0$, on a :

$$P\left(|M - 50| \geq \delta\right) \leq \frac{25}{7\delta^2}$$

En particulier :

$P\left(|M - 50| \geq 10\right) \leq \dfrac{25}{700}$, soit $P\left(|M - 50| \geq 10\right) \leq \dfrac{1}{28}$.

REMARQUE

La propriété précédente porte le nom d'inégalité de concentration.

REMARQUE

Soit n un entier naturel non nul.

Soit une loi de probabilité et $(X_1; X_2; ...; X_n)$ un échantillon de cette loi constitué de variables aléatoires d'espérance μ et de variance V .

Soit $M_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$ la variable aléatoire moyenne.

En choisissant $\delta = m\sigma$ avec m entier naturel non nul dans l'inégalité de concentration, on obtient :

$$P\left(|M_n - \mu| \geq m\sigma\right) \leq \frac{1}{nm^2}$$

Ainsi, plus la taille de l'échantillon est grand, plus la probabilité que l'écart entre M_n et μ dépasse un certain nombre de fois σ est faible.

EXEMPLE

Jeanne joue tous les jours d'une semaine donnée 100 fois à « pile ou face » avec une pièce équilibrée.

Soit X_i la variable aléatoire comptant le nombre de « pile » obtenus durant ces 100 lancers le i ème jour.

On sait que quel que soit l'entier i compris entre 1 et 7 :

- X_i suit la loi binomiale $\mathcal{B}(100; 0{,}5)$;
- son espérance est $E(X_i) = \mu = 100 \times 0{,}5 = 50$;
- sa variance est $V(X_i) = V = 100 \times 0{,}5 \times 0{,}5 = 25$;
- son écart-type est $\sigma(X_i) = \sqrt{V} = \sqrt{100 \times 0{,}5 \times 0{,}5} = \sqrt{25} = 5$.

Soit $M = \dfrac{X_1 + X_2 + \ldots + X_7}{7}$ la variable aléatoire moyenne.

Ainsi, pour tout entier naturel non nul, m , on a :

$$P\left(|M - 50| \geq 5m\right) \leq \frac{1}{7m^2}$$

REMARQUE

L'inégalité précédente permet de rechercher la taille n d'un échantillon pour majorer la probabilité que l'écart à l'espérance dépasse une valeur donnée.

EXEMPLE

Jeanne joue n jours de suite au même jeu :
Lancer 100 fois une pièce équilibrée et noter le nombre de « pile » obtenus.

Soit X_i la variable aléatoire comptant le nombre de « pile » obtenus durant ces 100 lancers le i ème jour.

On sait que quel que soit l'entier i compris entre 1 et n :

- X_i suit la loi binomiale $\mathcal{B}(100; 0{,}5)$;
- son espérance est $E(X_i) = \mu = 100 \times 0{,}5 = 50$;
- sa variance est $V(X_i) = V = 100 \times 0{,}5 \times 0{,}5 = 25$.

Soit $M_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$ la variable aléatoire moyenne.

Ainsi, pour tout réel $\delta > 0$, on a :

$$P\left(|M_n - 50| \geq \delta\right) \leq \frac{25}{n\delta^2}$$

En particulier :

$P\left(|M_n - 50| \geq 10\right) \leq \dfrac{25}{100n}$, soit $P\left(|M_n - 50| \geq 10\right) \leq \dfrac{1}{4n}$.

Si l'on souhaite majorer la probabilité $P\left(|M_n - 50| \geq 10\right)$ par $\dfrac{1}{100}$, on peut choisir n de telle sorte que $\dfrac{1}{4n} \leq \dfrac{1}{100}$.

Or :

$$\frac{1}{4n} \leq \frac{1}{10} \Leftrightarrow \frac{4n}{1} \geq \frac{100}{1}$$
$$\frac{1}{4n} \leq \frac{1}{10} \Leftrightarrow 4n \geq 100$$
$$\frac{1}{4n} \leq \frac{1}{10} \Leftrightarrow n \geq 25$$

Ainsi, si Jeanne joue au moins 25 jours de suite, la probabilité que l'écart entre la moyenne et l'espérance dépasse 10 sera inférieure à $\dfrac{1}{100}$.

III La loi des grands nombres

On étudie ensuite l'écart entre M_n et μ lorsque la taille de l'échantillon devient très grande.

THÉORÊME

Loi (faible) des grands nombres

Soit n un entier naturel non nul.

Soit une loi de probabilité et $\left(X_1; X_2; ...; X_n\right)$ un échantillon de cette loi constitué de variables aléatoires d'espérance μ et de variance V.

Soit $M_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$ la variable aléatoire moyenne.

Alors pour tout réel $\delta > 0$, on a :

$$\lim_{n \to +\infty} P\left(|M_n - \mu| \geq \delta\right) = 0$$

EXEMPLE

Soit X une variable aléatoire dont la loi de probabilité est la suivante :

x_i	0	10
$P(X=x_i)$	$\frac{1}{3}$	$\frac{2}{3}$

On a donc $E(X) = 0 \times \frac{1}{3} + 10 \times \frac{2}{3} = \frac{20}{3}$.

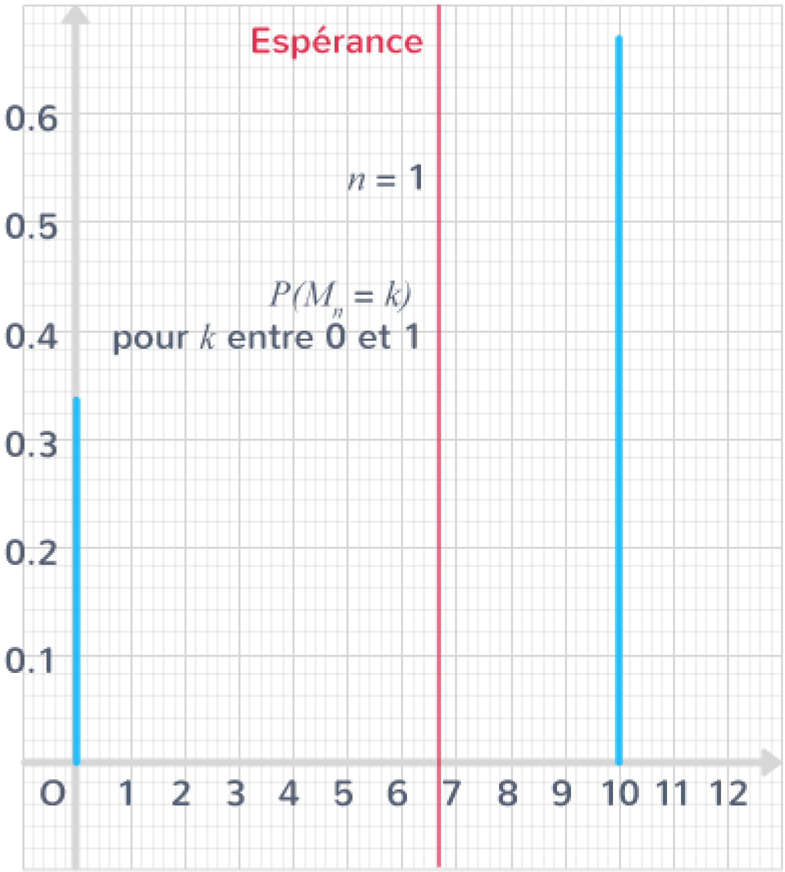
Soit n un entier naturel non nul, $(X_1; X_2; ...; X_n)$ un échantillon de cette loi de probabilité et M_n la variable aléatoire moyenne de cet échantillon.

On a donc, pour tout réel $\delta > 0$:

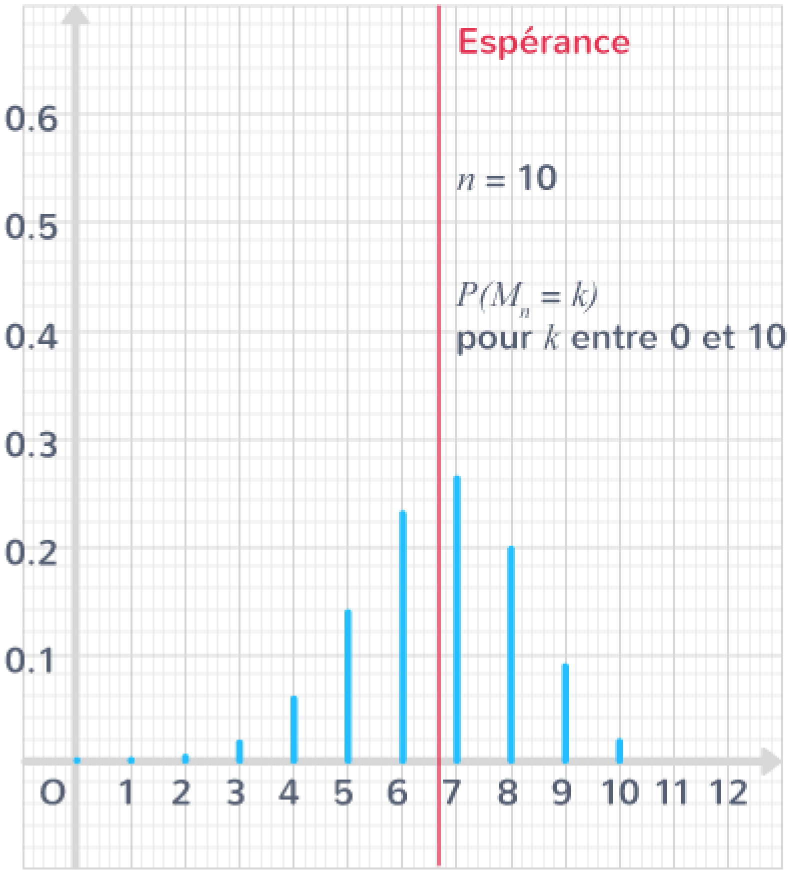
$$\lim_{n \to +\infty} P\left(\left|M_n - \frac{20}{3}\right| \geq \delta\right) = 0$$

Autrement dit, plus n est grand, plus la probabilité que la variable aléatoire M_n prenne des valeurs « éloignées » de l'espérance est faible.

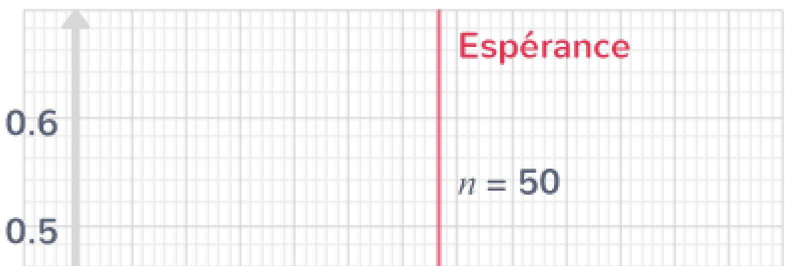
Voici les graphiques donnant les valeurs $P(M_n = k)$ pour toutes les valeurs possibles de k dans les cas $n = 1$, $n = 10$ et $n = 50$:

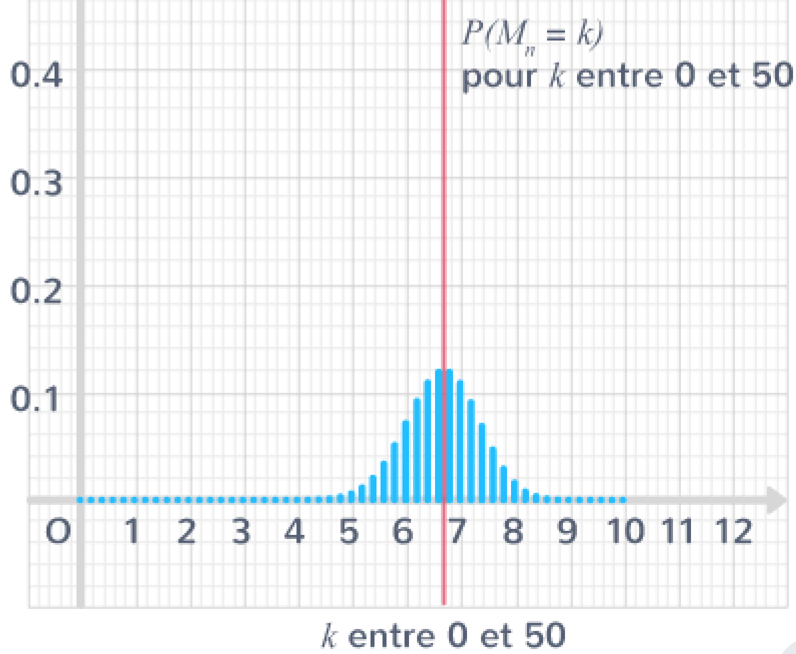


k entre 0 et 1



k entre 0 et 10




$P(M_n = k)$
pour k entre 0 et 50
0.4
0.3
0.2
0.1
O
1 2 3 4 5 6 7 8 9 10 11 12
k entre 0 et 50