Lycée Lucie Aubrac - 1re

Chapitre 5
Variables aléatoires

Table des matières

1 Rappels de Seconde (Indicateurs statistiques)

Dans l'ensemble de cette partie, on considère l'exemple de la série statistiques suivante, donnant les distances des planètes du système solaire au Soleil :

Planète	Distance au soleil en millions de km
Mercure	58
Vénus	108
Terre	150
Mars	228
Jupiter	778
Saturne	1429
Uranus	2871
Neptune	4503

1.1 Indicateurs de position

1.1.1 Moyenne

Définition 1

La **moyenne** d'une série statistique $x_1, \ldots, x_n$ se note $\overline{x}$ et est donné par la formule suivante :

$$\overline{x} = \frac{x_1 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Exemple.
En moyenne, la distance d'une planète du système solaire au Soleil est d'environ 1266 millions de km.

$$\overline{x} = \frac{58 + 108 + 150 + 228 + 778 + 1429 + 2871 + 4503}{8} \simeq 1266$$

1.1.2 Médiane

Définition 2

La **médiane** d'une série statistique $x_1, \ldots, x_n$ est une valeur qui permet de couper l'ensemble des valeurs en deux parties égales : la moitié des valeurs étant plus grande que la médiane et l'autre moitié étant moins grande.

Remarque.

- Lorsque la série comprend un nombre impair de valeurs, la médiane est simplement une des valeurs de la série.
- Lorsque la série comprend un nombre pair de valeurs, la médiane est la moyenne entre les deux valeurs du « milieu ».

Exemple.
Dans le cas des distances planète-Soleil, la médiane est la moyenne entre la 4e et la 5e plus grande valeur étant donné que la série comporte 8 valeurs.

$$m = \frac{228 + 778}{2} = 503 \text{ millions de km}$$

1.1.3 Quartiles

Définition 3

- Le **premier quartile** d'une série statistique est une valeur, notée Q_1 telle qu'au moins 25% des valeurs soient plus petites que Q_1.
- Le **troisième quartile** d'une série statistique est une valeur, notée Q_3 telle qu'au moins 75% des valeurs soient plus petites que Q_3.

Exemple.
Dans le cas des distances planète-Soleil,
Q_1 correspond à la moyenne entre la deuxième et la troisième valeur (car $8 \times \frac{1}{4} = 2$) et Q_3 correspond à lamoyenne entre la sixième et la sixième valeur (car $8 \times \frac{3}{4} = 6$).

$$Q_1 = 129$$

$$Q_3 = 2150$$

Remarque.
L'intervalle interquartile est l'intervale $[Q_1\,;Q_3]$. Il contient donc 50% des valeurs de la série statistique.

1.1.4 Valeur minimale et maximale

Définition 4

La **valeur minimale** d'une série statistique est la plus petite valeur de la série. la **valeur maximale** est la plus grande valeur de la série.

Exemple.
Dans le cas des distances planète-Soleil, la valeur minimale est 58 et la valeur maximale est 4503.

1.2 Indicateurs de dispersion

1.2.1 Étendue

Définition 5

L'**étendue** d'une série statistique est la différence entre la valeur maximale et la valeur minimale.

Exemple.
Dans le cas des distances planète-Soleil, l'étendue est de $4503 - 58 = 4445$ millions de km.

1.2.2 Écart interquartile

Définition 6

L' **écart interquartile** est $Q_3 - Q_1$.

Exemple.
Dans le cas des distances planète-Soleil, l'écart interquartile est $2150 - 129 = 2021$ millions de km.

1.2.3 Écart-type

Définition 7

- La **variance** V d'une série statistique $x_1, \ldots, x_n$ est définie par :

$$V = \frac{(x_1 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n} = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}{n}$$

- L'**écart type** σ d'une série statistique est définie par :

$$\sigma = \sqrt{V}$$

Remarque.
On retiendra que la variance est « la moyenne des carrés des écarts à la moyenne »

Exemple.
Dans le cas des distances planète-Soleil, on a :

$\overline{x} = 1266$

$$V = \frac{(58 - 1266)^2 + (108 - 1266)^2 + (150 - 1266)^2 + \ldots + (4503 - 1266)^2}{8} \simeq 2304324$$

$\sigma = \sqrt{V} \simeq 1518$

Remarque.
La donnée de la médiane et de l'écart interquartile n'est pas sensible aux valeurs extrêmes, contrairement à la donnée de la moyenne et de l'écart-type.

1.3 Échantillonnage

Proposition 1 – (admise)

Si l'effectif d'une série statistique est suffisamment grand, et si le phénomène étudié suit une loi de Gauss (la répartition des valeurs se fait selon une courbe en cloche) :

- L'intervalle $\left[\overline{x} - \sigma\,;\overline{x} + \sigma\right]$ contient environ 68% des valeurs.
- L'intervalle $\left[\overline{x} - 2\sigma\,;\overline{x} + 2\sigma\right]$ contient environ 95% des valeurs.
- L'intervalle $\left[\overline{x} - 3\sigma\,;\overline{x} + 3\sigma\right]$ contient environ 99% des valeurs.

Exemples.
Les phénomènes suivants suivent une loi de Gauss :

- Pourcentage de Pile obtenus lors du lancer de 50 pièces ;
- La taille des femmes dans la population française ;
- Le nombre de votants par bureau de vote lors d'une élection ;
- La taille des tiges des coquelicots dans un champ ;
- Les notes d'une évaluation.

Ainsi, si la moyenne a une évaluation est $\overline{x} = 10$ et l'écart-type est $\sigma = 3$, cela signifie qu'environ 68% des notes sont comprises entre 7 et 13.

Proposition 2 – Loi des Grands Nombres (admise)

On répète n fois une même expérience aléatoire de façon indépendante et dont la probabilité de succès est p. Expérimentalement, plus n sera grand, plus la fréquence du succès se rapprochera de p. De plus, on peut considérer que la fréquence de succès suit une loi de Gauss et on a :

$$\sigma \simeq \sqrt{\frac{p(1-p)}{n}}.$$

Remarque.
En général, lorsque $n \geqslant 25$, on considère que n est suffisament grand pour utiliser la loi des Grands Nombres.

Exemple.
On lance 200 fois un dé équilibré à six faces et on compte le nombre de 6 obtenus. Déterminer avec une marge d'erreur de 5%, un intervalle contenant la fréquence de 6 obtenus.

Solution :
Les lancers étant indépendants les uns des autres, on peut utiliser la loi des Grands Nombres. On a $p = \frac{1}{6}$ donc :

$$\sigma \simeq \sqrt{\frac{\frac{1}{6}(1-\frac{1}{6})}{200}} \simeq 0{,}0264.$$

Par conséquent, cela signifie que l'intervalle suivant contient 95% des fréquences expérimentales :

$$\left[p-2\sigma\,;p+2\sigma\right] \simeq \left[\frac{1}{6}-2\times 0{,}0264\,;\frac{1}{6}+2\times 0{,}0264\right] \simeq \left[0{,}114\,;0{,}219\right].$$

Autrement dit, avec une marge d'erreur de 5%, l'intervalle $\left[0{,}114\,;0{,}219\right]$ contient la fréquence de 6 obtenus.

2 Variables aléatoires

Introduction

Une association étudiante souhaite organiser un jeu de dés afin de récolter de l'argent. Les règles du jeu sont les suivantes. Le joueur lance un dé équilibré à six faces.

- Il gagne s'il obtient un 6 et l'association lui donne 12 euros.
- Il perd dans tous les autres cas et il donne 6 euros à l'association.

L'association organise 500 parties de ce jeu. Combien peut-elle espérer gagner en moyenne ?

2.1 Notion de variable aléatoire

Définition 8

Une **variable aléatoire** sur Ω est une fonction définie sur Ω à valeur dans $\mathbb{R}$.
À chaque issue de Ω, on associe un nombre réel.

Définition 9

Soit x un réel. L'événement « $\mathrm{X} = x$ » est l'ensemble des issues de Ω auxquelles on associe le réel x.

Exemple.
Avec l'exemple d'introduction :
L'univers est $\Omega = \{1\,;2\,;3\,;4\,;5\,;6\}$.
On note X la variable aléatoire correspondant au gain en euros de l'association sur une partie.
L'événement « $\mathrm{X} = 6$ » correspond aux issues 1, 2, 3, 4, 5.
L'événement « $\mathrm{X} = -12$ » correspond à l'issue 6.

2.2 Loi de probabilité d'une variable aléatoire

Définition 10

Pour définir la loi de probabilité d'une variable aléatoire X :

- On liste les valeurs x_i possibles pour X ;
- On détermine les probabilités $\mathrm{P}(\mathrm{X} = x_i)$.

Méthode – Déterminer la loi de probabilité d'une variable aléatoire

On forme le tableau suivant :

x_i	x_1	x_2	$\dots$	x_n
$P(\mathrm{X} = x_i)$	p_1	p_2	$\dots$	p_n

Remarque.
Dans le tableau, on a :

$$p_1 + p_2 + \dots + p_n = 1.$$

Exemple.
Avec l'exemple d'introduction : Si X est le gain de l'association sur une partie, la loi de probabilité de X est donnée par le tableau suivant :

x_i	-12	6
$P(\mathrm{X} = x_i)$	$\frac{1}{6}$	$\frac{5}{6}$

2.3 Espérance

Définition 11

Soit X une variable aléatoire. On appelle **espérance** de X le nombre :

$$E(X) = p_1x_1 + p_2x_2 + \ldots + p_nx_n = \sum_{i=1}^{n} p_ix_i$$

Remarque.
Le mot « espérance » vient du langage des jeux : lorsque X désigne le gain, $E(X)$ est le gain moyen que peut espérer un joueur lorsque le nombre de parties est important.

Exemple.
Avec l'exemple d'introduction :

$$E(X) = p_1x_1 + p_2x_2 = (-12) \times \frac{1}{6} + 6 \times \frac{5}{6} = 3.$$

Cela signifie que le jeu est défavorable au joueur et qu'en moyenne, l'association gagnera 3 euros par partie. Par conséquent, sur 500 parties, elle peut espérer gagner environ 1500 euros.

2.4 Variance et Écart-type

Définition 12

Soit X une variable aléatoire.

- On appelle **variance** de X le nombre :

$$V(X) = p_1(x_1-E(X))^2+p_2(x_2-E(X))^2+\ldots+p_n(x_n-E(X))^2 = \sum_{i=1}^{n} p_i(x_i-E(X))^2$$

- On appelle **écart-type** de X le nombre $\sigma(X) = \sqrt{V(X)}$.

Exemple.
Avec l'exemple d'introduction :

$$\begin{aligned} V(X) &= p_1(x_1 - E(X))^2 + p_2(x_2 - E(X))^2 \\ &= \frac{1}{6}(-12-3)^2 + \frac{5}{6}(6-3)^2 \\ &= 45 \end{aligned}$$

On a ainsi, $\sigma(X) = \sqrt{V(X)} = \sqrt{45} \simeq 6{,}71$.

2.5 Transformation affine d'une variable aléatoire

Proposition 3

Si X est une variable aléatoire prenant les valeurs $x_1, x_2, \ldots, x_n$ avec les probabilités $p_1, p_2, \ldots, p_n$ et si a et b des réels alors la loi de la variable aléatoire $Y = aX + b$ est donnée par le tableau suivant :

y_i	$ax_1 + b$	$ax_2 + b$	$\ldots$	$ax_n + b$
$P(Y = y_i)$	p_1	p_2	$\ldots$	p_n

Exemple.
Avec l'exemple d'introduction, supposons que l'on multiplie les gains et les pertes par 2 pour chaque partie, on obtient ainsi une nouvelle variable aléatoire $Y = 2X$.

Proposition 4

Si X est une variable aléatoire et a et b des réels alors

$$E(aX + b) = aE(X) + b.$$

Démonstration.

$$\begin{aligned}
E(aX+b) &= p_1(ax_1+b) + p_2(ax_2+b) + \ldots + p_n(ax_n+b) \\
&= p_1ax_1 + p_1b + p_2ax_2 + p_2b + \ldots + p_nax_n + p_nb \\
&= a\Big(p_1x_1 + p_2x_2 + \ldots + p_nx_n\Big) + \Big(p_1 + p_2 + \ldots + p_n\Big)b \\
&= aE(X) + b \quad (\text{car } p_1 + p_2 + \ldots + p_n = 1).
\end{aligned}$$

□

Proposition 5

Si X est une variable aléatoire et a et b des réels alors

$$V(aX + b) = a^2V(X).$$

Démonstration.

$$\begin{aligned}
&V(aX+b) \\
&= p_1\Big(ax_1 + b - E(aX+b)\Big)^2 + p_2\Big(ax_2 + b - E(aX+b)\Big)^2 + \ldots + p_n\Big(ax_n + b - E(aX+b)\Big)^2 \\
&= p_1\Big(ax_1 + b - aE(X) - b\Big)^2 + p_2\Big(ax_2 + b - aE(X) - b\Big)^2 + \ldots + p_n\Big(ax_n + b - aE(X) - b\Big)^2 \\
&= p_1\Big(ax_1 - aE(X)\Big)^2 + p_2\Big(ax_2 - aE(X)\Big)^2 + \ldots + p_n\Big(ax_n - aE(X)\Big)^2 \\
&= p_1\Big(a(x_1 - E(X))\Big)^2 + p_2\Big(a(x_2 - E(X))\Big)^2 + \ldots + p_n\Big(a(x_n - E(X))\Big)^2 \\
&= a^2\Bigg(p_1(x_1 - E(X))^2 + p_2(x_2 - E(X))^2 + \ldots + p_n(x_n - E(X))^2\Bigg) \\
&= a^2V(X)
\end{aligned}$$

$\square$

Proposition 6

Si X est une variable aléatoire et a et b des réels alors

$$\sigma(aX+b) = |a|\sigma(X).$$

Démonstration.

$$\begin{aligned}\sigma(aX+b) &= \sqrt{V(aX+b)}\\ &= \sqrt{a^2V(X)}\\ &= \sqrt{a^2}\sqrt{V(X)}\\ &= |a|\sigma(X)\end{aligned}$$

$\square$

QCM d'entrainement

Savoir-faire du chapitre

- Modéliser une situation à l'aide d'une variable aléatoire.
- Déterminer la loi de probabilité d'une variable aléatoire.
- Calculer l'espérance, la variance et l'écart-type d'une variable aléatoire.
- Interpréter l'espérance d'une variable aléatoire.