
Chapitre 16 : Variables aléatoires - Loi des grands nombres

Table des matières

Dans tout le chapitre, les variables aléatoires considérées sont définies sur un univers fini.

I Variables aléatoires	2
I.1 Quelques rappels : espérance, variance, écart-type	2
I.2 Opérations sur les variables aléatoires	2
I.3 Propriétés de l'espérance et de la variance	2
II Échantillon d'une loi de probabilité	3
II.1 Échantillon de taille n	3
II.2 Espérance, variance, écart-type	3
III Loi des grands nombres	3
III.1 Inégalité de Markov	3
III.2 Inégalité de Bienaymé-Tchebychev	4
III.3 Inégalité de concentration	4
III.4 Loi faible des grands nombres	5

Savoirs-faire

-

I Variables aléatoires

I.1 Quelques rappels : espérance, variance, écart-type

On considère une variable aléatoire X prenant un nombre fini de valeurs x_1, x_2, \dots, x_n .

x_i	x_1	x_2	\dots	x_n
$p(X = x_i)$	p_1	p_2	\dots	p_n

On définit :

- l'**espérance** de X : $E(X) = \sum_{i=1}^n p_i \times x_i$

L'espérance d'une variable aléatoire est, intuitivement, la valeur que l'on s'attend à trouver, **en moyenne**, si l'on répète un grand nombre de fois la même expérience aléatoire.

Elle correspond à une moyenne pondérée des valeurs que peut prendre cette variable.


- la **variance** de X : $V(X) = \sum_{i=1}^n p_i \times (x_i - E(X))^2 = \left(\sum_{i=1}^n p_i \times x_i^2 \right) - (E(X))^2$

La variance est un **caractère de dispersion**. Plus une variance est élevée plus la dispersion des observations est importante par rapport à la moyenne; elle est très sensible aux valeurs extrêmes.

- l'**écart-type** de X : $\sigma(X) = \sqrt{V(X)}$

En pratique c'est l'écart type qui est le plus utilisé; il s'exprime en effet avec les mêmes unités que les observations; la variance, quant à elle, s'exprime avec les unités au carré. L'écart type est la mesure la plus courante de la **dispersion ou de la répartition des données sur la moyenne**.

Exercice 1

On lance un dé cubique bien équilibré, possédant une face numérotée 1, deux faces numérotées 2 et trois faces numérotées 3. On appelle X le numéro de la face obtenue. 

1. Donner la loi de probabilité de X dans le tableau suivant :

Valeurs x_i de X			
$p(X = x_i)$			

2. Calculer l'espérance de X .
3. Calculer la variance et l'écart-type de X .

I.2 Opérations sur les variables aléatoires

Définition 1. Soient X et Y deux variables aléatoires, et soit a un réel. Les variables aléatoires $X + Y$ et aX prennent respectivement comme valeur pour un événement donné la somme des valeurs de X et Y et le produit de a par X .

Propriété 2. Soient X et Y deux variables aléatoires. Elles sont indépendantes si quelles que soient les valeurs x et y prises respectivement par X et Y , on a :

$$p((X = x) \cap (Y = y)) = p(X = x) \times p(Y = y)$$

I.3 Propriétés de l'espérance et de la variance

Propriété 3. Soient X et Y deux variables aléatoires. Soient $a, b \in \mathbb{R}$. Alors :

$$E(aX) = aE(X) \quad E(X+Y) = E(X)+E(Y) \quad (\text{Linéarité de l'espérance})$$

$$V(aX) = a^2V(X) \quad V(aX + b) = a^2V(X) \quad \sigma(aX) = |a|\sigma(X)$$

☞ Plus généralement, on a $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$

⚠ En général, $V(X + Y) \neq V(X) + V(Y)$!

Propriété 4. Soient X et Y deux variables aléatoires **indépendantes**. Alors :

$$V(X + Y) = V(X) + V(Y)$$

Exercice 2

On lance successivement deux fois le dé de l'exercice précédent, et on note X_1 et X_2 les nombres obtenus. On admet que X_1 et X_2 sont deux variables aléatoires indépendantes.

À l'issue des deux lancers, on calcule un score S , égal à deux fois le numéro obtenu au premier lancer, moins le numéro obtenu au second lancer.

1. Donner l'espérance et la variance des variables aléatoires X_1 et X_2 .
2. Exprimer S en fonction de X_1 et X_2 , puis déterminer son espérance et sa variance.

II Échantillon d'une loi de probabilité

II.1 Échantillon de taille n

Définition 5. Soit X une variable aléatoire.

Un **échantillon de taille n** de la loi de X est une liste (X_1, X_2, \dots, X_n) de variables aléatoires **indépendantes** et **identiques** suivant cette loi.

Définition 6.

- La variable aléatoire **somme** d'un échantillon de taille n de la loi de X est la variable aléatoire définie sur l'ensemble des échantillons de taille n par :

$$S_n = X_1 + X_2 + \dots + X_n$$

- La variable aléatoire **moyenne** d'un échantillon de taille n de la loi de X est la variable aléatoire définie sur l'ensemble des échantillons de taille n par :

$$M_n = \frac{1}{n} S_n$$

II.2 Espérance, variance, écart-type

Propriété 7. Avec les notations précédentes :

$$E(S_n) = n \times E(X) \quad V(S_n) = n \times V(X) \quad \sigma(S_n) = \sqrt{n} \times \sigma(X)$$

$$E(M_n) = E(X) \quad V(M_n) = \frac{1}{n} \times V(X) \quad \sigma(M_n) = \frac{1}{\sqrt{n}} \times \sigma(X)$$

☞ Comme $V(M_n) = \frac{1}{n} V(X)$, la variance de la moyenne diminue quand la taille de l'échantillon augmente. Cette variance quantifie la **fluctuation d'échantillonnage**, c'est à dire l'écart-moyen entre les valeurs prises par la variable aléatoire et son espérance.

☞ Si X suit une loi de Bernoulli de paramètre p , alors S_n suit une **loi binomiale** de paramètres (n, p) . Sachant qu'alors $E(X) = p$ et $V(X) = p(1-p)$, on retrouve les valeurs de $E(S_n) = np$ et $V(S_n) = np(1-p)$

Exercice 01 On lance 10 fois le dé de l'exercice 01, et on note X le nombre de 2 obtenus à l'issue des 10 lancers. On considère un échantillon $(X_1, X_2, \dots, X_{20})$ de taille 20 de X . On note :

$$S_{20} = X_1 + X_2 + \dots + X_{20} \quad M_{20} = \frac{X_1 + X_2 + \dots + X_{20}}{20}$$

1. Quelle est la loi de probabilité de X ? Préciser ses paramètres.
2. Donner l'espérance et la variance de X .
3. Donner l'espérance de S_{20} et M_{20} . Interpréter ces deux résultats dans le contexte de l'exercice.
4. Donner la variance de S_{20} et M_{20} .

III Loi des grands nombres

Lorsqu'on repète plusieurs fois un grand nombre de lancers d'une pièce bien équilibrée, on constate que la fréquence d'apparition de Pile semble très proche de $\frac{1}{2}$. Ce résultat peut-il se démontrer ? La réponse est (presque) oui.

III.1 Inégalité de Markov

Propriété 8 (Inégalité de Markov).

Soit X une variable aléatoire prenant uniquement des valeurs positives ou nulles.

Alors :

$$\forall \alpha > 0, \quad p(X \geq \alpha) \leq \frac{E(X)}{\alpha}$$

Démonstration. Soit $\alpha > 0$. On suppose que X prend n valeurs $0 \leq x_1 < x_2 < \dots < x_n$. Pour tout entier naturel k compris entre 1 et n , on pose $p_k = p(X = x_k)$.

1er cas : $\alpha > x_n$. Dans ce cas, $p(X \geq \alpha) = 0$ et l'inégalité est clairement vraie.

2ème cas : $\alpha < x_1$. Dans ce cas, $p(X \geq \alpha) = 1$ et pour tout entier k entre 1 et n , $\alpha < x_k$ d'où :

$$\begin{aligned} E(X) = p_1x_1 + p_2x_2 + \dots + p_nx_n &> \alpha p_1 + \alpha p_2 + \dots + \alpha p_n \\ &\geq \alpha (p_1 + p_2 + \dots + p_n) \\ &\geq \alpha \times 1 \\ &\geq \alpha \times p(X \geq \alpha) \end{aligned}$$

D'où (comme $\alpha > 0$) : $p(X \geq \alpha) \leq \frac{E(X)}{\alpha}$.

3ème cas : il existe un entier naturel i tel que $x_i < \alpha \leq x_{i+1}$. En particulier, pour tout entier k compris entre $i + 1$ et n , $x_k \geq \alpha$. D'où :

$$\begin{aligned} E(X) = p_1x_1 + p_2x_2 + \dots + p_nx_n &> p_1x_1 + p_2x_2 + \dots + p_ix_i + p_{i+1}x_{i+1} + \dots + p_nx_n \\ &\geq p_1x_1 + p_2x_2 + \dots + p_ix_i + \alpha (p_{i+1} + \dots + p_n) \\ &\geq p_1x_1 + p_2x_2 + \dots + p_ix_i + \alpha \times p(X \geq \alpha) \\ &\geq \alpha \times p(X \geq \alpha) \end{aligned}$$

Cette dernière inégalité est vraie car les x_i sont tous positifs ou nuls.

On retrouve ainsi l'inégalité souhaitée. □

Exercice 3

On lance 12 fois un même dé cubique bien équilibré, dont les faces sont numérotées de 1 à 6.

- Démontrer sans calculs, à l'aide de l'inégalité de Markov, que la probabilité d'obtenir au moins 10 fois le nombre 6 est inférieure à 20%.
- Déterminer la valeur de cette probabilité.

☞ L'inégalité de Markov n'est pas très intéressante en elle-même, mais elle permet de démontrer d'autres inégalités très importantes.

III.2 Inégalité de Bienaymé-Tchebychev

Propriété 9 (Inégalité de Bienaymé-Tchebychev).

Soit X une variable aléatoire. Alors :

$$\forall \alpha > 0, \quad p(|X - E(X)| \geq \alpha) \leq \frac{V(X)}{\alpha^2}$$

Démonstration. Soit $\alpha > 0$. On applique l'inégalité de Markov à la variable aléatoire

$Y = (X - E(X))^2$, qui est bien à valeurs positives ou nulles. L'inégalité reste vraie pour α^2 à la place de α :

$$p(Y \geq \alpha^2) \leq \frac{E(Y)}{\alpha^2}$$

Or :

- $Y \geq \alpha^2 \iff (X - E(X))^2 \geq \alpha^2 \iff |X - E(X)| \geq \alpha$ (car $\alpha > 0$)
- $E(Y) = E((X - E(X))^2) = V(X)$ par définition de la variance

D'où l'inégalité cherchée. □

Exercice 4

Le nombre de pièces sortant d'une usine en une journée est une variable aléatoire X d'espérance $E(X) = 100$ et d'écart-type $\sigma(X) = 10$. On cherche à estimer la probabilité $p(X \geq 150)$ que la production d'un jour donné dépasse 150 pièces.

- À l'aide de l'inégalité de Markov, donner une majoration de $p(X \geq 150)$.
- Justifier que $|X - 100| \geq 50 \implies X \geq 150$; en déduire une majoration de cette même probabilité à l'aide de l'inégalité de Bienaymé-Tchebychev.
- À l'aide de l'inégalité de Bienaymé-Tchebychev, démontrer que $p(|X - 100| < 20)$ est supérieure ou égale à 75%. Donner une interprétation de ce résultat dans le contexte de l'exercice.

III.3 Inégalité de concentration

Propriété 10 (Inégalité de concentration).

Soit M_n une variable aléatoire moyenne d'un échantillon de taille n d'une v.a. X .

Alors :

$$\forall \alpha > 0, \quad p(|M_n - E(X)| \geq \alpha) \leq \frac{V(X)}{n\alpha^2}$$

Démonstration. La démonstration est immédiate : on applique l'inégalité de Bienaymé-Tchebychev à la variable aléatoire M_n , et on sait que $E(M_n) = E(X)$ et $V(M_n) = \frac{1}{n}V(X)$. D'où le résultat. □

Exercice 5

Soit X une variable qui suit une loi de Bernoulli de paramètre $p = 0,5$.

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

Déterminer la valeur de n pour que la probabilité que M_n soit dans l'intervalle $[0,45; 0,55]$ soit supérieure à 0,95.

☞ On peut vérifier à l'aide de simulations que la valeur de n trouvée est très largement surestimée. Toutes ces inégalités sont loin d'être optimales.

☞ Et si $n \rightarrow +\infty$?

III.4 Loi faible des grands nombres

Théorème 11 (Loi (faible) des grands nombres).

Soit M_n une variable aléatoire moyenne d'un échantillon de taille n d'une v.a. X .

Alors :

$$\forall \alpha > 0, \quad \lim_{n \rightarrow +\infty} p(|M_n - E(X)| \geq \alpha) = 0$$

Démonstration. C'est le théorème des gendarmes appliqué à l'inégalité de concentration ! (une probabilité est toujours positive) \square

☞ Ce dernier théorème nous affirme que lorsque n devient grand, la probabilité que M_n soit hors de l'intervalle $[E(X) - \alpha; E(X) + \alpha]$ tend vers 0.

En d'autres termes, il nous permet de justifier (en termes de probabilités) que lorsqu'on répète un grand nombre de fois une expérience aléatoire, la valeur moyenne d'une variable aléatoire n'est jamais trop loin de son espérance mathématique...