Les urnes de Pòlya
Analyse mathématique

Équipe DREAM

12 juillet 2020

Table des matières

1 L'énoncé du problème

1.1 L'énoncé de Pòlya

Une urne contient originalement N boules dont R sont rouges et S noires, $R+S=N$. Nous faisons de l'urne des tirages successifs en ajoutant à l'urne, après chaque tirage, à la place de la boule tirée $1+\Delta$ boules *de la même couleur*. Si Δ est positif, les nombre des boules augmente après chaque tirage, chaque succès obtenu favorise les chances des succès à obtenir, chaque insuccès gâte encore les chances des épreuves suivantes, le succès ainsi que l'insuccès sont *contagieux*. Ajouter un nombre négatif signifie enlever ; donc si Δ est négatif, le nombre de boules diminue au cours des tirages, chaque succès diminue les chances de succès ultérieurs ; mais aussi les insuccès dont de la même nature et chaque coup tend à amener un revirement de fortune. (Pòlya, 1930, page 136)

1.2 Une variante

On dispose d'une urne contenant une boule rouge et d'une boule noire. L'expérience consiste à choisir une boule dans l'urne, à noter sa couleur, à la remettre dans l'urne en ajoutant une boule de même couleur, puis de réitérer le processus.
La question qui se pose est de déterminer la composition de l'urne après n tirages.

2 Solution(s), piste(s) de solution(s)

2.1 Le raisonnement de Pòlya

D'après Pòlya (1930) page 137.
On associe à chaque tirage les variables aléatoires x_1, x_2, $\ldots x_n$ tels que, par exemple, $x_k=1$ si le tirage fait sortir une boule rouge, $x_k=0$ si la boue est noire.
Après k tirages, l'urne contiendra un nombre total de $N+k\Delta$ boules dont $N+(\Sigma_1^k x_i)\Delta$ rouges.
La probabilité d'obtenir une boule rouge au tirage $k+1$ est donc :

$$p_{k+1}=\frac{N+(\Sigma_1^k x_i)\Delta}{N+k\Delta}$$

En posant $\rho=\frac{R}{N}$ et $\sigma=1-\rho=\frac{S}{N}$ et $\delta=\frac{\Delta}{N}$ on obtient :

$$p_{k+1}=\frac{\rho+\left(\Sigma_{i=1}^k\right)\delta}{1+k\delta}$$

Cette probabilité contient deux paramètres, δ qui correspond à la proportion initiale de boules rouges dans l'urne et δ indique l'intensité de la dépendance et la direction de l'évolution de l'urne. C'est un nombre compris entre 0 et 1.
En particulier si $R=S=1$ et que l'on rajoute une boule de la même couleur que celle choisie, alors $\rho=\frac{1}{2}$, $\delta=\frac{1}{2}$ et

$$p_{k+1}=\frac{1+\Sigma_{i=1}^k x_i}{k+2}$$

Sous une hypothèse d'équiprobabilité, le nombre de boules rouges dans l'urne à l'étape k est uniformément distribué sur $\{1,2,\ldots,k+1\}$. Autrement dit, en recommençant plusieurs fois l'expérience jusqu'à une étape donnée, toutes les compositions d'urne possibles sont équiprobables.

2.2 Quelques résultats dans le cas de l'énoncé

Une schématisation peut permettre de mieux comprendre le phénomène étudié :

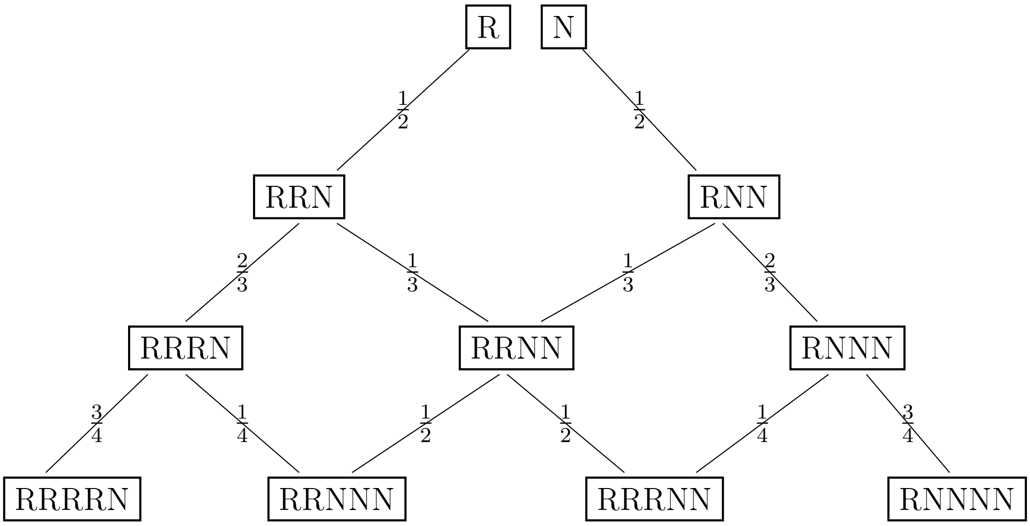

FIGURE 1: Quatre étapes : composition de l'urne

Le schéma de la figure 1 permet de conjecturer le résultat annoncé dans le paragraphe précédent : à l'étape 1, la probabilité d'obtenir une des deux urnes possibles est de $\frac{1}{2}$. A l'étape 2, les trois urnes ont également la même probabilité d'apparaître, ce que l'on vérifie encore à l'étape 4. Et, à chaque étape, la probabilité d'obtenir une boule rouge (resp.noire) vaut $\frac{1}{2}$.
On peut alors penser à un raisonnement par récurrence. Démontrons dans un premier temps qu'à chaque étape la probabilité d'obtenir une boule noire (resp. rouge) est de $\frac{1}{2}$.
C'est vrai à l'étape 1.
Supposons que ce soit vrai à l'étape n. D'après la formule des probabilités totales, la probabilité d'obtenir une boule rouge à l'étape $n+1$ est égale à la somme des probabilités d'obtenir une boule rouge sachant que la composition de l'urne contenait $1, 2, 3, n-1$ boules rouges. C'est à dire :

$$\begin{aligned} p(\text{obtenir une boule rouge}) &= \frac{1}{n+1} \times \frac{1}{n+2} + \frac{1}{n+1} \times \frac{2}{n+2} + \cdots + \frac{1}{n+1} \times \frac{n+1}{n+2} \\ &= \frac{1}{n+1} \times \left(\frac{(n+1)(n+2)}{2(n+2)} \right) = \frac{1}{2} \end{aligned}$$

Montrons alors que chaque composition d'urne est équiprobable, c'est à dire que la probabilité d'avoir une composition d'urne à l'étape $n+1$ vaut $\frac{1}{n+2}$.
A l'étape $n+1$ les urnes ont pour composition :
1 rouge n noires, 2 rouges, $n-1$ noires, 3 rouges, $n-2$ noires,...n rouges, 1 noire.
Notons B_i^k l'urne ayant i boules rouges à l'étape k.
La probabilité d'obtenir l'urne B_i^{k+1} est la probabilité d'obtenir à l'étape précédente une urne ayant $i-1$ boules rouges (B_{i-1}^k) et de tirer une boule rouge ou d'obtenir à l'étape précédente i boules rouges (B_i^k) et de tirer une boule noire. Donc :

$$\begin{aligned} p(B_i^{k+1}) &= p(B_{i-1}^k) \times \frac{i-1}{k+1} + p(B_i^k) \times \frac{k+1-i}{k+1} \\ &= \frac{1}{k+1} \left(\frac{i-1+k+2-i}{k+2} \right) \end{aligned}$$

$$= \frac{1}{k+2}$$

Démontrons maintenant que la proportion de boules rouges dans l'urne converge presque surement vers une variable aléatoire qui est uniformément distribué sur [0,1].
Soit X_n le nombre de boules rouges à l'étape n. Soit la variable aléatoire $Y_n = \frac{X_n}{n+2}$. Elle correspond à la proportion de boules rouges dans l'urne. $Y_n \to Y$ presque sûrement et Y est une variable aléatoire uniformément distribuée sur [0,1].
La convergence presque sûre affirme qu'il peut exister des suites de tirages qui ne converge pas vers une valeur unique mais que la probabilité de telles suites est nulle.
La suite des v.a. X_n n'est pas une suite de v.a. indépendantes mais elles sont échangeables, c'est à dire que l'ordre ne compte pas. Autrement dit, quelque soit la permutation σ la loi de $X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)}$ est la même que la loi de $X_1, X_2, \ldots, X_n$.
Par exemple :
La probabilité d'avoir la séquence une boule rouge (r), une boule rouge une boule noire (b) est donnée par

$$p(rrb) = \frac{r}{r+b} \times \frac{r+1}{r+b+1} \times \frac{b}{r+b+2}$$

Et c'est la même probabilité d'avoir la suite une boule rouge, une boule noire une boule rouge :

$$p(rbr) = \frac{r}{r+b} \times \frac{b}{r+b+1} \times \frac{r+1}{r+b+2}$$

C'est ce que l'on voit sur le schéma 1 puisque les feuilles de l'arbre proviennent de deux racines (sauf celles sur les bords) et donc avoir b boules noires à une certaine étape peut peut provenir du fait qu'on en avait b à l'étape précédente et qu'on a tiré une boule rouge ou bien qu'on en avait $b-1$ et qu'on en tire une à cette étape.
Le théorème de de Finetti établit que :
Si une suite infinie de v.a. X_i est échangeable alors il existe une distribution F telle que :

$$p_k = p(X_1 + X_2 + \cdots + X_n = k) = \int_0^1 \begin{pmatrix} n \\ k \end{pmatrix} x^k (1-x)^{n-k} dF(x)$$

Dans le cas des urnes de Pòlya, la suite x_k suit ainsi la même loi que : tirer U une loi uniforme sur $[0,1]$ puis effectuer une suite de tirages indépendants de Bernouilli de paramètre U.
On peut en déduire, bien sûr que la suite de variable converge en loi. Cependant, on peut aussi démontrer cette convergence directement (ce qui peut être intéressant pour des élèves de lycée).
En effet :
Soit donc F_n la fonction de répartition de la v.a. $Y_n = \frac{X_n}{n+2}$.
Pour $x \in]0,1[$,

$$F(x) = p(Y_n \leq x) = p(X_n \leq (n+2)x = p(X_n \leq \llcorner (n+2)x \lrcorner) = \frac{\llcorner (n+2)x \lrcorner}{n+2}$$

Bien sûr $f(x) = 0$pour $x < 0$ et $F(x) = 1$ pour x > 1
On a donc pour $0 \leq 1$:

$$\frac{(n+2)x - 1}{n+2} \leq \frac{\llcorner (n+2)x \lrcorner}{n+2} \leq x$$

Et par conséquent :

$$\lim_{n \to \infty} F_n(x) = x$$

Et par conséquent, la loi de Y_n tend en loi vers une variable aléatoire Y de loi uniforme sur $[0,1]$.

Référence

Pòlya, G. (1930). Sur quelques points de la théorie des probabilités, *Annales de l'IHP*, tome 1, n°2,117-161