Concentration et loi des grands nombres

I Inégalité de Markov et de Bienaymé-Tchebychev

Théorème - Inégalité de Markov

Soit X une variable aléatoire réelle positive ou nulle d'espérance μ.

$$\text{pout tout } \delta > 0, P\big(X \geqslant \delta\big) \leqslant \frac{\mu}{\delta}$$

Exemple

En 2015, le salaire brut mensuel moyen en france était de 2442 euros. On choisit un salarié au hasard et on note X la variable aléatoire qui associe à chaque issue le salaire du salarié.
Les salaires étant positifs, X est une variable aléatoire positive, donc d'après l'inégalité de Markov :

$$P(X \geqslant 7326) \leqslant \frac{2442}{7326} \quad \text{d'où} \quad P(X \geqslant 7326) \leqslant \frac{1}{3}$$

Démonstration

Soient n réels **positifs** $x_1, x_2, ..., x_n$ et X une variable aléatoire réelle **positive** ou nulle d'espérance μ et telle que $X(\Omega) = \{x_1, x_2, ..., x_n\}$.
On a donc $\mu = \sum_{i=1}^{n} x_i P\left(X = x_i\right)$, d'où pour tout $\delta > 0$:

$$\mu = \sum_{x_i \geqslant \delta} x_i P\left(X = x_i\right) + \sum_{x_i < \delta} x_i P\left(X = x_i\right)$$

donc, X étant une variable aléatoire à valeur positive, on a :

$$\begin{aligned}
\mu &\geqslant \sum_{x_i \geqslant \delta} x_i P\left(X = x_i\right) \\
&\geqslant \sum_{x_i \geqslant \delta} \delta P\left(X = x_i\right) \qquad \text{car } x_i \geqslant \delta \\
&\geqslant \delta \times \sum_{x_i \geqslant \delta} P\left(X = x_i\right) \quad \text{en factorisant par } \delta \\
&\geqslant \delta \times P\left(X \geqslant \delta\right)
\end{aligned}$$

donc

$$P\left(X \geqslant \delta\right) \leqslant \frac{\mu}{\delta}$$

Théorème - Inégalité de Bienaymé-Tchebychev
Soit X une variable aléatoire d'espérance μ et de variance V.

$$\text{pour tout } \delta > 0, P\big(|X-\mu| \geqslant \delta\big) \leqslant \frac{V}{\delta^2}$$

Interprétation La probabilité que les valeurs prisent par X s'écarte de plus de δ de l'espérance $\mu = E(X)$ est d'autant plus petite que δ est grand.

Exemple

Le taux moyen de glycémie dans une population est de 1 gramme par litre, avec une variance de 0.1.
On choisit aux hasard une personne de cette population.
Soit X la variable aléatoire qui associe à chaque issue son taux de glycémie.
Une personne présente un taux de glycémie critique si ce dernier ne se situe pas dans l'intervalle $]0,5;1,5[$. Cet événement ce traduit par l'inégalité $|X - E(X)| \geqslant 0,5$.
D'après l'inégalité de Bienaymé-Tchebychev, on a :

$$P(|X-E(X)| \geqslant 0,5) \leqslant \frac{0,1}{0,5^2} \quad \text{d'où} \quad P(|X-E(X)| \geqslant 0,5) \leqslant 0.4$$

La probabilité qu'une personne présente un taux de glycémie critique est donc inférieur ou égale à 0,4.

Démonstration

Soient X une variable aléatoire d'espérance μ et de variance V et δ un réel strictement positif.
$\delta > 0$, donc $|X-\mu| \geqslant \delta \Leftrightarrow (X-\mu)^2 \geqslant \delta^2$
La variable aléatoire $(X-\mu)^2$ étant positive ou nulle, on a, d'après l'inégalité de Markov,

$$P\big((X-\mu)^2 \geqslant \delta^2\big) \leqslant \frac{E\big((X-\mu)^2\big)}{\delta^2}$$

or $E\big((X-\mu)^2\big) = V$, donc $P\big((X-\mu)^2 \geqslant \delta^2\big) \leqslant \frac{V}{\delta^2}$ et par conséquent :

$$P\big(|X-\mu| \geqslant \delta\big) \leqslant \frac{V}{\delta^2}$$

II Loi des grands nombres

Théorème - Inégalité de concentration
Soit n variables aléatoires (avec $n \geqslant 1$) $X_1, X_2, ..., X_n$ un échantillon de taille n d'une variable aléatoire X. C'est à dire indépendantes et identiquement distribuées, suivant une même loi de probabilité d'espérance μ et de variance V.
On note $M_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$ la variable aléatoire moyenne.

$$\text{pout tout } \delta > 0, P\big(|M_n - \mu| \geqslant \delta\big) \leqslant \frac{V}{n\delta^2}$$

En effet, $E(M_n) = \mu$ et $V(M_n) = \dfrac{V}{n}$

Exemple

Dans une société de démarchage par téléphone, on estime que 40% des personnes appelées répondent effectivement.
On appelle n personnes. Soit X_k la variable aléatoire qui associe 1 à toute issue dont la k^{ime} personne appelée répond et 0 sinon.
les variables aléatoires $X_1, X_2, ..., X_n$ sont indépendantes et identiquement distribuées d'espérance $\mu = 0,6 \times 0 + 0,4 \times 1 = 0,4$ et de variance $V = 0,6 \times (0 - 0,4)^2 + 0,4 \times (1 - 0,4)^2 = 0,24$
On a alors, pour tout réel $\delta > 0$, $P\big(|M_n - 0,4| \geqslant \delta\big) \leqslant \dfrac{0,24}{n\delta^2}$
Par exemple pour $\delta = 0,1$ et $n = 1000$
$P\big(|M_n - 0,4| \geqslant \delta\big) \leqslant \dfrac{24}{1000 \times 0,1^2}$
Donc $P\big(|M_n - 0,4| \geqslant \delta\big) \leqslant 0,024$. on dit que l'on obtient pour M_n une précision de $0,1$ avec un risque de 0,024.
Lorsque l'on appelle 1000 personnes, la probabilité que le nombre de personne qui répondent soit en dehors de l'intervalle $]300; 500[$ est inférieur à $0,024 = 2,4\%$

Démonstration

Soit n variables aléatoires (avec $n \geqslant 2$) $X_1, X_2, ..., X_n$ indépendantes et identiquement distribuées, suivant une même loi de probabilité d'espérance μ et de variance V, ($X_1, X_2, ..., X_n$ est un échantillon de taille n).
On a donc, $E(M_n) = \mu$ et $V(M_n) = \dfrac{V}{n}$ et par conséquent, l'inégalité de Bienaymé-Tchebychev donne pour tout $\delta > 0$:

$$P\big(|M_n - \mu| \geqslant \delta\big) \leqslant \frac{V}{n\delta^2}$$

Propriété - Loi faible des grands nombres

Soient $n \in \mathbb{N}$ et $(X_i)_{1 \leqslant i \leqslant n}$ un échantillon de taille n d'une variable aléatoire X d'espérance μ et de variance V.
Soit $M_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i$ la variable aléatoire moyenne.
Pour tout réel $\delta > 0$

$$\lim_{n \to +\infty} P\big(|M_n - \mu| \geqslant \delta\big) = 0$$

Démonstration

Soient $n \in \mathbb{N}$ et $(X_i)_{1 \leqslant i \leqslant n}$ un échantillon de taille n d'une variable aléatoire X d'espérance μ et de variance V.
Soit $M_n = \displaystyle\sum_{i=1}^{n} X_i n$ la variable aléatoire moyenne.
Pour tout réel $\delta > 0$ D'après l'inégalité de Bienaymé-Tchebychef,

$$0 \leqslant P\big(|M_n - \mu| \geqslant \delta\big) \leqslant \frac{V}{n\delta^2}$$

d'où, par encadrement et par passage à la limite quand n tend vers $+\infty$,

$$\lim_{n \to +\infty} P(|M_n - \mu| \geqslant \delta\big) = 0$$