

Une urne contient deux boules jaunes et huit boules rouges. Un joueur tire successivement cinq boules dans cette urne. Pour chaque boule jaune tirée, il gagne deux points et pour chaque boule rouge tirée, il perd trois points. On note  $X$ , le nombre aléatoire de boules jaunes tirées et  $Y$ , le nombre aléatoire de points obtenus par le joueur au terme de la partie.

**1**• Proposer un modèle probabiliste en supposant que les tirages sont faits avec remise. Déterminer les lois de  $X$  et de  $Y$  pour ce modèle.

**2**• Et si les tirages sont faits sans remise ?

**1**• Lorsque les tirages sont effectués *avec remise*, la composition de l'urne est la même au cours de l'expérience et tous les tirages sont donc effectués dans les mêmes conditions. Il est donc raisonnable de modéliser la suite des tirages par une suite  $(U_n)_{n \geq 1}$  de variables aléatoires indépendantes et de même loi définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .

• On ne s'intéresse qu'à la couleur des boules tirées et il n'y a que deux couleurs : on peut donc supposer que chaque variable aléatoire  $U_n$  suit une loi de Bernoulli, en convenant par exemple que l'évènement  $[U_n = 1] \in \mathcal{A}$  correspond au fait de tirer une boule jaune ("succès") lors du  $n$ -ième tirage et que l'évènement contraire

$$[U_n = 0] = [U_n = 1]^c \in \mathcal{A}$$

correspond au fait de tirer une boule rouge lors du  $n$ -ième tirage. Il reste alors à choisir la valeur du paramètre  $p = \mathbf{P}(U_n = 1)$  de ces variables aléatoires de Bernoulli.

• Comme on ne dispose d'aucune information particulière, il est raisonnable de choisir le modèle le plus simple pour décrire un tirage : on suppose que chaque boule a la même probabilité d'être tirée et donc que la probabilité de tirer une boule d'une couleur donnée est égale à la proportion de boules de cette couleur dans l'urne. Autrement dit, on convient de poser

$$\forall n \in \mathbb{N}^*, \quad p = \mathbf{P}(U_n = 1) = \frac{2}{2+8} = \frac{1}{5} \quad \text{et} \quad q = \mathbf{P}(U_n = 0) = \frac{4}{5}.$$

• Il est important de retenir qu'on a fait ici un certain nombre d'hypothèses, sur lesquelles les calculs ultérieurs vont reposer, et qu'aucun argument mathématique ne peut justifier ces hypothèses : elles ne sont justifiées, et ne peuvent être justifiées, que par le bon sens.

• Le nombre  $X$  de boules jaunes tirées lors des cinq tirages consécutifs est donc égal au nombre de succès :

$$X = \sum_{n=1}^5 U_n.$$

D'après ce qui précède,  $X$  est donc une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbf{P})$  et, en tant que somme de 5 variables aléatoires de Bernoulli indépendantes de même loi  $\mathcal{B}(1/5)$ , la loi de  $X$  est la loi binomiale  $\mathcal{B}(5, 1/5)$ .

En particulier,  $\mathbf{E}(X) = np = 1$  et  $\mathbf{V}(X) = npq = 4/5$ .

• Le gain  $Y$  du joueur se déduit du nombre  $X$  de boules jaunes tirées et du nombre  $5 - X$  de boules rouges tirées :

$$Y = 2X + (-3)(5 - X) = -15 + 5X.$$

Par conséquent,  $Y$  est bien une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbf{P})$  et d'après les règles de calcul habituelles :

$$\mathbf{E}(Y) = -15 + 5\mathbf{E}(X) = -10, \quad \mathbf{V}(Y) = 5^2 \mathbf{V}(X) = 20.$$

**2**• Lorsque les tirages sont effectués *sans remise*, la composition de l'urne varie au cours de l'expérience et il n'est plus raisonnable de modéliser les différents tirages par des variables aléatoires indépendantes et de même loi.

• Quel que soit le modèle choisi, il est clair que la variable aléatoire  $X$  ne peut prendre que trois valeurs : 0, 1 et 2 (on ne peut pas tirer plus de boules jaunes qu'il n'y en a initialement dans l'urne). Il reste à choisir les valeurs de  $\mathbf{P}(X = 0)$ ,  $\mathbf{P}(X = 1)$  et  $\mathbf{P}(X = 2)$  d'une manière qui paraisse raisonnable.

• **Modèle A**

Un premier modèle repose sur la formule de Vandermonde :

$$\binom{10}{5} = \binom{2}{0} \binom{8}{5} + \binom{2}{1} \binom{8}{4} + \binom{2}{2} \binom{8}{3}.$$

On suppose ici que  $\Omega$  est l'ensemble des parties constituées de 5 boules prises dans un ensemble de 10 boules, dont 2 sont jaunes et 8 sont rouges. Il y a en tout  $\binom{10}{5} = 252$  possibilités, dont  $\binom{2}{0} \binom{8}{5} = 56$  parties constituées uniquement de boules rouges;  $\binom{2}{1} \binom{8}{4} = 140$  parties constituées d'une boule jaune (et donc de quatre boules rouges) et  $\binom{2}{2} \binom{8}{3} = 56$  parties constituées de deux boules jaunes (et donc de trois boules rouges). En munissant l'espace probabilisable  $(\Omega, \mathfrak{F}(\Omega))$  de la mesure de probabilité uniforme, on peut supposer que

$$\mathbf{P}(X = 0) = \frac{56}{252} = \frac{2}{9}, \quad \mathbf{P}(X = 1) = \frac{140}{252} = \frac{5}{9}, \quad \mathbf{P}(X = 2) = \frac{2}{9}.$$

• **Modèle B**

Un deuxième modèle va décrire en détail l'expérience aléatoire en s'appuyant sur la formule des probabilités composées. On obtiendra finalement la même loi pour  $X$ , mais après des calculs considérablement plus longs.

On décrit les cinq tirages successifs par cinq variables aléatoires  $(U_k)_{1 \leq k \leq 5}$  définies sur un même espace probabilisé, qui suivent toutes une loi de Bernoulli : comme plus haut, on convient d'interpréter l'évènement  $[U_k = 1]$  comme le fait d'obtenir une boule jaune au  $k$ -ième tirage. D'après la formule des probabilités composées,

$$\begin{aligned} \mathbf{P}(U_1 = \varepsilon_1, U_2 = \varepsilon_2, U_3 = \varepsilon_3, U_4 = \varepsilon_4, U_5 = \varepsilon_5) \\ = \mathbf{P}(U_1 = \varepsilon_1) \mathbf{P}(U_2 = \varepsilon_2 \mid U_1 = \varepsilon_1) \\ \times \mathbf{P}(U_3 = \varepsilon_3 \mid U_1 = \varepsilon_1, U_2 = \varepsilon_2) \\ \times \mathbf{P}(U_4 = \varepsilon_4 \mid U_1 = \varepsilon_1, U_2 = \varepsilon_2, U_3 = \varepsilon_3) \\ \times \mathbf{P}(U_5 = \varepsilon_5 \mid U_1 = \varepsilon_1, U_2 = \varepsilon_2, U_3 = \varepsilon_3, U_4 = \varepsilon_4) \end{aligned}$$

(avec la convention habituelle si l'un des évènements considérés était négligeable).

On convient d'invoquer l'hypothèse d'équiprobabilité pour définir chaque des ces probabilités : à chaque fois, on connaît exactement la composition de l'urne et on convient que la probabilité de tirer une boule jaune est égale à la proportion de boules jaunes dans l'urne.

On en déduit que

$$\begin{aligned} \mathbf{P}(U_1 = 0, U_2 = 0, U_3 = 0, U_4 = 0, U_5 = 0) &= \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{6}{8} \cdot \frac{5}{7} \cdot \frac{4}{6} = \frac{2}{9}, \\ \mathbf{P}(U_1 = 1, U_2 = 0, U_3 = 0, U_4 = 0, U_5 = 0) &= \frac{2}{10} \cdot \frac{8}{9} \cdot \frac{7}{8} \cdot \frac{6}{7} \cdot \frac{5}{6} = \frac{1}{9}, \\ \mathbf{P}(U_1 = 1, U_2 = 1, U_3 = 0, U_4 = 0, U_5 = 0) &= \frac{2}{10} \cdot \frac{1}{9} \cdot \frac{8}{8} \cdot \frac{7}{7} \cdot \frac{6}{6} = \frac{1}{45}. \end{aligned}$$

Il est clair que  $[X = 0] = [U_1 = 0, U_2 = 0, U_3 = 0, U_4 = 0, U_5 = 0]$ , donc  $\mathbf{P}(X = 0) = 2/9$ .

L'évènement  $[X = 1]$  est l'union de  $\binom{5}{1} = 5$  évènements analogues à l'évènement

$$[U_1 = 1, U_2 = 0, U_3 = 0, U_4 = 0, U_5 = 0]$$

et deux à deux disjoints. En appliquant les mêmes règles de calcul, on constate que ces cinq évènements ont la même probabilité. Par additivité de  $\mathbf{P}$ , on en déduit que

$$\mathbf{P}(X = 1) = 5 \cdot \frac{1}{9} = \frac{5}{9}.$$

De même, l'évènement  $[X = 2]$  est l'union de  $\binom{5}{2} = 10$  évènements deux à deux disjoints, analogues à l'évènement

$$[U_1 = 1, U_2 = 1, U_3 = 0, U_4 = 0, U_5 = 0].$$

Là encore, les calculs montrent que ces vingt évènements ont tous la même probabilité et donc que

$$P(X = 2) = \frac{2}{9}.$$

☞ On constate que  $P(X = 0) + P(X = 1) + P(X = 2) = 1$ , ce qui prouve que notre modèle a un sens et cela suffit pour la suite.

On constate par ailleurs que le modèle ainsi obtenu est le même que le précédent. Cela ne signifie pas que ce modèle soit le seul possible ou le seul acceptable, mais qu'on a présenté de deux manières différentes une même description d'une expérience aléatoire!

### • Modèle C

Un troisième modèle consiste à supposer que la variable aléatoire  $X$  suit une loi uniforme :

$$P(X = 0) = P(X = 1) = P(X = 2) = \frac{1}{3}.$$

On obtient cette fois un modèle différent des modèles précédents.

☞ *Aucun argument mathématique ne permet de préférer les modèles précédents au modèle C! Absolument aucun!*

*La seule certitude "mathématique" dans ce second cadre d'étude est que  $X$  est une variable aléatoire à valeurs dans  $E = \{0; 1; 2\}$ : on a donc une contrainte objective sur la loi de  $X$  (le support de cette loi est imposée), mais la loi de  $X$  n'est pas déterminée par l'énoncé.*

• Restons dans un cadre *mainstream* et supposons dans ce qui suit que

$$P(X = 0) = \frac{2}{9}, \quad P(X = 1) = \frac{5}{9}, \quad P(X = 2) = \frac{2}{9}.$$

Alors

$$E(X) = 0 \cdot \frac{2}{9} + 1 \cdot \frac{5}{9} + 2 \cdot \frac{2}{9} = 1,$$

$$E(X^2) = 0^2 \cdot \frac{2}{9} + 1^2 \cdot \frac{5}{9} + 2^2 \cdot \frac{2}{9} = \frac{13}{9},$$

$$V(X) = E(X^2) - E(X)^2 = \frac{4}{9}.$$

☞ *Avec ou sans remise, l'espérance de  $X$  est la même, mais la variance diminue : c'est normal, il y a moins de valeurs possibles, donc les valeurs de  $X$  sont plus concentrées autour de la valeur moyenne.*

Comme plus haut,  $Y = -15 + 5X$  et

$$E(Y) = -10, \quad V(Y) = \frac{100}{9}.$$