Forum scientifique et pédagogique

Forum mathématique

Envoyez vos contributions de préférence à Claude MORIN (claude.morin@ac-limoges.fr) par courrier électronique ou par courrier postal au secrétariat.

Le problème du collectionneur de vignettes

par Gérard LAVAU¹ et Arnaud BÉGYN²

On souhaite collectionner des vignettes des héros de l'espace, vendues avec des tablettes de chocolat. Chaque tablette contient une vignette, et la collection complète comporte n vignettes. La probabilité de tirer une vignette donnée dans une tablette est supposée uniforme sur $\{1, \dots, n\}$. Quand il s'agit de confectionner une unique collection complète de vignettes, l'espérance du nombre de tablettes à acheter est bien connue. Si on note X_1 la variable aléatoire égale au nombre de tablettes pour acquérir une première vignette (de sorte que X_1 est la variable certaine égale à 1), X_2 la variable aléatoire égale au nombre de tablettes supplémentaires pour acquérir une deuxième vignette différente de la première, ..., X_k la variable aléatoire égale au nombre de tablettes supplémentaires pour acquérir une k -ème vignette différente des $k-1$ premières vignettes distinctes déjà acquises, etc., alors le nombre total de tablettes à acquérir pour confectionner une collection complète de toutes les vignettes est $\sum_{k=1}^n X_k$. Pour tout k , X_k suit une loi géométrique de paramètre

$\frac{n-k+1}{n}$ et son espérance est $\frac{n}{n-k+1}$. L'espérance du nombre de tablettes à acheter est $\sum_{k=1}^n \frac{n}{n-k+1} = \sum_{k=1}^n \frac{n}{k} = n \ln(n) + O(n)$.

En 1960, Donald J. Newman, Lawrence Shepp ont donné l'espérance du nombre de tablettes à acheter pour confectionner un nombre m quelconque de collections complètes (*The Double Dixie Cup Problem*, Am. Math. Monthly, Vol. 67, n° 1, Jan. 1960, pp. 58-61), ainsi

1. Lycée Carnot, Dijon

2. Lycée Leconte de Lisle, Saint-Denis de la Réunion

qu'un développement asymptotique du nombre de tablettes à acheter lorsque n tend vers l'infini, à savoir

$$n(\ln(n) + (m - 1)\ln(\ln(n)) + O(1)).$$

On se borne ci-dessous à adapter leur article dans le cas où $m = 2$, situation éventuellement adaptable en problème auprès de nos étudiants.

Deux enfants souhaitent donc confectionner deux collections complètes. On note Z_n le nombre de tablettes à acheter pour y parvenir. Pour tout entier $k \geq 0$, on note p_k la probabilité d'acheter k tablettes et de ne pas obtenir avec ces k tablettes deux collections complètes.

On a donc $p_k = P(Z_n > k)$ et $E(Z_n) = \sum_{k=1}^{\infty} P(Z_n \geq k) = \sum_{k=0}^{\infty} P(Z_n > k) = \sum_{k=0}^{\infty} p_k$ (la convergence de la série est prouvée dans la démonstration de la proposition 1 ci-dessous). On a par ailleurs $p_k = N_k/n^k$, où N_k est le nombre de façons de ne pas avoir deux collections complètes de n vignettes après l'achat de k tablettes. N_k se dénombre en considérant qu'on a k vignettes au total et au moins une des n images de la collection n'est qu'en quantité 0 ou 1, de sorte que

$$N_k = \sum \frac{k!}{i_1!i_2! \dots i_n!},$$

où la somme porte sur les indices $i_1, \dots, i_n$ tels que $i_1 + \dots + i_n = k$ et au moins un des indices i_j est strictement inférieur à 2 (i_j est égal au nombre de représentants de la j -ième vignette de la collection). En complément, la somme $\sum \frac{k!}{i_1!i_2! \dots i_n!}$ portant sur les indices $i_1, \dots, i_n$ tous supérieurs ou égaux à 2 et tels que $i_1 + \dots + i_n = k$ vaut $n^k - N_k$.

On dispose du premier résultat suivant.

Proposition 1. Soit F la fonction définie par $F(t) = e^{nt} - (e^t - 1 - t)^n$. L'espérance du nombre Z_n de tablettes à acheter pour confectionner deux collections complètes de n vignettes est :

$$E(Z_n) = n \int_0^{\infty} F(t)e^{-nt} dt. \tag{1}$$

En effet, $e^t - 1 - t = \sum_{i=2}^{\infty} t^i/i!$ de sorte que, en effectuant le produit de Cauchy n fois de cette série :

$$(e^t - 1 - t)^n = \sum_{k=0}^{\infty} \sum \frac{1}{i_1!i_2! \dots i_n!} t^k = \sum_{k=0}^{\infty} \sum \frac{k!}{i_1!i_2! \dots i_n!} \frac{t^k}{k!}, \tag{2}$$

la somme intérieure $\sum \frac{k!}{i_1!i_2! \dots i_n!}$ portant sur les $i_1, \dots, i_n$ tous supérieurs ou égaux à 2 et tels que $i_1 + \dots + i_n = k$. On a donc :

$$(e^t - 1 - t)^n = \sum_{k=0}^{\infty} \frac{n^k - N_k}{k!} t^k \tag{3}$$

et

$$F(t) = \sum_{k=0}^{\infty} \frac{N_k}{k!} t^k. \tag{4}$$

Or, pour tout $k \geq 0$ et tout $n \geq 1$, $\frac{1}{n^k} = n \int_0^\infty \frac{t^k}{k!} e^{-nt} dt$, donc

$$\sum_{k=0}^\infty \frac{N_k}{n^k} = n \int_0^\infty F(t) e^{-nt} dt.$$

On peut appliquer le théorème d'intégration terme à terme et permuter série et intégrale car, la série définissant $F(t)$ étant à termes positifs, il suffit de vérifier que la série $\sum \frac{N_k}{n^k}$ converge et, pour cela, il suffit de vérifier que ses sommes partielles sont majorées, or elles le sont précisément par $n \int_0^\infty F(t) e^{-nt} dt$ dont on pourra vérifier qu'il s'agit d'une intégrale convergente. Comme on a vu que $\sum_{k=0}^\infty \frac{N_k}{n^k}$ n'était autre que $E(Z_n)$, on a montré le résultat annoncé.

Montrons maintenant que :

Proposition 2. *Quand n tend vers l'infini, on a :*

$$E(Z_n) = n \ln(n) + n \ln(\ln(n)) + O(n).$$

Autrement dit, le nombre de tablettes à acheter pour confectionner une deuxième collection est négligeable devant le nombre de tablettes pour confectionner la première collection, ce qui satisfera les parents de plusieurs enfants.

Pour tout entier $n \geq 1$, posons

$$I_n = \int_0^\infty 1 - (1 - (1 + t)e^{-t})^n dt, \tag{5}$$

de sorte que $E(Z_n) = nI_n$. Le but est de montrer que

$$I_n = \ln(n) + \ln(\ln(n)) + O(1)$$

quand n tend vers $+\infty$. On obtient ce résultat par une suite d'inégalités conduisant à un encadrement adéquat.

Remarquons d'abord que $I_n - I_{n-1} = \int_0^\infty (1 - (1 + t)e^{-t})^{n-1} (1 + t)e^{-t} dt$. Considérons la fonction $f : t \in [0, +\infty[\mapsto 1 - (1 + t)e^{-t} : f$ est de classe $\mathcal{C}^1$, strictement croissante et bijective de $[0, +\infty[$ sur $[0, 1[$. Si on effectue le changement de variable $x = f(t)$, on obtient $dx = te^{-t} dt$ et

$$I_n - I_{n-1} = \int_0^1 x^{n-1} \left(1 + \frac{1}{f^{-1}(x)} \right) dx. \tag{6}$$

Lemme 1. *Pour tout x élément de $[0, 1[$, $f^{-1}(x) \geq -\ln(1 - x)$.*

En effet :

$$\begin{aligned} f^{-1}(x) \geq -\ln(1 - x) &\iff x \geq f(-\ln(1 - x)) = 1 - (1 - \ln(1 - x))(1 - x) \\ &\iff x \geq 1 - (1 - x - (1 - x)\ln(1 - x)) \iff 0 \geq (1 - x)\ln(1 - x), \end{aligned}$$

qui est vrai car x est compris entre 0 et 1.

On en déduit également que :

Lemme 2. Pour tout entier $r \geq 1$ et tout x dans $[0, 1[$, $f^{-1}(x) \geq x^r \ln(r+1)$.

En effet :

$$f^{-1}(x) \geq -\ln(1-x) = \sum_{k=1}^{\infty} \frac{x^k}{k} \geq \sum_{k=1}^r \frac{x^k}{k} \geq x^r \sum_{k=1}^r \frac{1}{k} \geq x^r \ln(r+1). \quad (7)$$

On en déduit que :

Lemme 3. Pour tout entier $r \geq 1$, avec $r < n$:

$$\int_0^1 \frac{x^{n-1}}{f^{-1}(x)} dx \leq \frac{1}{(n-r) \ln(r+1)}. \quad (8)$$

En effet, $\int_0^1 \frac{x^{n-1}}{f^{-1}(x)} dx \leq \int_0^1 \frac{x^{n-1}}{x^r \ln(r+1)} dx = \frac{1}{(n-r) \ln(r+1)}$.

D'autre part :

Lemme 4. Si on prend un réel $u \geq 1$ et $a = f(u)$, alors :

$$\int_0^1 \frac{x^{n-1}}{f^{-1}(x)} dx \geq \frac{1}{u} \int_0^a x^{n-1} dx \geq \frac{1}{nu} - \frac{(1+u)e^{-u}}{u} \geq \frac{1}{nu} - 2e^{-u}. \quad (9)$$

En effet, comme a appartient à $[0, 1[$ et qu'on intègre des fonctions positives, on a :

$$\int_0^1 \frac{x^{n-1}}{f^{-1}(x)} dx \geq \int_0^a \frac{x^{n-1}}{f^{-1}(x)} dx. \quad (10)$$

Or, sur $[0, a]$, $f^{-1}(x) \leq f^{-1}(a) = u$, donc

$$\int_0^1 \frac{x^{n-1}}{f^{-1}(x)} dx \geq \frac{1}{u} \int_0^a x^{n-1} dx = \frac{a^n}{nu} \quad (11)$$

Il suffit donc ensuite de montrer que $\frac{a^n}{nu} \geq \frac{1}{nu} - \frac{(1+u)e^{-u}}{u}$, ce qui équivaut à $\frac{1-a^n}{n} \leq (1+u)e^{-u} = 1 - f(u) = 1 - a$, ou à $\frac{1-a^n}{1-a} \leq n$, ou encore à $1 + a + \dots + a^{n-1} \leq n$, ce qui est vrai car $a \leq 1$. Enfin $\frac{1}{nu} - \frac{(1+u)e^{-u}}{u} \geq \frac{1}{nu} - 2e^{-u}$ équivaut à $\frac{1+u}{u} \leq 2$ qui est vrai car, comme $u \geq 1$, $\frac{1+u}{u} = 1 + \frac{1}{u} \leq 2$.

Lemme 5. Pour $n \geq 3$, posons maintenant $r = \left\lfloor \frac{n}{\ln(n)} \right\rfloor$ (où $\lfloor \cdot \rfloor$ désigne la partie entière) et $u = \ln(n) + 2 \ln(\ln(n))$. Alors il existe une constante C telle que, pour tout $n \geq 3$:

$$\frac{1}{nu} - 2e^{-u} \geq \frac{1}{n \ln(n)} - \frac{C \ln(\ln(n))}{n \ln(n)^2} \quad (12)$$

et

$$\frac{1}{(n-r) \ln(r+1)} \leq \frac{1}{n \ln(n)} + \frac{C \ln(\ln(n))}{n \ln(n)^2}. \quad (13)$$

On a :

$$\begin{aligned} \frac{1}{nu} - 2e^{-u} &= \frac{1}{n(\ln(n) + 2\ln(\ln(n)))} - \frac{2}{n\ln(n)^2} \\ &= \frac{1}{n\ln(n)} \left(1 + \frac{2\ln(\ln(n))}{\ln(n)} \right)^{-1} - \frac{2}{n\ln(n)^2} \\ &= \frac{1}{n\ln(n)} \left(1 - \frac{2\ln(\ln(n))}{\ln(n)} + o\left(\frac{\ln(\ln(n))}{\ln(n)}\right) \right) - \frac{2}{n\ln(n)^2} \\ &= \frac{1}{n\ln(n)} - \frac{2\ln(\ln(n))}{n\ln(n)^2} + o\left(\frac{\ln(\ln(n))}{n\ln(n)^2}\right) - \frac{2}{n\ln(n)^2} \\ &= \frac{1}{n\ln(n)} - \frac{2\ln(\ln(n))}{n\ln(n)^2} + o\left(\frac{\ln(\ln(n))}{n\ln(n)^2}\right), \end{aligned}$$

car $\frac{2}{n\ln(n)^2} = o\left(\frac{\ln(\ln(n))}{n\ln(n)^2}\right)$. La constante C cherchée doit être telle que :

$$\frac{1}{n\ln(n)} - \frac{2\ln(\ln(n))}{n\ln(n)^2} + o\left(\frac{\ln(\ln(n))}{n\ln(n)^2}\right) \geq \frac{1}{n\ln(n)} - \frac{C\ln(\ln(n))}{n\ln(n)^2}, \tag{14}$$

ce qui équivaut à $2 + o(1) \leq C$, ce qui est possible car $o(1)$ est une quantité bornée.

En ce qui concerne l'inégalité (13), comme $r + 1 \geq \frac{n}{\ln(n)}$ et $n - r \geq n - \frac{n}{\ln(n)}$, on a :

$$\begin{aligned} \frac{1}{(n-r)\ln(r+1)} &\leq \frac{1}{\left(n - \frac{n}{\ln(n)}\right) \ln \frac{n}{\ln(n)}} \\ &\leq \frac{1}{n\ln(n)} \left(1 - \frac{1}{\ln(n)} \right)^{-1} \left(1 - \frac{\ln(\ln(n))}{\ln(n)} \right)^{-1} \\ &\leq \frac{1}{n\ln(n)} \left(1 + \frac{1}{\ln(n)} + o\left(\frac{1}{\ln(n)}\right) \right) \left(1 + \frac{\ln(\ln(n))}{\ln(n)} + o\left(\frac{\ln(\ln(n))}{\ln(n)}\right) \right) \\ &\leq \frac{1}{n\ln(n)} \left(1 + o\left(\frac{\ln(\ln(n))}{\ln(n)}\right) \right) \left(1 + \frac{\ln(\ln(n))}{\ln(n)} + o\left(\frac{\ln(\ln(n))}{\ln(n)}\right) \right) \\ &\leq \frac{1}{n\ln(n)} \left(1 + \frac{\ln(\ln(n))}{\ln(n)} + o\left(\frac{\ln(\ln(n))}{\ln(n)}\right) \right) \\ &\leq \frac{1}{n\ln(n)} + \frac{\ln(\ln(n))}{n\ln(n)^2} + o\left(\frac{\ln(\ln(n))}{n\ln(n)^2}\right). \end{aligned}$$

On prendra également C tel que $C \geq 1 + o(1)$.

En regroupant (8), (9), (12) et (13), on obtient, pour tout $n \geq 3$:

$$\frac{1}{n\ln(n)} - \frac{C\ln(\ln(n))}{n\ln(n)^2} \leq \int_0^1 \frac{x^{n-1}}{f^{-1}(x)} dx \leq \frac{1}{n\ln(n)} + \frac{C\ln(\ln(n))}{n\ln(n)^2}. \tag{15}$$

On en déduit :

Lemme 6. *Il existe une constante D telle que, pour tout $n \geq 3$:*

$$\sum_{k=3}^n \frac{1}{k\ln(k)} - D \leq \sum_{k=3}^n \int_0^1 \frac{x^{k-1}}{f^{-1}(x)} dx \leq \sum_{k=3}^n \frac{1}{k\ln(k)} + D. \tag{16}$$

En effet, la série $\sum \frac{\ln(\ln(n))}{n \ln(n)^2}$ converge (son terme général est un $O\left(\frac{1}{n \ln(n)^{\frac{3}{2}}}\right)$, et l'on tombe sur une série de Bertrand dont la convergence se prouve par une comparaison série-intégrale). Il suffit donc de sommer l'encadrement (15) et de prendre D égal à C fois la somme de la dite série pour conclure.

On en tire ensuite :

Lemme 7. *Il existe une constante E telle que, pour tout $n \geq 3$:*

$$\ln(\ln(n)) - E \leq \sum_{k=3}^n \int_0^1 \frac{x^{k-1}}{f^{-1}(x)} dx \leq \ln(\ln(n)) + E. \quad (17)$$

Pour cela, on encadre $\sum_{k=3}^n \frac{1}{k \ln(k)}$ au moyen d'une comparaison série intégrale :

$$\int_3^n \frac{1}{x \ln(x)} dx + \frac{1}{n \ln(n)} \leq \sum_{k=3}^n \frac{1}{k \ln(k)} \leq \int_2^n \frac{1}{x \ln(x)} dx, \quad (18)$$

ou encore

$$\ln(\ln(n)) - \ln(\ln(3)) + \frac{1}{n \ln(n)} \leq \sum_{k=3}^n \frac{1}{k \ln(k)} \leq \ln(\ln(n)) - \ln(\ln(2)). \quad (19)$$

Il suffit alors de prendre $E = \ln(\ln(3)) + D$.

Enfin, on conclut la démonstration de la proposition 2 au moyen de :

Lemme 8. *Il existe une constante F telle que, pour tout $n \geq 3$:*

$$\ln(n) + \ln(\ln(n)) - F \leq I_n \leq \ln(n) + \ln(\ln(n)) + F. \quad (20)$$

Comme $I_k - I_{k-1} = \int_0^1 x^{k-1} \left(1 + \frac{1}{f^{-1}(x)}\right) dx = \frac{1}{k} + \int_0^1 \frac{x^{k-1}}{f^{-1}(x)} dx$, on a, en sommant de 3 à n et en utilisant l'encadrement (17) :

$$\sum_{k=3}^n \frac{1}{k} + \ln(\ln(n)) - E \leq I_n - I_2 \leq \sum_{k=3}^n \frac{1}{k} + \ln(\ln(n)) + E. \quad (21)$$

Mais, comme ci-dessus, $\ln(n) - \ln(3) + \frac{1}{n} \leq \sum_{k=3}^n \frac{1}{k} \leq \ln(n) - \ln(2)$. On prend $F = E + \ln(3) + I_2$.