48. Combat de dés

Probabilités et statistique : loi uniforme discrète ; espérance conditionnelle ; loi géométrique ; formule des probabilités totales ; *loi uniforme continue*
Analyse : comparaison série-intégrale ; *série harmonique*
Difficulté : ⋆ ⋆

Énoncé. Soit $n \geqslant 1$ un entier. Deux joueurs utilisent un dé à n faces (numérotées $1, 2, \ldots, n$) de la manière suivante :

- le joueur A lance le dé, obtient le résultat X, et verse 3 euros au joueur B ;
- le joueur B lance alors le dé jusqu'à ce qu'il obtienne une valeur supérieure ou égale à X (la partie s'arrête alors) ; à chaque lancer, il donne 1 euro au joueur A.

On note Y la somme versée au cours de la partie par le joueur B au joueur A.

1. Pour $k \in \{1, \ldots, n\}$ et $j \in \mathbb{N}^\star$, que vaut $\mathbb{P}(Y = j \mid X = k)$?
2. Montrer qu'une variable aléatoire Z à valeurs dans $\mathbb{N}^\star$ admet une espérance si et seulement si la série de terme général $\mathbb{P}(Z \geqslant j)$ converge, et que dans ce cas
$$\mathbb{E}[Z] = \sum_{j=1}^{+\infty} \mathbb{P}(Z \geqslant j)\,.$$
3. En déduire un encadrement simple de $\mathbb{E}[Y]$ en fonction de n, puis un équivalent lorsque $n \to +\infty$.
4. Selon la valeur de n, à qui le jeu est-il favorable en moyenne ? On pourra utiliser les valeurs approchées $\mathrm{e}^2 \approx 7{,}39$ et $\mathrm{e}^3 \approx 20{,}09$.

Conseils. L'énoncé ne précise pas explicitement que les résultats des lancers de dés suivent des lois uniformes sur $\{1, \ldots, n\}$ et ne souligne pas non plus les indépendances entre différents lancers. Qu'à cela ne tienne, nous ferons preuve d'initiative et supposerons ces faits vérifiés : le jury appréciera que nous pointions ces hypothèses naturelles mais implicites. La question 1 revient alors à identifier une situation connue, vue en cours : un temps d'attente dans une suite d'épreuves de Bernoulli.

La formulation de la question 2 suggère qu'il s'agit de retrouver un résultat général (classique et peut-être vu durant l'année) ; en cas de difficulté ici, on peut commencer par réfléchir au cas où Z est bornée, qui est traité à la question 2 de l'exercice 44. Et en dernier ressort, on peut sauter sans souci la question 2 lors de la préparation. C'est en un sens davantage une indication pour la suite. Une démonstration possible repose sur une interversion de sommations. On rappelle qu'une telle interversion est légitime dans deux cas : lorsque d'une part, tous les termes sont positifs (même si ce résultat n'est pas toujours formulé comme étant explicitement au programme) ; et d'autre part, en cas de convergence absolue (en cas de convergence de la somme des valeurs absolues des termes).

La première partie de la question 3 doit découler assez rapidement des questions 1 et 2, la seconde partie est classique et a sans nul doute été vue durant l'année (elle a en tout cas été rencontrée dans un exercice précédent, mais pour maintenir le suspense, nous ne révélons pas encore lequel !).

Enfin, la question 4 a été jugée difficile : elle nécessite de bien comprendre le déroulé du jeu. Il s'agit de comparer des espérances de gains (ce sont les mots « en moyenne » qui indiquent cela). Celle de B se lit dans l'énoncé, celle de A est d'abord à calculer, et il s'agit ensuite de déterminer, en fonction de n, qui est la plus grande des deux. S'il n'est pas possible de traiter *in extenso* cette question lors du temps de préparation, il est bon de présenter au jury au moins les grandes lignes d'un raisonnement ; il sera toujours temps de se pencher sur les détails du calcul si le temps de passage à l'oral le permet.

Corrigé. Commençons par préciser et formaliser davantage l'énoncé. Il est naturel de modéliser les lancers du joueur B comme les réalisations de variables aléatoires T_1, T_2, … indépendantes entre elles et de X, toutes (y compris X) de loi uniforme sur l'ensemble $\{1, \ldots, n\}$. La somme Y versée par B à A est égale au nombre de lancers nécessaires pour égaler ou dépasser le score X : formellement,

$$Y = \min\Big\{j \geqslant 1 : \quad T_j \geqslant X\Big\} ,$$

où la résolution de l'exercice montrera bien qu'avec probabilité 1, un entier $j \geqslant 1$ tel que $T_j \geqslant X$ existe. La variable aléatoire Y est donc ce que l'on appelle un temps d'attente. La valeur de Y dépend évidemment de X (les deux variables aléatoires ne sont pas indépendantes) : plus le score X est petit, plus la somme Y aura tendance à être petite.

1. Pour $k = 1$, un seul lancer suffit pour égaler ou dépasser k :

$$\mathbb{P}(Y = 1 \,|\, X = 1) = 1 .$$

Pour les valeurs $k \in \{2, \ldots, n\}$, on raisonne conditionnellement à l'événement $\{X = k\}$: par indépendance des T_j par rapport à X, la loi de Y sachant $X = k$ est la loi de

$$Y_k = \min\Big\{j \geqslant 1 : \quad T_j \geqslant k\Big\} .$$

Les événements $\{T_j \geqslant k\}$ étant indépendants et de même probabilité, la loi de Y_k (et donc celle de Y sachant $X = k$) est une loi géométrique sur $\mathbb{N}^\star$, de paramètre noté $p_k = \mathbb{P}(T_j \geqslant k)$. Les valeurs prises par T_j étant équiprobables, on calcule p_k comme le rapport entre le nombre de cas favorables (les entiers de k à n : il y en a $n-k+1$) sur le nombre total de cas (les entiers de 1 à n : il y en a n). Ainsi,

$$p_k = \mathbb{P}(T_j \geqslant k) = \frac{n-k+1}{n} = 1 - \frac{k-1}{n}\,.$$

Par définition de la loi géométrique, on sait alors que pour tout entier $j \geqslant 1$,

$$\mathbb{P}(Y = j \,|\, X = k) = \mathbb{P}(Y_k = j) = p_k(1-p_k)^{j-1}\,.$$

Par ailleurs, calculons tout de suite la probabilité pour Y de dépasser j sachant $X = k$; cette quantité est très souvent utile lorsqu'il est question de loi géométrique, et l'énoncé de la question 2 signale que l'on va certainement s'en servir dans la suite. Pour tous $k \geqslant 2$ et $j \geqslant 1$,

$$\mathbb{P}(Y \geqslant j \,|\, X = k) = \mathbb{P}(Y_k \geqslant j) = \mathbb{P}\left(\bigcap_{\tau=1}^{j-1} \left\{T_\tau \leqslant k-1\right\}\right) = (1-p_k)^{j-1}\,, \tag{48.1}$$

où l'on a utilisé $\mathbb{P}(T_\tau \leqslant k-1) = 1 - p_k$ et l'indépendance des T_τ.

2. Lorsque Z admet une espérance, la formule de transfert pour les variables discrètes à valeurs dans $\mathbb{N}^\star$ assure que

$$\mathbb{E}[Z] = \sum_{s=1}^{+\infty} s\,\mathbb{P}(Z = s)\,.$$

On peut artificiellement faire apparaître une seconde somme :

$$\mathbb{E}[Z] = \sum_{s=1}^{+\infty} \sum_{j=1}^{s} \mathbb{P}(Z = s)\,.$$

On voudrait alors échanger les sommations : on a le droit de le faire et on peut le justifier soit en notant que tous les termes de la somme double sont positifs, soit en notant que la condition d'intégrabilité de Z revient précisément ici (tous les termes de la somme étant positifs) à la convergence absolue de la série double. Ainsi,

$$\mathbb{E}[Z] = \sum_{j=1}^{+\infty} \underbrace{\sum_{s=j}^{+\infty} \mathbb{P}(Z = s)}_{\mathbb{P}(Z \geqslant j)} = \sum_{j=1}^{+\infty} \mathbb{P}(Z \geqslant j)\,,$$

et la série de terme général $\mathbb{P}(Z \geqslant j)$ converge. Réciproquement, si la série des $\mathbb{P}(Z \geqslant j)$ converge, on note (en faisant le raisonnement ci-dessus en sens inverse) que Z admet une espérance, égale à la somme de cette série.

On peut aussi prouver ce résultat sans faire appel à l'interversion des sommations dans une série double : en utilisant la formule de sommation du lemme de la page 303 avec $a_k = \mathbb{P}(Z \geqslant k)$ pour tout $k \geqslant 0$ d'une part, $b_0 = 0$ et $b_k = 1$ pour tout $k \geqslant 1$ d'autre part, on obtient que pour tout $n \geqslant 1$,

$$\sum_{k=1}^{n} \mathbb{P}(Z \geqslant k) = n\,\mathbb{P}(Z \geqslant n) + \sum_{k=0}^{n-1} k\,\mathbb{P}(Z = k)\,. \tag{48.2}$$

Ainsi, si Z admet une espérance, la série de terme général $k\mathbb{P}(Z = k)$ converge et

$$n\,\mathbb{P}(Z \geqslant n) \leqslant \sum_{k=n}^{+\infty} k\,\mathbb{P}(Z = k) \xrightarrow[n\to+\infty]{} 0\,,$$

si bien que (48.2) montre que la série de terme général positif $\mathbb{P}(Z \geqslant k)$ converge, et en passant à la limite $n \to +\infty$ dans (48.2) on obtient que

$$\sum_{k=0}^{+\infty} \mathbb{P}(Z \geqslant k) = \mathbb{E}[Z]\,.$$

Réciproquement, si la série de terme général $\mathbb{P}(Z \geqslant k)$ converge, alors, puisque

$$\sum_{k=0}^{n-1} k\,\mathbb{P}(Z = k) \leqslant \sum_{k=1}^{n} \mathbb{P}(Z \geqslant k)$$

d'après (48.2), la série de terme général positif $k\,\mathbb{P}(Z = k)$ converge, et donc Z admet une espérance. On pourra noter la similitude entre cette démonstration et celle de la version continue de ce résultat, présentée à l'exercice 56.

3. Vu le résultat général prouvé à la question précédente, on commence par calculer $\mathbb{P}(Y \geqslant j)$ pour tout entier $j \geqslant 1$: la formule des probabilités totales indique que

$$\mathbb{P}(Y \geqslant j) = \sum_{k=1}^{n} \mathbb{P}(Y \geqslant j \mid X = k)\,\mathbb{P}(X = k)\,,$$

où pour tout $k \in \{1, \dots, n\}$, on a $\mathbb{P}(X = k) = 1/n$. En injectant l'égalité (48.1), valable pour $k \geqslant 2$, et $\mathbb{P}(Y \geqslant j \mid X = 1) = 0$ pour $j \geqslant 2$, il vient

$$\mathbb{P}(Y \geqslant j) = \frac{1}{n} \sum_{k=2}^{n} (1 - p_k)^{j-1} \qquad \text{pour} \qquad j \geqslant 2\,.$$

Par définition de Y, on a par ailleurs $\mathbb{P}(Y \geqslant 1) = 1$.
En particulier, $\mathbb{P}(Y \geqslant j)$ est bien le terme général d'une série convergente. Ainsi, par la question précédente, Y admet une espérance et

$$\begin{aligned} \mathbb{E}[Y] = \sum_{j=1}^{+\infty} \mathbb{P}(Y \geqslant j) &= 1 + \sum_{j=2}^{+\infty} \left(\frac{1}{n} \sum_{k=2}^{n} (1 - p_k)^{j-1} \right) \\ &= 1 + \frac{1}{n} \sum_{k=2}^{n} \sum_{j=2}^{+\infty} \left(\frac{k-1}{n} \right)^{j-1}. \end{aligned}$$

En découpant 1 en une somme de n termes égaux à $1/n$ et en utilisant la formule de sommation d'une série géométrique, on calcule finalement

$$\begin{aligned} \mathbb{E}[Y] = \frac{1}{n} + \frac{1}{n} \sum_{k=2}^{n} \sum_{j=1}^{+\infty} \left(\frac{k-1}{n} \right)^{j-1} &= \frac{1}{n} + \frac{1}{n} \sum_{k=2}^{n} \frac{1}{1 - \dfrac{k-1}{n}} \\ &= \frac{1}{n} + \sum_{k=2}^{n} \frac{1}{n-(k-1)} = \sum_{\ell=1}^{n} \frac{1}{\ell}\,, \end{aligned}$$

après le changement de variable de sommation $\ell = n - (k-1)$.
On reconnaît la somme partielle de la série harmonique. Une comparaison série-intégrale, comme dans la question 2 de l'exercice 36, montre que

$$\begin{aligned} \ln(n+1) = \int_1^{n+1} \frac{1}{x}\,\mathrm{d}x \leqslant \mathbb{E}[Y] = \sum_{\ell=1}^{n} \frac{1}{\ell} \\ \leqslant 1 + \int_1^{n} \frac{1}{x}\,\mathrm{d}x = 1 + \ln(n)\,. \end{aligned} \tag{48.3}$$

Bien sûr, d'autres encadrements légèrement différents sont possibles, notamment lorsque l'on traite les premiers termes de la somme à part ; on renvoie le lecteur à l'exercice 36 et à la question 4 ci-dessous.
Comme $\ln(n+1) = \ln(n) + \ln(1 + 1/n)$, où $\ln(1 + 1/n) \to 0$ lorsque $n \to +\infty$, il vient

$$1 + \frac{\ln(1+1/n)}{\ln(n)} \leqslant \frac{\mathbb{E}[Y]}{\ln(n)} \leqslant 1 + \frac{1}{\ln(n)}\,,$$

d'où par encadrement $\mathbb{E}[Y]/\ln(n) \to 1$, c'est-à-dire l'équivalent demandé $\mathbb{E}[Y] \sim \ln(n)$ lorsque $n \to +\infty$.

4. Ici encore, la question est volontairement formulée de manière à ce qu'il faille interpréter techniquement sa signification : comparer les gains « en moyenne » veut dire comparer les espérances de gains des deux joueurs. Le transfert de A vers B (le gain de B) est toujours égal à 3. L'espérance du transfert de B vers A (le gain de A) est précisément $\mathbb{E}[Y]$, une quantité que nous avons déterminée à la question précédente. Il s'agit donc de comparer $\mathbb{E}[Y]$ à 3.

   Or $\mathbb{E}[Y]$ est une quantité strictement croissante en n, car sa valeur pour un n donné est la somme partielle au rang n d'une suite de termes tous strictement positifs. Par ailleurs, nous avons montré qu'elle tendait vers $+\infty$, de sorte qu'elle dépasse en particulier la valeur 3. Le seuil critique n_0 est l'entier tel que

$$\sum_{k=1}^{n_0} \frac{1}{k} < 3 \qquad \text{mais} \qquad \sum_{k=1}^{n_0+1} \frac{1}{k} > 3\,;$$

   nous allons voir ci-dessous qu'aucune de ces deux quantités ne peut être exactement égale à 3 ; nous avons anticipé ce fait dans les inégalités ci-dessus en n'écrivant que des inégalités strictes. Ainsi, pour $n \leqslant n_0$ le jeu est favorable au joueur B, et pour $n \geqslant n_0 + 1$ c'est au joueur A qu'il l'est.

   En particulier, les encadrements de la question précédente conduisent à ce que $n_0 \leqslant n^\star$, pour tout entier $n^\star$ tel que

$$\sum_{k=1}^{n^\star+1} \frac{1}{k} \geqslant \ln\big((n^\star+1)+1\big) > 3\,, \qquad \text{soit} \qquad n^\star > \mathrm{e}^3 - 2 \approx 18{,}09\,.$$

   De même, $n_0 \geqslant n_\star$, pour tout entier $n_\star$ tel que

$$\sum_{k=1}^{n^\star} \frac{1}{k} \leqslant 1 + \ln(n_\star) < 3\,, \qquad \text{soit} \qquad n_\star < \mathrm{e}^2 \approx 7{,}39\,.$$

   Ainsi, on a déjà l'encadrement $7 \leqslant n_0 \leqslant 19$.

   En fait, on peut améliorer la minoration (48.3) en traitant le premier (ou les premiers) terme(s) à part. On obtient par exemple avec les deux premiers termes

$$\mathbb{E}[Y] \geqslant 1 + \frac{1}{2} + \int_3^{n+1} \frac{1}{x}\,\mathrm{d}x = \frac{3}{2} + \ln\Big(\frac{n+1}{3}\Big)\,,$$

ce qui conduit à la condition $n_0 \leqslant n^\star$ dès que

$$\frac{3}{2}+\ln\left(\frac{(n^\star+1)+1}{3}\right) > 3, \qquad \text{soit} \qquad n^\star > 3\,\exp\left(3-\frac{3}{2}\right)-2 \approx 11{,}45\,,$$

c'est-à-dire à l'encadrement $7 \leqslant n_0 \leqslant 12$. En calculant plus précisément (avec un ordinateur), on note que

$$\sum_{k=1}^{10}\frac{1}{k} \approx 2{,}93 \qquad \text{et} \qquad \sum_{k=1}^{11}\frac{1}{k} \approx 3{,}02\,,$$

ce qui montre que $n_0 = 10$.

Pour aller plus loin. On peut se demander ce qui se passerait pour la « limite continue » du modèle étudié ici (on appelle limite continue une version du problème discret qui apparaît comme sa limite quand le nombre n de valeurs discrètes tend vers l'infini). On va voir que cette version limite met en jeu des variables aléatoires de loi admettant une densité.

Si l'on numérote les faces du dé avec les nombres k/n, pour $k \in \{1,\ldots,n\}$, sans rien changer par ailleurs, la limite continue du tirage au hasard effectué par le dé sur cet ensemble de nombres est un tirage aléatoire uniforme sur $[0,1]$. Soit U et $V_1, V_2, \ldots$ des variables aléatoires indépendantes et de même loi donnée par cette loi uniforme sur $[0,1]$: la limite continue des temps d'attente étudiés dans le modèle discret est

$$Z = \min\Big\{j \geqslant 1 : \quad V_j \geqslant U\Big\}\,.$$

Nous avons ainsi défini notre modèle continu, et nous allons maintenant l'étudier en nous demandant quelle est l'espérance de Z (si elle existe) : est-ce la limite des espérances étudiées à la question 3, à savoir $+\infty$?

Nous allons à nouveau utiliser la formule

$$\mathbb{E}[Z] = \sum_{j=1}^{+\infty}\mathbb{P}(Z \geqslant j) \tag{48.4}$$

et nous devons donc déterminer

$$\mathbb{P}(Z \geqslant j) = \mathbb{P}\left(\bigcap_{1\leqslant\tau\leqslant j-1}\Big\{V_\tau < U\Big\}\right),$$

pour $j \geqslant 2$, cette probabilité valant 1 par construction de Z lorsque $j = 1$.

À notre niveau, il est difficile d'adapter le raisonnement utilisé dans cet exercice car on ne sait pas ce que cela veut dire que de conditionner par une valeur donnée de U ; en effet, pour tout $u \in [0,1]$, l'événement $\{U = u\}$ est de probabilité nulle ! Des notions plus avancées de calcul des probabilités montrent toutefois l'équivalent suivant de la formule des probabilités totales :

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{1\leqslant\tau\leqslant j-1}\big\{V_\tau < U\big\}\Big) &= \int_0^1 \mathbb{P}\Big(\bigcap_{1\leqslant\tau\leqslant j-1}\big\{V_\tau < u\big\}\Big)\mathrm{d}u \\ &= \int_0^1 \prod_{1\leqslant\tau\leqslant j-1} \mathbb{P}(V_\tau < u)\,\mathrm{d}u = \int_0^1 u^{j-1}\mathrm{d}u = \frac{1}{j}\,.\end{aligned}$$

De façon plus élémentaire, on peut recourir à un raisonnement par symétrie, qui lui, ne requiert aucune notion hors-programme. On définit l'événement

$$A(V_1,\ldots,V_{j-1},U) = \bigcap_{1\leqslant\tau\leqslant j-1}\big\{V_\tau < U\big\}\,.$$

C'est l'événement correspondant au fait que, parmi les j variables aléatoires indépendantes et identiquement distribuées $V_1,\ldots,V_{j-1},U$, ce soit la j-ème d'entre elles, U, qui soit la plus grande (au sens strict ou large : cela ne change pas la probabilité de l'événement). Or, par symétrie, cet événement a la même probabilité que celui où c'est la τ-ième variable aléatoire, avec $\tau \leqslant j-1$, qui est la plus grande. Comme la réunion de ces événements est l'univers tout entier, chacun de ces k événements est de probabilité $1/k$: on retrouve bien le résultat annoncé précédemment.

Finalement, en revenant à (48.4), on voit que Z n'admet pas d'espérance. Cela n'est pas une surprise, les espérances calculées à la question 3 tendant (logarithmiquement) vers $+\infty$ lorsque $n \to +\infty$...