Chapitre 14 – LOI DES GRANDS NOMBRES

La forme faible de la loi des grands nombres repose sur une convergence en probabilité, c'est-à-dire qu'elle affirme que parmi tous les échantillons de valeurs possibles, ceux dont la moyenne s'éloigne de la probabilité sont rares, et que cette rareté s'accentue avec la taille de l'échantillon. Ce phénomène était déjà remarqué dans l'étude statistique sur des jeux de hasard, en astronomie et en finance notamment.

Jacques Bernoulli, dans la partie quatre de son *Ars Conjectandi* (« Art de la conjecture ») publié en 1713, l'étudie dans le cas de la loi binomiale. Le résultat obtenu par ce dernier est appelé « loi des grands nombres » par Siméon Denis Poisson, et sera généralisé à toutes les lois de probabilités admettant une variance notamment grâce à l'inégalité de Markov au début du XX^e^ siècle.

Au milieu du XIX^e^ siècle, Irénée-Jules Bienaymé énonce l'inégalité en 1853 et Pafnouti Tchebychev la démontre ce qui permettra de prouver de façon rigoureuse cette loi des grands nombres.

1) INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV

a) L'inégalité de Markov

Définition

Une variable aléatoire est dite **positive ou nulle** dans un univers Ω, lorsque toutes les valeurs prises par celle-ci sont des réels positifs ou nuls.

Exemple
La variable aléatoire donnant le nombre de faces numérotées 1 obtenues sur dix lancers d'un dé est positive ou nulle.

Théorème (Inégalité de Markov)

Soit X une variable aléatoire réelle positive ou nulle d'espérance $E(\mathrm{X})$.
Alors, pour tout réel a strictement positif, $P(\mathrm{X} \geq \mathrm{a}) \leq \frac{E(\mathrm{X})}{a}$.

Démonstration
Soit X une variable aléatoire réelle positive ou nulle dont on note x_i les n valeurs pour l'entier i entre 1 et n.
Par définition de l'espérance, on a $E(\mathrm{X}) = \sum_{i=1}^{n} x_i P(X = x_i)$.
Séparons cette somme en deux blocs en considérant les valeurs supérieures ou égales à a et celles strictement inférieures à a.
On obtient $E(X) = \sum_{x_i \geq a} x_i P(X = x_i) + \sum_{x_i < a} x_i P(X = x_i)$.
Pour tout entier i compris entre 1 et n, on sait que $x_i \geq 0$ (car X est positive ou nulle) et $P(X = x_i) \geq 0$ (par définition d'une probabilité) donc $\sum_{x_i < a} x_i P(X = x_i) \geq 0$.
On en déduit que $E(X) \geq \sum_{x_i \geq a} x_i P(X = x_i)$.
Par ailleurs, dans cette partie de la somme, pour tout entier i compris entre 1 et n, $x_i \geq a$ donc $E(X) \geq \sum_{x_i \geq a} x_i P(X = x_i) \geq \sum_{x_i \geq a} a P(X = x_i)$.
Or $\sum_{x_i \geq a} a P(X = x_i) = a \sum_{x_i \geq a} P(X = x_i) = a P(X \geq a)$.
Par conséquent, $E(X) \geq\ a \sum_{x_i \geq a} P(X = x_i)$ ou encore $E(X) \geq\ a P(X \geq a)$.
Puisque $a > 0$, on obtient enfin $P(\mathrm{X} \geq \mathrm{a}) \leq \frac{E(\mathrm{X})}{a}$.

Exemple
En 2015, le salaire brut mensuel moyen en France était de 2442€. On choisit un salarié au hasard et on note X la variable aléatoire donnant son salaire. Les salaires étant positifs ou nuls, on sait que X est une variable aléatoire positive ou nulle.
On peut donc appliquer l'inégalité de Markov sur un exemple :

$P(X \geq 7326) \leq \frac{2442}{7326}$ soit $P(X \geq 7326) \leq \frac{1}{3}$.

Exemple

Une usine produit en moyenne 35 pièces par semaine.

On note X la variable aléatoire donnant le nombre de pièces produites par semaine.

Que peut-on dire de la probabilité que l'usine produise plus de 70 pièces par semaine.

Solution

D'après le contexte, on sait que la variable X est positive et que $E(X) = 35$.

L'inégalité de Markov implique que $P(X \geq 70) \leq \frac{35}{70} = \frac{1}{2}$.

Il y a donc une chance sur deux que l'usine produise plus de 70 pièces par semaine.

Méthode

- On vérifie que la variable aléatoire considérée soit positive ou nulle.
- On repère la valeur de l'espérance.
- On applique l'inégalité de Markov.

b) L'inégalité de Bienaymé-Tchebychev

Théorème (Inégalité de Inégalité de Bienaymé-Tchebychev)

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$.

Alors, pour tout réel a strictement positif, $P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$.

Démonstration

Comme $a > 0$, les inégalités $|X - E(X)| \geq a$ et $[X - E(X)]^2 \geq a^2$ sont équivalentes.

De plus, la variable $[X - E(X)]^2$ est positive ou nulle.

On applique donc l'inégalité de Markov à la variable $[X - E(X)]^2$ et au réel a^2.

Ainsi $P([X - E(X)]^2 \geq a^2) \leq \frac{E\left([X-E(X)]^2\right)}{a^2}$. Or $[X - E(X)]^2 = V(X)$ donc $P([X - E(X)]^2 \geq a^2) \leq \frac{V(X)}{a^2}$ et on a donc bien $P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$.

Exemple

Le taux moyen de glycémie dans la population est de $1\ g.L^{-1}$ avec une variance de $0{,}1$. Une personne présente un taux de X critique si son taux ne se situe pas dans l'intervalle $]0{,}5; 1{,}5[$. Cet évènement se traduit par l'inégalité $|X - E(X)| \geq 0{,}5$.

Sa probabilité vérifie donc $P(|X - E(X)| \geq 0{,}5) \leq \frac{0{,}1}{0{,}5^2} = 0{,}4$.

La probabilité qu'une personne présente un taux critique est inférieur ou égal à $0{,}4$.

Propriété

Sous les mêmes conditions que le théorème précédent, l'inégalité peut être écrite de la façon suivante :

$$P(|X - E(X)| < a) \leq 1 - \frac{V(X)}{a^2}$$

Démonstration

$P(|X - E(X)| \geq a) = 1 - P(|X - E(X)| < a)$ et en utilisant l'inégalité de Bienaymé-Tchebychev, on obtient aisément l'inégalité de la propriété.

Exemple

Lors d'une saison de football, le nombre moyen de buts par match est de 2,5 avec une variance de $1{,}1$.

Majorer la probabilité que le match suivant ne se termine pas avec deux ou trois buts.

Solution

Soit X la variable aléatoire donnant le nombre de buts.
On cherche ici la probabilité que $X \leq 1$ ou $X \geq 4$ et on sait que $E(X) = 2{,}5$ et $V(X) = 1{,}1$.
Cela revient donc à dire que $|X - E(X)| \geq 1{,}5$.
On applique l'inégalité de Bienaymé-Tchebychev pour obtenir $P(|X - E(X)| \geq 1{,}5) \leq \frac{1{,}1}{1{,}5^2} < 0{,}49$.
La probabilité que le match ne se termine pas par deux ou trois buts est inférieure à 0,49.

Méthode
- Définir une variable aléatoire correspondant à l'énoncé.
- Traduire les inégalités de l'énoncé sous d'écart par rapport à la moyenne.
- Appliquer l'inégalité de Bienaymé-Tchebychev pour conclure.

2) LOI DES GRANDS NOMBRES

a) L'inégalité de concentration

Théorème (Inégalité de concentration)

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$.
On pose M_n la variable aléatoire moyenne d'un échantillon de taille n de X ;autrement dit, $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, où les variables X_i sont indépendantes et de même loi de probabilité (celle de X).
Alors, pour tout réel a strictement positif, $P(|M_n - E(X)| \geq a) \leq \frac{V(X)}{na^2}$.

Démonstration
On applique l'inégalité de Bienaymé-Tchebychev à M_n :
$P(|M_n - E(M_n)| \geq a) \leq \frac{V(M_n)}{a^2}$
Or $E(M_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = E(X)$ et $V(M_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}V(X)$,
donc $P(|M_n - E(X)| \geq a) \leq \frac{\frac{1}{n}V(X)}{a^2}$ soit $P(|M_n - E(X)| \geq a) \leq \frac{V(X)}{na^2}$.

Exemple 1
On effectue n lancers successifs supposés indépendants d'une pièce équilibrée.
On associe à chaque tirage i la variable aléatoire X_i, prenant comme valeur 0 si on obtient face et 1 si on obtient pile. On pose $S_n = \sum_{i=1}^{n} X_i$ la variable aléatoire donnant le nombre de piles obtenu.
On pose $M_n = \frac{S_n}{n}$. On a, pour tout $i \in \{1; \dots; n\}$, $E(X_i) = \frac{1}{2}$ et $V(X_i) = \frac{1}{4}$.
Pour $n = 10000$, l'inégalité de concentration donne :

$$P\left(\left|M_n - \frac{1}{2}\right| \geq 0{,}01\right) \leq \frac{\frac{1}{4}}{10000 \times 0{,}01^2} \Leftrightarrow P\left(\left|M_n - \frac{1}{2}\right| \geq 0{,}01\right) \leq \frac{1}{4}$$

Pour 10000 lancers, la probabilité que la proportion de pile obtenue s'écarte de plus d'un centième de $\frac{1}{2}$ est inférieure ou égale à $\frac{1}{4}$.

Exemple 2
On reprend l'exemple précédent. On souhaite que l'écart entre la proportion de pile obtenue et $\frac{1}{2}$ soit inférieur ou égale à 0,01. Quelle est la valeur minimale de n pour que le risque d'erreur soit inférieur ou égal à 5% ?

Solution

En reprenant le travail effectué dans l'exemple avec n lancers à la place de 10000, on a :

$$\frac{\frac{1}{4}}{n \times 0{,}01^2} \leq 0{,}05 \Leftrightarrow n \geq \frac{\frac{1}{4}}{0{,}05 \times 0{,}01^2} \Leftrightarrow n \geq 50000$$

Pour que le risque d'erreur soit inférieur ou égal à 5% il faut effectuer au moins 50000 lancers.

Méthode
Dans un exercice complet :
- On pose convenablement les données pour écrire l'inégalité de concentration dans le cas général.
- On l'applique ensuite pour les valeurs souhaitées en résolvant l'inégalité.

b) Loi faible des grands nombres

Propriété

Soit (X_n) un échantillon d'une variable aléatoire.
On pose $M_n = \frac{1}{n}\sum_{i=0}^{n} X_k$.
Alors, pour tout réel a strictement positif, $\lim\limits_{n \to +\infty} P(|M_n - E(X)| \geq a) = 0$.

Démonstration
On applique l'inégalité de concentration à la variable aléatoire M_n :
$0 \leq P(|M_n - E(X)| \geq a) \leq \frac{V(X)}{na^2}$
Or $\lim\limits_{n \to +\infty} \frac{1}{n} = 0$ donc $\lim\limits_{n \to +\infty} \frac{V(X)}{na^2} = 0$
Alors d'après le théorème des gendarmes $\lim\limits_{n \to +\infty} P(|M_n - E(X)| \geq a) = 0$.

Exemple
On considère une urne contenant cinq boules noires et trois boules blanches.
On souhaite estimer la probabilité d'obtenir au moins six boules noires sur un ensemble de dix tirages avec remise.
Pour simuler cette expérience plusieurs fois, Adrien a écrit un programme Python.
En effectuant 100000 épreuves de dix tirages, il obtient une proportion $p \approx 0{,}6943$.
Sachant que la variance est 2,344, déterminer la probabilité de se tromper de plus d'un centième.

Solution
D'après la loi des grands nombres, la probabilité cherchée est proche de 0,6943.
La probabilité de s'écarter de plus de $a = 0{,}01$ est inférieure à $\frac{V(X)}{n \times a^2} = \frac{2{,}344}{100000 \times 0{,}01^2} = 0{,}2344$.