Quel hasard ?

Question : Quelle est la probabilité pour que r boules réparties au hasard dans n boîtes laissent exactement k boîtes vides ?

Première méthode. On différentie les boîtes et les boules. Une répartition des r boules dans les n boîtes s'identifie à une application $\{1,\ldots,r\} \to \{1,\ldots,n\}$. Soit Ω l'ensemble de ces applications et A la partie de Ω représentant l'évènement "k boîtes exactement sont vides". Si toutes les répartitions sont équiprobables, la probabilité est uniforme sur Ω et $\mathrm{P}(A)$ est le quotient $\frac{|A|}{|\Omega|}$. On a $|\Omega| = n^r$; on veut donc calculer $|A|$.

Un élément de A est une application évitant exactement k valeurs. Il y a C_n^k façons de choisir ces k valeurs. Une fois ces valeurs choisies, il faut aussi choisir une application surjective de $\{1,\cdots,r\}$ dans un ensemble à $n-k$ éléments.

Notons $a_{r,n}$ le nombre d'applications surjectives de $\{1,\cdots,r\}$ dans $\{1,\cdots,n\}$. La probabilité cherchée est

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{\mathrm{C}_n^k a_{r,n-k}}{n^r} \ .$$

Pour calculer $a_{r,n}$ on peut utiliser la formule de Poincaré : Le complémentaire dans Ω de l'ensemble des applications surjectives est constitué des applications évitant au moins une valeur. Notons, pour i dans $\{1,\cdots,n\}$, A_i l'ensemble des applications évitant la valeur i. Pour $i_1 < \cdots < i_j$ des éléments de $\{1,\cdots,n\}$, l'intersection $A_{i_1} \cap \cdots \cap A_{i_j}$ est l'ensemble des applications évitant les valeurs $i_1,\cdots,i_j$ donc est de cardinal $(n-j)^r$. On obtient

$$a_{r,n} = |\Omega| - |\cup_{i=1\ldots n} A_i| = n^r - \textstyle\sum_{j=1}^n \sum_{1\le i_1<\cdots<i_j\le n} (-1)^{j+1}(n-j)^r$$

$$= \textstyle\sum_{j=0}^n \mathrm{C}_n^j (-1)^{n+j} j^r.$$

Il n'est pas facile d'aller plus loin.

Deuxième méthode. On différentie les boîtes mais pas les boules. Une répartition des r boules dans les n boîtes s'identifie alors à une suite $(n_1,\ldots,n_r)$ d'entiers positifs de somme r. Soit Ω' l'ensemble de ces suites. Une répartion laissant exactement k boîtes vides s'identifie à une suite $(n_1,\ldots,n_r)$ telle que $n_i = 0$ pour exactement k indices i. A nouveau il y a C_n^k choix possibles des indices concernés et pour chacun de ces choix il faut dénombrer les suites de $n-k$ entiers strictement positifs de somme r, ce qui équivaut à dénombrer les suites de $n-k$ entiers positifs ou nuls et de somme $r-(n-k)$.

Soit $b_{n,r}$ le cardinal de Ω'. Si les répartitions sont équiprobables, la probabilité cherchée est donc

$$\mathrm{P} = \frac{\mathrm{C}_n^k\, b_{n-k,r-n+k}}{b_{n,r}} \ .$$

On peut calculer $b_{n,r}$ en l'interprétant comme le coefficient de x^r dans la série produit $(\sum_{i\ge 0} x^i)^n$. Or $\sum_{i\ge 0} x^i$ est le développement en série entière de $x \mapsto \frac{1}{1-x}$ en 0 et la dérivée r-ième de la fonction $x \mapsto \frac{1}{(1-x)^n}$ en 0 vaut $n(n+1)\cdots(n+r-1) = r!\,\mathrm{C}_{n+r-1}^r$, donc $b_{n,r}$ vaut C_{n+r-1}^r.

Comparons les résutats des deux méthodes. Le quotient des deux probabilités trouvées est

$$\frac{a_{r,n-k} b_{n,r}}{n^r b_{n-k,r-n+k}} \ .$$

Si les probabilités sont égales alors le quotient $\frac{b_{n,r}}{r^n}$ ne dépend que de $n-k$ donc ne dépend pas de n, ce qui semble très faux. En fait pour $n=3$, $r=3$ et k décrivant $\{0,1,2\}$ les valeurs des probabilités trouvées par les deux méthodes sont respectivement

$$\frac{2}{9}, \frac{2}{3}, \frac{1}{9} \text{ et } \frac{1}{10}, \frac{3}{5}, \frac{3}{10} \ .$$

Que s'est-il passé ?

dehon@math.unice.fr