Quelques calculs de probabilités dans la vie courante

Robert Ferréol

Souvent, les problèmes de mathématiques posés de façon concrète sont des problèmes abstraits ayant reçu un habillage ; ceux que je vais proposer et résoudre dans cet article sont au contraire de vrais problèmes que je me suis posé dans la vie courante. Puissent-ils vous servir à répondre à la sempiternelle question : à quoi servent les maths ?

J'utiliserai souvent deux propriétés très utiles reliant les probabilités et les espérances :

Propriété 1, élémentaire, mais plus utile qu'on ne pense : la probabilité d'un événement est aussi l'espérance de la variable aléatoire valant 1 si l'événement est réalisé et 0 sinon.

Propriété 2, très importante pour la compréhension des probabilités et rarement mentionnée : lors d'une succession d'épreuves aléatoires répétées indépendantes, si un événement a une probabilité p, cet événement arrive en moyenne tous les $1/p$ coups.

Cette propriété est très naturelle : elle signifie par exemple que si un événement a une chance sur 3 d'arriver, il arrivera en moyenne une fois sur 3…

Autre exemple : la probabilité que j'aie 4 as lors d'une distribution de 52 cartes entre 4 joueurs est égale à $\dfrac{C_{48}^{9}}{C_{52}^{13}} \approx 0{,}0027$; ce nombre n'est pas très parlant. Mais si l'on dit que cela signifie que j'aurai en moyenne 4 as tous les $\dfrac{C_{52}^{13}}{C_{48}^{9}} \approx 379$ coups, je trouve que ça l'est plus !

Voici la démonstration de cette propriété :

Un événement de probabilité p va arriver au bout de k coups avec la probabilité $(1-p)^{k-1}\,p$; le temps d'attente moyen (en nombre de coups) de cet événement est donc $T_A = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$; or on sait que $\sum_{k=1}^{\infty} kx^{k-1} = \dfrac{d}{dx}\left(\sum_{k=1}^{\infty} x^k\right) = \dfrac{1}{(1-x)^2}$; donc

$$T_A = \frac{p}{(1-1+p)^2} = \frac{1}{p}.$$

Voici donc les petits problèmes que je vous propose.

1) La consultation d'un dictionnaire en 15 volumes.

Je reconnais que ce problème, à l'heure des encyclopédies sur CD-ROM est un peu désuet. Lorsque vous avez, disons, n = 10 mots quelconques à consulter dans un dictionnaire en p = 15 volumes, vous avez espoir de ne pas avoir à ouvrir 10 volumes différents ! Mais combien de volumes aurez-vous à ouvrir en moyenne ?

Solution :

L'espérance du nombre d'ouvertures de volumes est la somme des espérances du nombre d'ouverture de *chacun* des volumes, pour la consultation des n mots. Or la contribution d'un volume donné au nombre de volumes à consulter a une espérance E_0 qui se confond d'après la propriété 1 avec la probabilité d'utiliser ce volume, puisque cette contribution est 1 quand on l'utilise, 0 sinon.
Comme pour chaque dictionnaire, j'ai une chance sur p qu'un mot donné se trouve dedans (on considère que tous les dictionnaires ont le même nombre de mots) , la probabilité d'utiliser un volume donné quand je cherche n mots est

$$E_0 = 1 - \left(1 - \frac{1}{p}\right)^n .$$

L'espérance cherchée est donc

$$E_0 = n\left(1 - \left(1 - \frac{1}{p}\right)^n\right) = n - \frac{C_n^2}{p} + \frac{C_n^3}{p^2} - \frac{C_n^4}{p^3} + \ldots ;$$

cela donne environ 7,5 volumes à ouvrir pour chercher 10 mots dans un dictionnaire en 15 volumes.

Il faut dire que cette rapide solution probabiliste a mis du temps à germer ; en 1990, j'avais posé ce problème dans la revue Quadrature et, dans la solution proposée, la valeur de E obtenue à partir de la loi de probabilité du nombre de volumes à consulter était

$$E = \frac{1}{p^n} \sum_{k=1}^{n} k\, S_n^k\, C_p^k$$

où S_n^k est le nombre de surjection d'un n-ensemble vers un k--ensemble ! Il était ensuite démontré que cela donnait bien la valeur ci-dessus, et J. Moreau de Saint Martin avait trouvé une démonstration directe utilisant les espérances conditionnelles. Mais ce n'est que très récemment que ce dernier m'a envoyé cette élégante démonstration.

2) Étude d'une loterie.

Ce problème m'a été inspiré par une loterie organisée par le village de Saint-Étienne de Tinée, près de Nice. Comme toute loterie, cette loterie possède n (= 1 000) billets dont p (= 30) sont gagnants ; combien faut-il que j'achète de billets pour que la probabilité d'en avoir au moins un gagnant soit supérieure à 1/2 ?
Et combien faut-il que j'achète de billets pour que l'espérance du nombre de mes billets gagnants soit supérieure à 1 ?
Sachant que chaque lot vaut en moyenne r euros et que chaque billet coûte s euros, quelle est mon espérance de gain quand j'achète q billets (à savoir l'espérance de

l'argent rapporté par les lots moins les qs euros). Et je souhaite surtout savoir pour quel nombre de billets achetés cette espérance de gain est maximale.

Ma solution.

Si j'achète q billets, la loi de probabilité du nombre G de billets gagnants est une loi hypergéométrique :

$$\mathrm{P}([\mathrm{G}=k]) = \frac{\mathrm{C}_p^k\,\mathrm{C}_{n-p}^{q-k}}{\mathrm{C}_n^q}.$$

La probabilité d'avoir au moins un billet gagnant est donc

$$1-\mathrm{P}([\mathrm{G}=0]) = 1-\frac{\mathrm{C}_{n-p}^q}{\mathrm{C}_n^q}.$$

Avec les données numériques, ce nombre est supérieur à 1/2 pour un nombre q de billets supérieur ou égal à 23.

Pour obtenir l'espérance de G, on peut utiliser l'expression ci-dessus de la loi et la formule de convolution, mais il est plus élégant de faire le raisonnement probabiliste suivant :

L'espérance du nombre de billets gagnants est la somme des espérances des q variables aléatoires valant 1 si le i-ème billet est gagnant et 0 sinon. Or ces espérances sont d'après la propriété 1 égales à la probabilité qu'un billet soit gagnant, soit $\frac{p}{n}$; l'espérance du nombre de billets gagnants est donc $\frac{pq}{n}$; avec les données numériques, ce nombre est supérieur à 1 pour $q \geq 34$. Ce résultat est en fait très intuitif : il y a, en moyenne dans les q billets que j'ai achetés, la même proportion de billets gagnants que la proportion globale de billets gagnants.

La suite est un peu décevante : mon espérance de gain est égale à :

$$\frac{pq}{n}r - qs = \frac{q}{n}(pr-ns).$$

Il n'y a donc pas de miracle : soit $ns < pr$, ce qui signifie que la loterie est à perte (le coût des lots est supérieur au rapport des billets) et plus j'achète de billets, plus je gagne, soit $ns \geq pr$ (cas normal !) et plus j'achète de billets, plus je perds, en moyenne…

3) Le jeu de Pierre Noir.

Ce jeu est plutôt connu en Suisse et dans les pays germaniques, mais il est très similaire au pouilleux. C'est un jeu constitué de n (= 15) familles de deux cartes ; si l'on joue à q (= 3) joueurs, quelle est la probabilité d'avoir déjà une famille juste après la distribution, et quelle est l'espérance du nombre de familles ?

Solution.

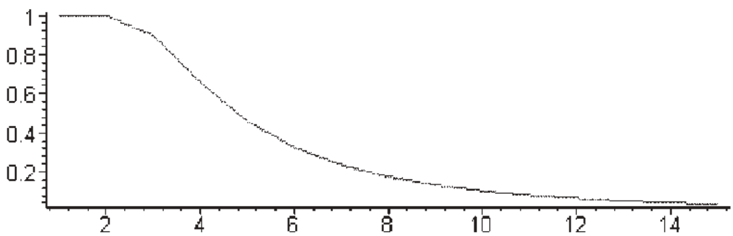
Après la distribution, chaque joueur a reçu $\frac{2n}{q}$ cartes (et on va supposer que q divise

$2n$). Le nombre de mains possibles est C_{2n}^p et le nombre de mains sans famille vaut $X_n^p = C_{2n}^p 2^p$ (je choisis d'abord p familles parmi n et ensuite une carte dans chaque famille) ; la probabilité d'avoir déjà une famille vaut donc

$$1-\frac{C_{2n}^p 2^p}{C_{2n}^p} = 1-\prod_{k=1}^{p-1}\frac{n-k}{n-\frac{k}{2}}.$$

Pour $n = 15$, voici comment varient ces probabilités en fonction du nombre de joueurs :

n:=15:p:=2*n/q:
plot([seq([q,1-2^p*binomial(n,p)/binomial(2*n,p)],q=1..n)]) ;



La probabilité d'avoir une famille est donc plus grande que 1/2 jusqu'à 4 joueurs.

Cherchons maintenant la probabilité d'avoir k familles après la donne.

Le nombre de choix de k familles parmi les n vaut C_n^k et il me reste à choisir parmi les $n-k$ familles restantes $p-2k$ cartes sans qu'il y ait de famille, soit :

$$X_{n-k}^{p-2k} = C_{n-k}^{p-2k} 2^{p-2k}.$$

La probabilité d'avoir k familles vaut donc :

$$\frac{2^{p-2k} C_n^k C_{n-k}^{p-2k}}{C_{2n}^p}.$$

L'espérance du nombre de familles vaut donc :

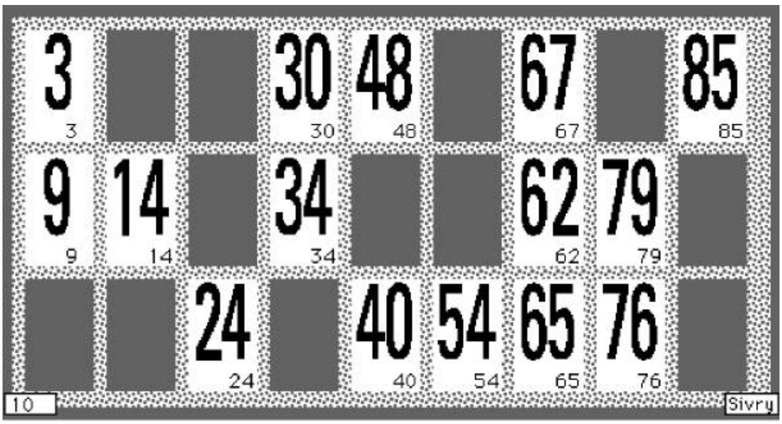
$$\frac{2^p}{C_{2n}^p}\sum_{k=0}^{[p/2]} k\frac{C_n^k C_{n-k}^{p-2k}}{2^{2k}}.$$

Là encore, un raisonnement direct (dû à Jean Moreau de Saint Martin) donne une expression bien plus sympathique !

Quand je prends deux cartes au hasard, la probabilité qu'elles forment une famille est $\frac{1}{2n-1}$: en effet, ayant pris la première carte, la carte qui complète la famille est une parmi $2n-1$ équiprobables.

Quand j'ai p cartes en main, elles forment $\frac{p(p-1)}{2}$ paires dont chacune a la probabilité $\frac{1}{2n-1}$ d'être une famille. Or l'espérance du nombre de familles est la somme des espérances des $\frac{p(p-1)}{2}$ variables aléatoires associées à chacune de ces paires, valant 1 si cette paire est une famille et 0 sinon. D'après la propriété 1, chacune a une espérance égale à $\frac{1}{2n-1}$, donc l'espérance du nombre de familles vaut $\frac{p(p-1)}{2(2n-1)}$; cela donne une espérance de 1,6 familles après la donne, pour un maximum de 5 familles possibles.

4) Le remplissage d'une carte de loto.



Le jeu de loto traditionnel se joue avec des cartes ayant trois lignes remplies chacune avec 5 nombres compris entre 1 et 90. Lorsque l'on tire successivement les nombres de 1 à $n = 90$, au bout de combien de coups en moyenne l'une de mes lignes sera-t-elle remplie, et au bout de combien de coups la carte entière ?

Je crois que ce problème possède autant de méthodes de résolutions qu'il y a de personnes capables de le résoudre…

En voici une première :

Raisonnons d'abord pour les $p = 15$ nombres de la carte entière ; la loi du nombre X de pions reçus sur ma carte après que le k-ième nombre ait été tiré est de nouveau hypergéométrique :

$$P([X = i]) = \frac{C_p^i \, C_{n-p}^{k-i}}{C_n^k}$$

et la probabilité que ma carte soit remplie vaut donc :

$$P([X = p]) = \frac{C_{n-p}^{k-p}}{C_n^k},$$

expression que l'on peut arranger en $\dfrac{C_k^p}{C_n^p}$. Mais ce que je cherche, c'est la probabilité d'avoir mes p pions *pour la première fois* à l'étape k ; or l'événement « avoir sa carte remplie à l'étape k » est la réunion disjointe des événements « avoir sa carte déjà remplie à l'étape $k-1$ » et « avoir sa carte remplie pour la première fois à l'étape k ».

La probabilité de remplir sa carte pour la première fois à l'étape k vaut donc :

$$\frac{C_{k-1}^p}{C_n^p} - \frac{C_k^p}{C_n^p} = \frac{C_{k-1}^{p-1}}{C_n^p}.$$

L'espérance du rang de remplissage de la carte vaut donc :

$$\sum_{k=p}^{n} k\frac{C_{k-1}^{p-1}}{C_n^p} = \frac{1}{C_n^p}\sum_{k=p}^{n} k\,C_{k-1}^{p-1} = \frac{p}{C_n^p}\sum_{k=p}^{n} C_k^p = p\frac{C_{n+1}^{p+1}}{C_n^p} = \frac{p}{p+1}(n+1),$$

soit, numériquement, 84 coups pour la carte complète (c'est très long !) et 75 coups pour une ligne de 5 nombres.
On peut aussi calculer la variance :

$$V = \frac{(n+1)(n-p)}{(p+1)^2(p+2)},$$

ce qui donne un écart-type de 14 coups pour une ligne et de 5 coups pour la carte.

Mais pour obtenir cette espérance, il est beaucoup plus élégant de tenir le raisonnement suivant (qu'a tenu Jean Moreau de St Martin dans un article de Quadrature de 1991) :

Le temps d'attente d'une nouvelle pièce est ici constant en moyenne. Or il y a $p+1$ laps de temps où les $n-p$ numéros non inscrits sur ma carte sortent : avant le premier de mes numéros à sortir, entre le premier et le deuxième, …, et après le dernier. D'après la remarque ci-dessus, ces laps de temps sont égaux en moyenne ce qui fait que le nombre moyen de mauvais numéros sortant entre deux numéros qui sont sur ma carte vaut $\dfrac{n-p}{p+1}$; l'espérance du rang de remplissage vaut donc :

$$n - \frac{n-p}{p+1} = \frac{p(n+1)}{p+1}.$$

Épilogue : ce n'était pas le but premier, mais nous avons obtenu lors de ce voyage en probabilités un certain nombre de formules non triviales que voici :

$$\sum_{k=1}^{n} k\,S_n^k\,C_p^k = p^n - (p-1)^n \quad \text{(dictionnaires)}$$

$$\sum_{k+\ell=q} k\,C_p^k\,C_{n-p}^{\ell} = p\,C_{n-1}^{q-1} \quad \text{(loterie)}$$

$$\sum_{k=0}^{[p/2]} k \frac{C_n^k \, C_{n-k}^{p-2k}}{2^{2k}} = \frac{p(p-1)}{2n-1} \frac{C_{2n}^p}{2^{p+1}} \quad \text{(Pierre Noir)}$$

$$\sum_{k=p}^{n} k \, C_{k-1}^{p-1} = p \, C_{n+1}^{p+1} \quad \text{(loto)}$$

Si vous avez des problèmes similaires à me proposer, je les rajouterai volontiers à ma collection (rferreol@noos.fr).