COURS HACHETTE

soutienscolaire@monsieurtahar.com

CHAPITRE 14

LOIS DES GRANDS NOMBRES

1. Transformations de variables aléato

1. Transformation affine

Définition

Soit X une variable aléatoire. Soient a et b deux nombres réels.
On note x_1 ; x_2 ; … ; x_n les valeurs prises par X.
La variable aléatoire Y définie par $\boldsymbol{Y = aX + b}$ est la variable aléatoire qui prend pour valeurs les réels $y_i = ax_i + b$, pour tout i allant de 1 à n.

Propriétés

Soit X une variable aléatoire et Y la variable aléatoire définie par $Y = aX + b$, où a et b sont deux réels.

- L'espérance de Y est $E(Y) = E(aX + b) = aE(X) + b$.
- La variance de Y est $V(Y) = V(ax + b) = V(aX) = a^2V(X)$.

Exemple

Soit X une variable aléatoire qui suit la loi binomiale de paramètres $n = 2$ et $p = 0{,}2$ et soit $Y = 4X - 0{,}6$.
On a $E(X) = n \times p = 0{,}4$ et $V(X) = n \times p \times (1 - p) = 0{,}32$.
On a alors :

$$E(Y) = E(4X - 0{,}6) = 4E(X) - 0{,}6 = 4 \times 0{,}4 - 0{,}6 = 1$$
$$\text{et } V(Y) = V(4X) = 4^2V(X) = 16 \times 0{,}32 = 5{,}12.$$

2. Somme de deux variables aléatoires

Définition

Lorsque X et Y sont deux variables aléatoires, $\boldsymbol{X + Y}$ est la variable aléatoire qui prend pour valeurs les sommes des valeurs possibles de X et de Y.

Propriété

Soient X et Y deux variables aléatoires.
On a $E(X + Y) = E(X) + E(Y)$.

Exemple

X est une variable aléatoire qui suit la loi binomiale de paramètres $n = 200$ et $p = 0{,}2$ et Y une variable aléatoire qui suit la loi binomiale de paramètres $n = 100$ et $p = 0{,}5$.
$X + Y$ est une variable aléatoire dont les valeurs possibles sont $\{0 ; 1 ; 2 ; \ldots ; 300\}$.

$$E(X + Y) = E(X) + E(Y) = 200 \times 0{,}2 + 100 \times 0{,}5 = 90$$

Remarque

Les propriétés $E(AX) = aE(X)$ et $E(X + Y) = E(X) + E(Y)$ s'appellent la linéarité de l'espérance.

Propriété

Soient X et Y deux variables aléatoires. On suppose que X et Y sont associées à deux expériences aléatoires dont les conditions de réalisation sont **indépendantes.**
On a $V(X + Y) = V(X) + V(Y)$.
On dit alors que les variables aléatoires sont indépendantes.

Remarque

Dans le cas où les expériences ne sont pas indépendantes, il se peut que $V(X + Y) \neq V(X) + V(Y)$.

Méthode 1 Appliquer la linéarité de l'espérance

Soit X une variable aléatoire telle que $E(X) = 10$ et $V(X) = 4$. Soit Y une autre variable aléatoire telle que $E(Y) = 0$ et $V(Y) = 1$.

Calculer l'espérance et la variance des variables aléatoires suivantes.

1. $Z = 2X - 15$
2. $T = X + Y$
3. $U = -5Y$

Solution commentée

1. Par linéarité de l'espérance, on a $E(Z) = 2E(X) - 15 = 2 \times 10 - 15 = 5$.
$V(Z) = 2^2\, V(X) = 16$.

2. Par linéarité de l'espérance, on a $E(T) = E(X) + E(Y) = 10 + 0 = 10$.
En l'absence d'information sur l'indépendance des expériences aléatoires associées aux variables X et Y, on ne peut appliquer la formule sur la variance de la somme. On ne peut donc pas déterminer $V(T)$.

3. Par linéarité de l'espérance, on a $E(U) = -5E(Y) = 0$ et $V(U) = (-5)^2\, V(Y) = 25 \times 1 = 25$.

Méthode 2 Calculer l'espérance et la variance

On considère le jeu consistant à lancer 20 fois un dé.

À chaque lancer, on gagne 6 € si on obtient « 5 » ou « 6 » et 3 € sinon.

La participation au jeu est de 70 €.

1. Quel gain peut-on espérer ?
2. Quelle est la variance du gain ?

Solution commentée

1. Pour tout i allant de 1 à 20, on note X_i la variable aléatoire qui prend pour valeurs la somme en euro gagnée au i-ème lancer, pour $1 \leqslant i \leqslant 20$. X_i prend pour valeurs 3 ou 6.
La loi de X_i est $P(X_i = 6) = \frac{2}{6} = \frac{1}{3}$, $P(X_i = 3) = \frac{4}{6} = \frac{2}{3}$.
Il y a 20 répétitions indépendantes de la même expérience aléatoire qui consiste à lancer le dé.
Le gain algébrique est la variable aléatoire $G = X_1 + \ldots + X_{20} - 70$.

On cherche $E(G)$.
Pour tout i allant de 1 à 20, on a :
$$E(X_i) = 6 \times \frac{1}{3} + 3 \times \frac{2}{3} = 4.$$
Par linéarité de l'espérance, on a :
$E(G) = E(X_1 + \ldots + X_{20} - 70) = E(X_1) + \ldots + E(X_{20}) - 70 = 20 \times E(X_1) - 100 = 80 - 70 = 10$.
On peut donc espérer gagner 10 €.

2. On cherche $V(G)$.
Les expériences aléatoires étant indépendantes, on peut utiliser la propriété sur la variance de la somme.
$V(G) = V(X_1 + \ldots + X_{20} - 70) = V(X_1 + \ldots + X_{20}) = V(X_1) + \ldots + V(X_{20})$
Pour tout i allant de 1 à 20, on a :
$$V(X_i) = (6-4)^2 \times \frac{1}{3} + (3-4)^2 \times \frac{2}{3} = 4 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{6}{3} = 2.$$
On a donc $V(G) = V(X_1) + \ldots + V(X_{20}) = 20 \times V(X_1) = 40$.

2. Inégalités de concentration

1. Inégalité de Markov

Propriété (inégalité de Markov)

DÉMO en ligne

Soit X une variable aléatoire à valeurs positives et soit a un nombre réel strictement positif.

On a $P(X \geqslant a) \leqslant \dfrac{E(X)}{a}$.

Interprétation

Ce résultat signifie que la probabilité que X prenne des valeurs plus grandes que a est d'autant plus petite que a est grand. Autrement dit, la probabilité que X prenne des valeurs plus grandes que a est d'autant plus petite que a est grand.

Remarque

Si $a \leqslant E(X)$, l'inégalité de Markov n'a pas d'intérêt (on dit qu'elle est triviale).

En effet la borne $\dfrac{E(X)}{a}$ est alors supérieure à 1 et donc, nécessairement, à la probabilité $P(X \geqslant a)$.

Exemple

Soit X une variable aléatoire positive d'espérence 1.
D'après l'inégalité de Markov, on a $P(X \geqslant 100) \leqslant 0{,}01$.
Autrement dit, une variable aléatoire positive dont l'espérance vaut 1 a au plus une chance sur 100 de dépasser 100.

2. Inégalité de Bienaymé-Tchebychev

Propriété

DÉMO p. 445

Soit X une variable aléatoire et soit t un nombre réel strictement positif.

On a $P(|X - E(X)| \geqslant t) \leqslant \dfrac{V(X)}{t^2}$.

Interprétation

La probabilité que les valeurs prises par X s'écartent d'au moins t de l'espérance $E(X)$ est d'autant plus petite que t est grand.

Remarques

- $1 - P(|X - E(X)| \geqslant t) = P(E(X) - t < X < E(X) + t)$.
- On dit que $[E(X) - t\ ;\ E(X) + t]$ est un **intervalle de fluctuation** de X.
- Les inégalités de Markov et de Bienaymé-Tchebychev font partie de ce qu'on appelle des inégalités de « concentration ».

Exemple

Si $t = 2\sigma(X)$ où $\sigma(X)$ est l'écart-type de la variable X, alors :

$$P(|X - E(X)| \geqslant 2\sigma(X)) \leqslant \frac{V(X)}{(2\sigma(X))^2} = \frac{1}{4}.$$

Autrement dit, la probabilité qu'une variable aléatoire prenne des valeurs éloignées de son espérance d'au moins le double de son écart type est inférieure à $\dfrac{1}{4} = 0{,}25$.

Méthode 1 Appliquer les inégalités de concentration

Le nombre de pièces fabriquées dans une usine en une journée suit une variable aléatoire d'espérance 50 et de variance 25.

- Donner deux majorations de la probabilité que la production dépasse sur une journée 75 pièces.

Solution commentée

On note X la variable aléatoire égale au nombre de pièces fabriquées un jour donné.
On cherche à majorer $P(X \geqslant 75)$.
D'après l'inégalité de Markov, avec $a = 75$, on a $P(X \geqslant 75) \leqslant \dfrac{E(X)}{75}$.

Or $\dfrac{E(X)}{75} = \dfrac{50}{75} = \dfrac{2}{3}$, ce qui donne une première majoration : $P(X \geqslant 75) \leqslant \dfrac{2}{3}$.

Une seconde majoration peut être obtenue à l'aide de l'inégalité de Bienaymé-Tchebychev. Pour cela, on fait apparaître la bonne « forme » : on dit que l'on centre la variable X en lui retranchant son espérance.

$P(X \geqslant 75) = P(X - 50 \geqslant 25) \leqslant P(|X - 50| \geqslant 25)$

Or $P(|X - 50| \geqslant 25) \leqslant \dfrac{V(X)}{25^2}$ d'après l'inégalité de Bienaymé-Tchebychev.

Donc $P(X \geqslant 75) \leqslant P(|X - E(X)| \geqslant 25) \leqslant \dfrac{V(X)}{25^2} = \dfrac{25}{25^2} = \dfrac{1}{25}$.

On a donc $P(X \geqslant 75) \leqslant \dfrac{1}{25}$ ce qui est une seconde majoration.

Remarque : Il apparaît clairement ici que l'inégalité de Bienaymé-Tchebychev fournit une majoration bien plus précise que l'inégalité de Markov. On privilégiera en général l'inégalité de Bienaymé-Tchebychev.

Méthode 2 Manipuler les inégalités de concentration

On lance 800 fois une pièce de monnaie non truquée.

- Donner une minoration de la probabilité que le nombre de « Pile » obtenus soit compris entre 381 et 419.

Solution commentée

On note X la variable aléatoire égale au nombre de Pile obtenus lors des 800 lancers de la pièce.
On cherche à minorer la probabilité $P(381 \leqslant X \leqslant 419)$.
Pour utiliser l'inégalité de Bienaymé-Tchebychev, il faut déterminer l'espérance et la variance de X.
La loi de X est la loi binomiale de paramètres $n = 800$ et $p = 0{,}5$.
Donc $E(X) = 800 \times 0{,}5 = 400$ et $V(X) = 800 \times 0{,}5 \times (1 - 0{,}5) = 200$.
Comme l'inégalité de Bienaymé-Tchebychev est une majoration de la probabilité, on utilise l'événement contraire :

$$P(381 \leqslant X \leqslant 419) = 1 - P(X \leqslant 380 \text{ ou } X \geqslant 420).$$

Il reste à faire apparaître $|X - E(X)|$ en « centrant » la variable X :

$$\begin{aligned} P(381 \leqslant X \leqslant 419) &= 1 - P(X - E(X) \leqslant 380 - E(X) \text{ ou } X - E(X) \geqslant 420 - E(X)) \\ &= 1 - P(X - E(X) \leqslant -20 \quad \text{ou } X - E(X) \geqslant 20) = 1 - P(|X - E(X)| \geqslant 20) \end{aligned}$$

L'inégalité de Bienaymé-Tchebychev donne $P(|X - E(X)| \geqslant 20) \leqslant \dfrac{V(X)}{20^2}$.

Or $\dfrac{V(X)}{20^2} = \dfrac{200}{400} = 0{,}5$, donc $P(|X - E(X)| \geqslant 20) \leqslant \dfrac{1}{2}$.
Par conséquent, $P(381 \leqslant X \leqslant 419) \geqslant 0{,}5$.
On a donc au moins une chance sur deux d'obtenir entre 381 et 419 Pile lors des 800 lancers de la pièce.

3. Loi des grands nombres

1. Cas particulier de la loi binomiale

Propriété

On considère un schéma de Bernoulli comportant n répétitions d'une épreuve de Bernoulli de paramètre p.
Pour tout i de 1 à n, on note X_i la variable aléatoire associée à la i-ème épreuve de Bernoulli qui prend la valeur 1 en cas de succès et 0 sinon. On a alors $P(X_i = 1) = p$.

- Chaque variable X_i (pour i allant de 1 à n) suit la **loi de Bernoulli** de paramètre p.
- La variable aléatoire $S_n = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n$ est égale au nombre de succès lors des n épreuves et suit la loi binomiale de paramètres n et p.

Définition et propriété

- On appelle **fréquence empirique** des variables aléatoires $X_1, X_2, \ldots, X_n$, la variable aléatoire M_n définie par $M_n = \frac{X_1 + \ldots + X_n}{n} = \frac{S_n}{n}$.
- Soit t un nombre réel strictement positif.

L'inégalité de Bienaymé-Tchebychev appliquée à S_n et M_n donne :

$$P(|S_n - np| \geqslant t) \leqslant \frac{np(1-p)}{t^2} \quad \text{et} \quad P(|M_n - p| \geqslant t) \leqslant \frac{p(1-p)}{nt^2}.$$

Remarques

- La première inégalité signifie que la probabilité que la variable S_n prenne des valeurs éloignées d'une valeur t de son espérance np est d'autant plus petite que t est grand.
- $|M_n - p| < t$ équivaut à $p - t < M_n < p + t$
- En utilisant l'événement contraire, si t est un nombre réel strictement positif, on a alors :

$$P(|S_n - np| < t) \geqslant 1 - \frac{np(1-p)}{t^2} \text{ et } P(|M_n - p| < t) \geqslant 1 - \frac{p(1-p)}{nt^2}.$$

2. Cas général

Définition

On considère n expériences aléatoires identiques et indépendantes. On note $X_1, X_2, \ldots, X_n$ les variables aléatoires associées à ces expériences, toutes de même loi.

On note $S_n = X_1 + \ldots + X_n$ et $M_n = \frac{S_n}{n}$. M_n s'appelle la **moyenne empirique** des variables $X_1, X_2, \ldots, X_n$.

Théorème : Loi des grands nombres

On considère une expérience aléatoire et X la variable aléatoire associée à cette expérience, d'espérance $E(X)$ et de variance $V(X)$. On répète n fois cette expérience de manière indépendante. On obtient un échantillon de taille n composé de n variables aléatoires $X_1, X_2, \ldots, X_n$.

- Les variables $X_1, X_2, \ldots, X_n$ ont la même loi (elles ont donc la même espérance $E(X)$ et la même variance $V(X)$)
- Pour tout nombre réel $t > 0$, $P(|M_n - E(X)| \geqslant t) \leqslant \frac{V(X)}{nt^2}$.

Autrement dit, on a $\lim_{n \to +\infty} P(|M_n - E(X)| \geqslant t) = 0$.

On dit que M_n **converge en probabilité** vers $E(X)$ lorsque n tend vers $+\infty$.

Méthode 1 Appliquer Bienaymé-Tchebychev avec une loi binomiale

On lance 100 fois un dé équilibré, à quatre faces de couleurs différentes et on note la couleur de la face cachée. On appelle S_n la variable aléatoire qui compte le nombre de fois où la face cachée est rouge au cours des 100 lancers.

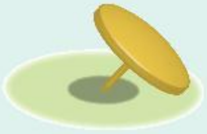
• En utilisant la loi des grands nombres dans le cas particulier d'une loi binomiale, déterminer une majoration de $P(21 \leqslant S_{100} \leqslant 29)$.

Solution commentée

• On reconnaît un schéma de Bernoulli : répétition d'une expérience aléatoire (lancer d'un dé) de manière indépendante (l'issue d'un lancer n'a pas d'influence sur l'issue suivante).
La probabilité de succès est $p = \frac{1}{4}$.
La variable aléatoire S_{100} suit la loi binomiale de paramètres $n = 100$ et $p = 0{,}25$.
On cherche à majorer $P(21 \leqslant S_{100} \leqslant 29)$. On applique la loi des grands nombres dans le cas particulier d'une loi binomiale : on cherche la valeur de t telle que $P(|S_{100} - 100p| \geqslant t) \leqslant \frac{100p(1-p)}{t^2}$.
On a $|S_{100} - 100p| < t \Leftrightarrow 100p - t < S_{100} < 100p + t \Leftrightarrow 25 - t < S_{100} < 25 + t$.
Comme on cherche la probabilité de l'événement $\{21 < S_n < 29\}$, on en déduit que $t = 5$.
On a $P(|S_{100} - 100 \times 0{,}25| \geqslant 5) \leqslant \frac{100 \times 0{,}25 \times (1 - 0{,}25)}{5^2}$, donc $P(|S_{100} - 100 \times 0{,}25| \geqslant 5) \leqslant 0{,}75$.
Donc $P(|S_{100} - 25| < 5) = P(21 \leqslant S_{100} \leqslant 29) \geqslant 1 - 0{,}75$.
La probabilité cherchée est donc d'au moins 25 %.

Méthode 2 Appliquer la loi des grands nombres dans le cas général

Justifier qu'en lançant une punaise un nombre de fois suffisamment grand, on peut estimer la probabilité qu'elle retombe en reposant sur sa pointe.



Solution commentée

On considère l'expérience aléatoire qui consiste à lancer une punaise et à regarder si elle retombe sur la « pointe » (succès de probabilité p inconnue), ou sur la « tête ». C'est une épreuve de Bernoulli de paramètre p. On répète cette expérience n fois de manière indépendante, où n est un entier naturel. On a alors un schéma de Bernoulli de paramètres n et p.
On note $X_1, X_2, \ldots, X_n$ les variables aléatoires associées à chaque répétition, qui prennent la valeur 1 en cas de « succès » et 0 sinon. Elles suivent toutes la loi de Bernoulli de paramètre p.
Les variables X_i ont toutes la même espérance : $E(X_1) = \ldots = E(X_n) = p$.
On pose $S_n = X_1 + \ldots + X_n$ et $M_n = \frac{S_n}{n}$
On peut appliquer la loi des grands nombres :

pour tout nombre réel t, on a $\lim\limits_{n \to +\infty} P(|M_n - p| \geqslant t) = 0$.

En lançant un grand nombre de fois la punaise, on peut donc estimer la valeur de p en observant la réalisation de la variable aléatoire M_n.