7

VARIABLES ALÉATOIRES

Résumé

Dans ce chapitre, nous améliorons notre formalisme pour décrire des événements probabilistes à l'aide des variables aléatoires et nous introduisons, l'espérance, l'un des outils les plus importants de toute la théorie des probabilités.

On se restreint à des expériences aléatoires sur un univers Ω **fini**.

1 Notions de variables aléatoires

1.1 Variable aléatoire réelle

Définition

Une **variable aléatoire réelle** X est une **fonction** définie sur Ω, à valeurs dans $\mathbf{R}$, de telle sorte que tout élément de Ω est associé à un unique nombre réel.

Exemples

- On lance une pièce de monnaie (équilibrée ou non) 10 fois à la suite et on note X le nombre de piles. X est une variable aléatoire réelle.
- Un professeur tire au sort un élève dans une classe de 32 élèves où chaque élève dispose de son propre numéro compris entre 1 et 32.

  On note X le numéro de l'élève tiré au sort : c'est une variable aléatoire réelle.

Remarques

- Pour une même expérience aléatoire, on peut définir différentes variables aléatoires. On aurait pu définir Y le nombre de faces dans le premier exemple donné précédemment. Ainsi, on aurait eu $X + Y = 10$.
- On peut définir des événements à partir d'une variable aléatoire X comme $\{X = a\}$ correspondant aux issues telles que X prenne la valeur a ou $\{X > a\}$ correspondant aux issues telles que X prenne des valeurs strictement supérieures à a.
- La probabilité de ces événements sera notée $\mathbb{P}(X = a)$ et $\mathbb{P}(X > a)$.

1.2 Loi de probabilité

Définition

Donner la **loi de probabilité** d'une variable aléatoire X, c'est donner la probabilités de tous les événements $\{X = a\}$ définis par X.
On la présente usuellement sous forme de tableau où x_i sont les valeurs prises par X et p_i les probabilités $\mathbb{P}(X = x_i)$.

x_i	x_1	x_2	$\cdots$	x_n
p_i	$\mathbb{P}(X = x_1)$	$\mathbb{P}(X = x_2)$	$\cdots$	$\mathbb{P}(X = x_n)$

Exemple Une station de lavage automobile a constaté que, parmi ses clients :

- 90% lavent la carrosserie de leur voiture ;
- 30% nettoient l'intérieur de leur voiture ;
- 20 en % lavent la carrosserie et nettoient l'intérieur de leur voiture.

Le coût du lavage de la carrosserie est de 5€, celui du nettoyage de l'intérieur est de 2€.
On note X la variable aléatoire modélisant la dépense, en euro, d'un client de la station choisi au hasard.

x_i	2	5	7
p_i	$\mathbb{P}(X = 2)$	$\mathbb{P}(X = 5)$	$\mathbb{P}(X = 7)$

On sait déja que $\mathbb{P}(X = 7) = 0{,}2$. De plus, il y a 10 % des clients qui nettoient l'intérieur sans laver la carrosserie, c'est-à-dire, $\mathbb{P}(X = 2) = 0{,}1$. Enfin, 70 % des clients lavent la carrosserie sans nettoyer l'intérieur donc $\mathbb{P}(X = 5) = 0{,}7$.

x_i	2	5	7
p_i	0,1	0,7	0,2

Propriété

La somme des probabilités p_i est égale à 1.
Autrement dit, $\sum_{i=0}^{n} \mathbb{P}(X = x_i) = 1$

Démonstration. Les évènements $\{X = x_i\}$ pour $1 \leqslant i \leqslant n$ forment une partition de l'univers. □

Exemple On vérifie bien dans l'exemple de la laverie que $0{,}1 + 0{,}7 + 0{,}2 = 1$.

2 Espérance, variance et écart-type

Définition | Espérance

L'**espérance** de la variable aléatoire X est le nombre réel $\mathbb{E}[X]$ défini par :

$$\mathbb{E}[X] = \sum_{i=0}^{n} \mathbb{P}(X = x_i) \times x_i = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n.$$

Remarque $\mathbb{E}[X]$ peut être vu comme une « **moyenne probabiliste** ». En effet, on peut définir la moyenne d'une série statistique par la formule $\overline{x} = \sum_{i=0}^{n} f_i x_i$ où les f_i sont les fréquences.

Définitions | Variance et écart-type

La **variance** de la variable aléatoire X est le réel **positif** $\mathrm{Var}(X)$ défini par :

$$\mathrm{Var}(X) = \sum_{i=0}^{n} p_i \left(x_i - \mathbb{E}[X]\right)^2 = p_1 \left(x_1 - \mathbb{E}[X]\right)^2 + \cdots + p_n \left(x_n - \mathbb{E}[X]\right)^2.$$

L'**écart-type** de X, $\sigma(X)$ est la racine carrée de la variance de X, c'est-à-dire :

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

Remarque Ici, le parallèle avec les statistiques est plus évident. Les deux indicateurs variance et écart-type mesurent la **dispersion** autour de l'espérance.

Exemple Revenons à l'exemple de la laverie et calculons $\mathbb{E}[X]$, $\mathrm{Var}(X)$ et $\sigma(X)$ à partir de la loi de probabilité.

x_i	2	5	7
p_i	0,1	0,7	0,2

On a :

$$\mathbb{E}[X] = 0{,}1 \times 2 + 0{,}7 \times 5 + 0{,}2 \times 7 = 5{,}1.$$

Ainsi, les clients dépensent en moyenne 5€10.
Enfin,

$$\mathrm{Var}(X) = 0{,}1 \times (2 - 5{,}1)^2 + 0{,}7 \times (5 - 5{,}1)^2 + 0{,}2 \times (7 - 5{,}1)^2 = 1{,}69$$

et

$$\sigma(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{1{,}69} = 1{,}3.$$

On a une dispersion moyenne de 1€ autour de l'espérance.

Remarque Soient a et b deux réels. On peut définir Y=aX+b comme étant la variable aléatoire vérifiant la loi de probabilité suivante.

y_i	$ax_1 + b$	$ax_2 + b$	$\cdots$	$ax_n + b$
$\mathbb{P}(Y = y_i)$	p_1	p_2	$\cdots$	p_n

Théorème

Soient a et b deux réels. On a :

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

et

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X).$$

Exemple Une station de lavage concurrente propose une tarification à moitié prix (lavage à 2€50 et nettoyage à 1€) mais demande 2€ pour entrer dans la station. On considère que les habitudes des clients sont les mêmes que pour la station précédente.

On note Y la variable aléatoire modélisant la dépense d'un client tiré au sort dans cette deuxième station de lavage.
Ainsi, $Y = \frac{1}{2}X + 2$ de loi de probabilité suivante.

y_i	3	4,5	6,5
p_i	0,1	0,7	0,2

$$\mathbb{E}[Y] = \frac{1}{2}\mathbb{E}[X] + 2 = \frac{1}{2} \times 5{,}1 + 2 = 4{,}55$$

Les clients dépensent donc en moyenne moins dans cette deuxième station.
Enfin, $\mathrm{Var}(Y) = \left(\frac{1}{2}\right)^2 \mathrm{Var}(X) = \frac{1}{4} \times 1{,}69 = 0{,}4225$ donc $\sigma(Y) = \sqrt{0{,}4225} = 0{,}65$.
La dispersion est plus faible que pour la première station.