Petit essai sur les tirages dans une urne

Yves-Noël Haubry(*)

En Premières et Terminales STI et S, je donne toujours des exercices de probabilités concernant les tirages de boules dans des urnes contenant un mélange bicolore du genre :

Dans une urne on place 10 boules blanches et 5 noires. On effectue 2 (ou 3) tirages successifs d'une boule avec remise. On suppose l'équiprobabilité de sortie des boules. On ne tient pas compte de l'ordre de sortie des couleurs. À l'aide d'un arbre, déterminer les événements et calculer leurs probabilités.

Ensuite on joue au jeu suivant : chaque blanche sortie donne un gain de 10 € et chaque noire donne une perte de 20 €. Déterminer les gains et leurs probabilités et calculer le gain moyen.

Puis je fais reprendre l'exercice avec cette fois des tirages sans remise, le contenu initial de l'urne et les valeurs attribuées aux couleurs étant identiques.

Or, j'ai constaté qu'à chaque fois on obtenait le même gain moyen (ou espérance mathématique de gain) que les tirages soient faits avec ou sans remise alors que je pensais avoir des résultats différents. Je me suis posé la question de savoir si cette particularité était générale, c'est à dire :

– quels que soient le nombre de tirages avec ou sans remise, le nombre de blanches et de noires et les valeurs qui leur sont attribuées, a-t-on toujours la même espérance mathématique de gain ?

Après quelques vérifications pour 3 et 4 (et plus) tirages à l'aide d'EXCEL où j'ai toujours constaté l'égalité, je me suis attaqué au problème sur le plan général et je l'ai résolu pour des tirages bicolores (j'enfonce peut-être des portes ouvertes, mais je n'ai jamais vu cette propriété évoquée dans les manuels de Terminale). La démonstration de cette propriété n'est évidemment pas au programme des classes terminales STI ou S, mais on peut amener les élèves à conjecturer ce résultat par des expériences successives.

L'énoncé du problème étant donné, je fournis aux élèves des tableaux photocopiés (j'ai remarqué qu'en général les élèves mettent beaucoup de temps à tracer un tableau sur leur cahier et on a intérêt à fournir des photocopies de tableaux tout préparés, le gain de temps est appréciable).

(*) Lycée « Les Lombards », Troyes.

Nb B	10	Nb N	5	Tot b	15	P(1B)	
		Gain B	10	Gain N	-20		
		3 tirages avec remises			3 tirages sans remises		
Évént	nb B	proba	gain	P × G	proba	gain	P × G
	3						
	2						
	1						
	0						
somme							

Ils le remplissent en utilisant un arbre pour déterminer les événements et calculer leurs probabilités, puis les espérances mathématiques dont on constate l'égalité.

Dans un deuxième temps on reprend l'exercice en ne modifiant que les valeurs attribuées aux couleurs et on refait calculer l'espérance mathématique dans les deux cas de tirages. Ceci peut être fait assez rapidement en fournissant aux élèves des tableaux de calculs préparés où ils n'ont plus qu'à calculer les gains, les $g_i \times p_i$ et leurs sommes, tableaux créés par un tableur comme celui-ci :

Nb B	10	Nb N	5	Tot b	15	P(1B)	
		Gain B		Gain N			
		3 tirages avec remises			3 tirages sans remises		
Évént	nb B	proba	gain	P × G	proba	gain	P × G
3B	3	0.296			0.26374		
2B 1N	2	0.444			0.49451		
1B 2N	1	0.222			0.21978		
3N	0	0.037			0.02198		
somme		1			1		

Dans un troisième temps on modifie la composition de l'urne et les valeurs des gains et, toujours avec l'aide de tableaux préparés, on refait constater que le mode de tirage ne modifie pas l'espérance mathématique (on peut donner les tableaux à faire en exercices à la maison et contrôler les résultats au cours suivant). Il n'est pas question de faire en classe la démonstration de cette propriété. J'indique aux élèves que cette démonstration existe et que la propriété est vérifiée.

On peut aussi poser la question des valeurs à donner aux gains pour que le jeu soit équitable, c'est à dire à espérance nulle, pour une composition donnée de l'urne.

Revenons à notre préoccupation : l'espérance mathématique ne dépend pas du mode de tirage.

PROBLÈME

Dans une urne on place n boules blanches et k noires. On effectue t tirages successifs d'une boule soit avec remise, soit sans remise. On suppose l'équiprobabilité de sortie des boules et que n et k sont supérieurs à t. On ne tient pas compte de l'ordre de sortie des couleurs.

Chaque blanche donne un gain x et chaque noire donne un gain y.

Montrons que l'espérance mathématique du gain est la même dans les deux cas de tirage et que l'on a :

$$\mathrm{EG} = t(px + (1-p)y)$$

où $p = \dfrac{n}{n+k}$ est la proportion de blanches dans l'urne.

1°) Les tirages avec remise

Le nombre X de boules blanches sorties sur les t tirages suit une loi binomiale de paramètres t, le nombre de tirages, et $p = \dfrac{n}{n+k}$, la proportion de blanches dans l'urne. La proportion de noires est donc $1 - p = \dfrac{k}{n+k}$.

On a donc $\mathrm{P}(\mathrm{X} = i) = \mathrm{C}_t^i\, p^i (1-p)^{t-i}$ avec i variant de 0 à t.

Pour i blanches sorties, il y a $(t - i)$ noires et le gain d'un tel tirage est donc $\mathrm{G} = ix + (t - i)y$.

L'espérance mathématique de gain est alors :

$$\mathrm{EG} = \sum_{i=0}^{t} (ix + (t-i)y)\mathrm{C}_t^i\, p^i (1-p)^{t-i},$$

$$\mathrm{EG} = x\sum_{i=0}^{t} i\,\mathrm{C}_t^i\, p^i (1-p)^{t-i} + y\sum_{i=0}^{t} (t-i)\mathrm{C}_t^i\, p^i (1-p)^{t-i}.$$

En tenant compte que pour $i = 0$ le premier terme de la somme du coefficient de x est nul et que pour $i = t$ c'est le dernier de la somme du coefficient de y qui est nul, on peut réécrire :

$$\mathrm{EG} = x\sum_{i=1}^{t} i\,\mathrm{C}_t^i\, p^i (1-p)^{t-i} + y\sum_{i=0}^{t-1} (t-i)\mathrm{C}_t^i\, p^i (1-p)^{t-i}.$$

En remarquant que $i\,\mathrm{C}_t^i = t\,\mathrm{C}_{t-1}^{i-1}$ et $(t-i)\mathrm{C}_t^i = t\,\mathrm{C}_{t-1}^i$ (remarque 1), on a alors :

$$\mathrm{EG} = x\sum_{i=1}^{t} t\,\mathrm{C}_{t-1}^{i-1}\, p^i (1-p)^{t-i} + y\sum_{i=0}^{t-1} t\,\mathrm{C}_{t-1}^{i}\, p^i (1-p)^{t-i}.$$

On peut alors factoriser tp dans la première somme et $t(1-p)$ dans la seconde :

$$\mathrm{EG} = tpx\sum_{i=1}^{t} \mathrm{C}_{t-1}^{i-1}\, p^{i-1}(1-p)^{t-i} + t(1-p)y\sum_{i=0}^{t-1} \mathrm{C}_{t-1}^{i}\, p^{i}(1-p)^{(t-1)-i}.$$

On effectue le changement d'indice suivant : $j = i - 1$ dans la somme facteur de x. Quand i varie de 1 à t, j varie de 0 à $t - 1$. Cette somme se récrit alors :

$$\sum_{j=0}^{t-1} \mathrm{C}_{t-1}^{j}\, pj(1-p)^{(t-1)-j}.$$

On reconnaît que les deux sommes sont le développement de $((1-p)+p)^{t-1}$ selon la formule du binôme de Newton et comme $(1-p)+p = 1$ ces sommes valent donc 1 (ou la somme des probabilités pour la loi binomiale de paramètres $(t-1)$ et p). On a alors :

$$\mathrm{EG} = t(px + (1-p)y),$$

ce qui est bien le résultat escompté.

2°) Tirages sans remise

Pour t tirages successifs d'une boule sans remise (on ne retient que le nombre de blanches et de noires obtenues sans tenir compte de l'ordre des sorties), appelons i le nombre de blanches tirées parmi les n et $(t-i)$ celui de noires parmi les k. Il y a $(n+k)$ boules au total. On peut aussi considérer que l'on effectue un tirage de t boules simultanément parmi les $(n+k)$ boules, ce tirage donnant i boules blanches et $(t-i)$ noires. La probabilité d'un tel tirage est :

$$\mathrm{P}(\mathrm{X} = i) = \mathrm{P}(\mathrm{Y} = k - i) = \frac{\mathrm{C}_n^i\, \mathrm{C}_k^{t-i}}{\mathrm{C}_{n+k}^t}$$

et le gain pour i blanches et $(t-i)$ noires est, comme au 1º, $\mathrm{G} = ix + (t-i)y$.

L'espérance mathématique du gain est donnée par :

$$\mathrm{EG} = \sum_{i=0}^{t} (ix + (t-i)y)\frac{\mathrm{C}_n^i\, \mathrm{C}_k^{t-i}}{\mathrm{C}_{n+k}^t}.$$

Avec la remarque 1 du 1º, nous pouvons écrire :

$$\mathrm{EG} = \frac{1}{\mathrm{C}_{n+k}^t}\left(nx\sum_{i=1}^{t} \mathrm{C}_{n-1}^{i-1}\, \mathrm{C}_k^{t-i} + ky\sum_{i=0}^{t-1} \mathrm{C}_n^i\, \mathrm{C}_{k-1}^{t-i-1} \right).$$

Considérant que $\dfrac{1}{\mathrm{C}_{n+k}^t} = \dfrac{t}{n+k}\dfrac{1}{\mathrm{C}_{n+k-1}^{t-1}}$ et $p = \dfrac{n}{n+k}$ et $1 - p = \dfrac{k}{n+k}$, on obtient alors :

$$\mathrm{EG} = \frac{t}{C_{n+k-1}^{t-1}}\left(px\sum_{i=1}^{t} C_{n-1}^{i-1}\,C_k^{t-i} + (1-p)y\sum_{i=0}^{t-1} C_n^{i}\,C_{k-1}^{t-i-1}\right).$$

Posons : $A = \frac{1}{C_{n+k-1}^{t-1}}\sum_{i=1}^{t} C_{n-1}^{i-1}\,C_k^{t-i}$ et $B = \frac{1}{C_{n+k-1}^{t-1}}\sum_{i=0}^{t-1} C_n^{i}\,C_{k-1}^{t-i-1}$. On effectue le changement d'indice suivant : $j = i - 1$ dans la somme A. Quand i varie de 1 à t, j varie de 0 à $t - 1$. La somme se récrit alors :

$$A = \frac{1}{C_{(n-1)+k}^{t-1}}\sum_{j=0}^{t-1} C_{n-1}^{j}\,C_k^{(t-1)-j}.$$

Le produit $a = C_{n-1}^{j}\,C_k^{(t-1)-j}$ représente le nombre de façons de choisir j boules blanches parmi $(n - 1)$ et $(t - 1) - j$ boules noires parmi k. Le terme $b = C_{(n-1)+k}^{i-1}$ représente le nombre de façons de choisir $(t - 1)$ boules parmi $(n - 1) + k$. Le quotient a/b des deux termes est donc la probabilité de sortir j boules blanches et $(t - 1) - j$ noires en effectuant $(t - 1)$ tirages successifs d'une boule sans remises (ou dans un tirage de $(t - 1)$ boules simultanément) dans un ensemble de $(n - 1) + k$ boules.

La somme A est donc la somme des probabilités de tous les tirages pour j variant de 0 à $(t - 1)$. Cette somme vaut donc 1.

De même, la somme B s'interprète comme la somme des probabilités de sortir i boules blanches parmi n et $(t - 1) - i$ noires parmi $(k - 1)$ dans un ensemble de $(n - 1) + k$ boules pour i variant de 0 à $(t - 1)$. Elle vaut également 1.

Nous obtenons encore :

$$\mathrm{EG} = t(px + (1-p)y)$$

ou en reprenant le nombre n de blanches et k de noires :

$$\mathrm{EG} = t\left(\frac{nx + ky}{n + k}\right).$$

EG est donc la moyenne des gains de l'ensemble des boules multipliée par le nombre de tirages.

L'espérance mathématique du gain ne dépend donc pas du mode de tirage employé.

Jeu équitable.

Pour que le jeu soit équitable il faut que EG soit nulle. Ce qui nous donne :

$$px + (1 - p)y = 0,$$

soit :

$$\frac{x}{y} = -\frac{1-p}{p} = -\frac{k}{n} = -\frac{1}{\frac{n}{k}}.$$

Il faut donc choisir les gains x et y inversement proportionnels au rapport – blanches/noires et cela ne dépend pas du nombre de tirages.

Cette démonstration est assez lourde du fait de l'utilisation des coefficients du binôme.

Paul-Louis Hennequin (que je remercie pour ses conseils) m'en a indiqué une autre, plus simple, que voici :

Au départ l'urne contient n boules blanches et k noires. On peut effectuer dans cette urne :
– soit t tirages avec remise (t entier quelconque),
– soir t tirages sans remise ($t \leq n + k$).
On appelle A_i l'événement « au i-ème tirage la boule est blanche ».
Pour tout i compris entre 1 et t on a :

$$P(A_i) = \frac{n}{n+k}$$

dans les deux cas.
a) Si le tirage est avec remise, tous les tirages sont faits dans une urne où il y a toujours la même composition de n blanches et k noires et $P(A_i)$ ne dépend pas de i et vaut $\frac{n}{n+k}$.

b) Si les tirages sont faits sans remise, on a $P(A_1) = \frac{n}{n+k}$.

Au lieu d'effectuer t tirages sans remise, on peut considérer que l'on fait un tirage de t boules simultanément et qu'on les numérote 1, 2, …, t. Un tirage de t boules numérotées est un arrangement de t boules parmi $(n + k)$. Il y a alors $(n + k)(n + k - 1) \ldots (n + k - t + 1)$ tirages possibles (nombre d'arrangements) et il y a $n\,(n + k - 1) \ldots (n + k - t + 1)$ tirages où la i-ème boule est blanche (n choix de la i-ème boule chacun étant associé à un arrangement de $t - 1$ boules parmi les $(n + k - 1)$ restantes.

La probabilité de A_i est bien $\frac{n}{n+k}$ et ne dépend pas du rang i.

Donc dans les deux cas on a toujours $P(A_i) = \frac{n}{n+k}$.

Si on note X_t le nombre de blanches obtenues en t tirages, ce nombre peut s'écrire :

$$X_t = \sum_{i=1}^{t} 1_{A_t}$$

avec

$$1_{A_t} = \begin{cases} 1 & \text{si } A_t \text{ est réalisé} \\ 0 & \text{sinon} \end{cases}$$

et, par linéarité de l'espérance mathématique :

$$E(X_t) = \sum_{i=1}^{t} E(1_{A_t}) = \sum_{i=1}^{t} P(A_i) = \frac{tn}{n+k} = tp$$

avec

$$p = \frac{n}{n+k}.$$

Si chaque blanche rapporte x, l'espérance sera alors tpx.

De même la probabilité de sortie d'une noire étant $1-p$ et son gain y, l'espérance des gains des noires sera $t(1-p)y$ et, par linéarité, l'espérance totale est bien :

$$EG = t(px + (1-p)y),$$

Cette propriété peut se généraliser à un nombre quelconque de couleurs et de tirages. Pour une couleur c_k donnée parmi les couleurs c_1, c_1, …, c_n on établit que la probabilité que la i-ème boule tirée soit de cette couleur est égale à

$$\frac{b_k}{b_1 + b_2 + \ldots + b_n}$$

où b_k est le nombre de boules de la couleur c_k en considérant une partition de l'ensemble des boules en deux sous-ensembles : celles de la couleur c_k d'une part et toutes les autres d'autre part, on est ainsi ramené au problème de deux couleurs. L'espérance se calcule par linéarité et ne dépend pas du mode de tirage.

Pour conclure, j'ai eu l'idée de cet essai pour donner une base de travaux pratiques de probabilités, abordables avec des moyens élémentaires : un arbre des choix. On peut arriver à amener les élèves à une conclusion en deux séances de TD d'une heure, en préparant les tableaux de calculs et en se limitant à deux et trois tirages. J'ai tenté de le faire avec trois couleurs, mais les calculs deviennent longs pour plus de deux tirages.