Vecteurs aléatoires [56.=CCP 95]

Modélisation

On modélise les tirages successifs par cinq variables aléatoires de Bernoulli : $(B_k)_{1\leqslant k\leqslant 5}$ définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. Ces variables aléatoires sont caractérisées par leur **loi conjointe** ou loi du vecteur aléatoire

$$B = (B_1, B_2, B_3, B_4, B_5),$$

c'est-à-dire par la famille des $2^5 = 32$ réels positifs

$$\mathbf{P}(B_1 = \varepsilon_1, B_2 = \varepsilon_2, B_3 = \varepsilon_3, B_4 = \varepsilon_4, B_5 = \varepsilon_5)$$

lorsque la liste $(\varepsilon_k)_{1\leqslant k\leqslant 5}$ parcourt l'ensemble $\{0;1\}^5$.

On convient d'interpréter ces variables aléatoires de la manière suivante : l'événement $[X_k = 1]$ signifie qu'on a tiré une boule jaune au k-ième tirage et l'événement contraire $[X_k = 0]$ signifie qu'on a tiré une boule rouge au k-ième tirage.

❧ D'après la formule des probabilités composées, la probabilité

$$\mathbf{P}(B_1 = \varepsilon_1, B_2 = \varepsilon_2, B_3 = \varepsilon_3, B_4 = \varepsilon_4, B_5 = \varepsilon_5)$$

est égale au produit

$$\begin{aligned}&\mathbf{P}(B_1 = \varepsilon_1) \times \mathbf{P}(B_2 = \varepsilon_2 \mid B_1 = \varepsilon_1)\\ &\quad\times \mathbf{P}(B_3 = \varepsilon_3 \mid B_1 = \varepsilon_1, B_2 = \varepsilon_2)\\ &\quad\times \mathbf{P}(B_4 = \varepsilon_4 \mid B_1 = \varepsilon_1, B_2 = \varepsilon_2, B_3 = \varepsilon_3)\\ &\quad\times \mathbf{P}(B_5 = \varepsilon_5 \mid B_1 = \varepsilon_1, B_2 = \varepsilon_2, B_3 = \varepsilon_3, B_4 = \varepsilon_4)\end{aligned}$$

et il reste à donner une valeur à ces probabilités.

❧ Pour fixer la valeur de chaque probabilité conditionnelle, nous allons recourir à la règle intuitive habituelle : l'hypothèse d'équiprobabilité. Ici, cela signifie que la probabilité de tirer une boule jaune est, en toutes circonstances, égale à la proportion de boules jaunes dans l'urne au moment du tirage.

❧ Il faut remarquer que la *somme*

$$s = \varepsilon_1 + \cdots + \varepsilon_5$$

est aussi le *nombre* d'indices k tels que $\varepsilon_k = 1$ (puisqu'on somme des termes qui sont égaux à 0 ou à 1). Cela signifie que le nombre X de boules jaunes est bien une variable aléatoire discrète :

$$X = B_1 + B_2 + B_3 + B_4 + B_5$$

et que, pour tout $0 \leqslant s \leqslant 5$,

$$\mathbf{P}(X = s) = \sum_{\substack{(\varepsilon_1,\dots,\varepsilon_5)\in\{0;1\}^5\\ \varepsilon_1+\cdots+\varepsilon_5=s}} \mathbf{P}\big(B = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5)\big)$$

Nous allons maintenant discuter sur la valeur de s.

❧ *Premier cas*

Si la somme s est nulle, alors tous les ε_k sont nuls : on a tiré cinq boules rouges. Compte-tenu de la composition de l'urne au cours des tirages,

$$\mathbf{P}(B_1 = 0) = \frac{8}{10}, \quad \mathbf{P}(B_2 = 0 \mid B_1 = 0) = \frac{7}{9},$$
$$\mathbf{P}(B_3 = 0 \mid B_1 = 0, B_2 = 0) = \frac{6}{8},$$
$$\mathbf{P}(B_4 = 0 \mid B_1 = B_2 = B_3 = 0) = \frac{5}{7},$$
$$\mathbf{P}(B_5 = 0 \mid B_1 = B_2 = B_3 = B_4 = 0) = \frac{4}{6}$$

et par conséquent, il paraît raisonnable de poser :

$$\begin{aligned}\mathbf{P}(B_1 = B_2 = B_3 = B_4 = B_5 = 0) &= \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{6}{8} \cdot \frac{5}{7} \cdot \frac{4}{6} \\ &= \frac{2}{9}\end{aligned}$$

ce qui nous donne $\mathbf{P}(X = 0) = 2/9$.

❧ *Deuxième cas*

Si la somme s est égale à 1, alors tous les ε_k sont nuls sauf un. En discutant sur l'indice de la variable B_k qui prend la valeur 1, on constate que ce cas concerne $\binom{5}{1} = 5$ événements. En raisonnant comme plus haut,

$$\mathbf{P}(B_1 = 0) = \frac{8}{10}, \quad \mathbf{P}(B_2 = 0 \mid B_1 = 0) = \frac{7}{9},$$
$$\mathbf{P}(B_3 = 0 \mid B_1 = 0, B_2 = 0) = \frac{6}{8},$$
$$\mathbf{P}(B_4 = 0 \mid B_1 = B_2 = B_3 = 0) = \frac{5}{7},$$
$$\mathbf{P}(B_5 = 1 \mid B_1 = B_2 = B_3 = B_4 = 0) = \frac{2}{6}$$

et cette fois, il paraît raisonnable de poser :

$$\begin{aligned}&\mathbf{P}(B_1 = B_2 = B_3 = B_4 = 0, B_5 = 1) \\ &\qquad = \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{6}{8} \cdot \frac{5}{7} \cdot \frac{2}{6} = \frac{1}{9}.\end{aligned}$$

Avec quelques calculs au brouillon, on finit par deviner que les cinq événements étudiés ici sont équiprobables (les dénominateurs sont les mêmes dans le même ordre, les numérateurs sont les mêmes dans un ordre différent, donc tous les quotients sont égaux : c'est un *constat* que l'on fait sur les calculs, ce n'est pas une *hypothèse*). Par conséquent, $\mathbf{P}(X = 1) = 5/9$.

❧ *Troisième cas*

Si la somme s est égale à 2, alors tous les ε_k sont nuls sauf deux. En discutant sur les indices des variables B_k qui prennent la valeur 1, ce cas concerne

$\binom{5}{2} = 10$ événements. En raisonnant comme plus haut,

$$\mathbf{P}(B_1 = 0) = \frac{8}{10}, \quad \mathbf{P}(B_2 = 0 \mid B_1 = 0) = \frac{7}{9},$$
$$\mathbf{P}(B_3 = 0 \mid B_1 = 0, B_2 = 0) = \frac{6}{8},$$
$$\mathbf{P}(B_4 = 1 \mid B_1 = B_2 = B_3 = 0) = \frac{2}{7},$$
$$\mathbf{P}(B_5 = 1 \mid B_1 = B_2 = B_3 = 0, B_4 = 1) = \frac{1}{6}$$

et cette fois, il paraît raisonnable de poser :

$$\mathbf{P}(B_1 = B_2 = B_3 = 0, B_4 = B_5 = 1) = \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{6}{8} \cdot \frac{2}{7} \cdot \frac{1}{6} = \frac{2}{90}.$$

Comme dans le cas précédent, on peut s'assurer que les dix événements considérés sont équiprobables et par suite, $\mathbf{P}(X = 2) = 2/9$.

❧ *Dernier cas*

Si la somme s était supérieure à 3, alors le nombre de boules jaunes tirées serait plus grand que le nombre de boules jaunes contenues dans l'urne ! Ces $32 - (1 + 5 + 10) = 16$ événements sont donc impossibles et leur probabilité est donc nulle.

Cohérence du modèle

Le modèle que nous venons de définir est le résultat d'hypothèses successives, chacune d'elles étant raisonnables (et les calculs intermédiaires ont suivi les règles de calcul établies en cours). Il reste donc à vérifier que le modèle qui en résulte est cohérent, c'est-à-dire que la famille

$$\Big(\mathbf{P}\big[B = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5)\big]\Big)_{(\varepsilon_k)_{1 \leqslant k \leqslant 5} \in \{0;1\}^5}$$

est bien une famille sommable de réels positifs dont la somme est égale à 1 (caractérisation des lois de probabilités discrètes).

Il est clair que la famille est sommable (nombre *fini* de termes) et que ces réels sont positifs (!). Quant à la somme de ces nombres, en suivant la discussion précédente, elle est égale à

$$\frac{2}{9} + \frac{5}{9} + \frac{2}{9} + 0 = 1.$$

Nous venons de prouver que notre modèle probabiliste a bien un sens.

Loi de X

Selon notre modèle, le nombre X de boules jaunes tirées est une variable aléatoire discrète (en tant que somme de cinq variables aléatoires discrètes), qui prend les valeurs 0, 1 ou 2, avec les fréquences respectives :

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

Autre modélisation

Bien entendu, le modèle proposé n'est pas le seul modèle acceptable et rien n'assure que tous les modèles acceptables aboutiront à la même loi pour la variable aléatoire X...

❧ Par exemple, nous aurions pu modéliser un tirage de cinq boules successives sans remise comme le choix d'une partie de cinq éléments dans un ensemble de dix éléments. Cela suppose que les dix boules soient discernables (par exemple numérotées en plus d'être colorées en jaune ou en rouge) et occulte le fait que les boules soient tirées les unes après les autres (on s'intéresse au résultat global des cinq tirages, pas aux résultats successifs).

L'ensemble E_5^{10} des parties de cinq éléments d'un ensemble de dix éléments est un ensemble fini, dont le cardinal est égal à $\binom{10}{5}$. Faute d'information plus précise, on peut se résoudre à munir cet ensemble de la mesure de probabilité uniforme.

❧ Dans ces conditions, il y a :
- $\binom{8}{5} \times \binom{2}{0}$ parties de cinq éléments rouges ;
- $\binom{8}{4} \times \binom{2}{1}$ parties de quatre éléments rouges et un élément jaune ;
- $\binom{8}{3} \times \binom{2}{2}$ parties de trois éléments rouges et deux éléments jaunes.

On retrouve alors que

$$\frac{\binom{8}{5} \times \binom{2}{0}}{\binom{10}{5}} = \frac{2}{9}, \quad \frac{\binom{8}{4} \times \binom{2}{1}}{\binom{10}{5}} = \frac{5}{9}, \quad \frac{\binom{8}{3} \times \binom{2}{2}}{\binom{10}{5}} = \frac{2}{9}.$$

Comme on a une famille de trois réels positifs dont la somme est égale à 1, on peut juger raisonnable de modéliser le nombre boules jaunes tirées par une variable aléatoire discrète

$$X : \Omega \to \{0; 1; 2\}$$

dont la loi est caractérisée par

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

❧ Avec deux modélisations sensiblement différentes, nous sommes arrivés à la même modélisation pour X : cela tient littéralement du miracle. Aucun argument mathématique ne permet de préférer un modèle à un autre et, normalement, la loi de X qu'on cherche ici à établir *dépend* du modèle choisi.

Loi de Y

Quelle que soit la loi de X (calculée comme pour le premier modèle ou choisie comme pour le second modèle), on peut facilement exprimer Y en fonction de X :

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

(Puisque X compte le nombre de boules jaunes, le nombre de boules rouges tirées est égal à $(5 - X)$ et on connaît le nombre de points attribués à chaque tirage.)

❧ Autrement dit, $Y = f(X)$ pour une certaine fonction affine f, ce qui prouve que Y est bien une variable aléatoire discrète. Comme la fonction f est bijective, on en déduit que

$$\begin{aligned}
[Y = -15] &= [X = 0], \\
[Y = -10] &= [X = 1], \\
[Y = -5] &= [X = 2].
\end{aligned}$$

❧ Par linéarité de l'espérance,

$$\mathbf{E}(Y) = 5\,\mathbf{E}(X) - 15$$

quelle que soit la loi de X (c'est-à-dire quel que soit le modèle retenu).