

R

ENDEZ-VOUS

P.80 Logique & calcul
 P.86 Art & science
 P.88 Idées de physique
 P.92 Chroniques de l'évolution
 P.96 Science & gastronomie
 P.98 À picorer

LES DURES LOIS DES COLLECTIONS

Compléter sa collection de vignettes est loin d'être simple.
 Les mathématiques apportent des explications inattendues!

L'AUTEUR



JEAN-PAUL DELAHAYE
 professeur émérite
 à l'université de Lille
 et chercheur au
 laboratoire Cristal
 (Centre de recherche
 en informatique, signal
 et automatique de Lille)



Jean-Paul Delahaye
 a récemment publié:
Au-delà du Bitcoin
 (Dunod, 2022).

V

ous lancez un dé jusqu'à ce que chacune des faces apparaisse au moins une fois. Combien vous faut-il, en moyenne, de lancés? Si le dé est truqué, faudra-t-il plus de temps?

Vous examinez les premières décimales de $\pi = 3,14159265358979323846264338327950\dots$ Est-il étonnant que, pour avoir vu chacune des décimales au moins une fois, il soit nécessaire d'attendre jusqu'à la 32^e décimale, première occurrence d'un zéro?

Ces questions sont équivalentes à celles que se pose le collectionneur d'une série de vignettes, par exemple les onze cartes des joueurs de son équipe nationale de foot qu'il trouve une par une dans les paquets de céréales achetés pour son petit déjeuner. Ce problème se nomme «problème du collectionneur de vignettes», en anglais «*coupon collector's problem*».

Dans sa formulation générale, il y a N vignettes acquises les unes après les autres lors d'un tirage au hasard. Le but est d'avoir au moins un exemplaire de chacune, ce qui est difficile, car on tire généralement plusieurs fois les mêmes vignettes et que, plus on s'approche du but, plus la probabilité de tirer celles qui manquent se réduit.

Le problème a été mentionné pour la première fois en 1708 dans l'ouvrage *De Mensura Sortis* (*Sur la mesure du hasard*), écrit par le mathématicien français Abraham de Moivre. Puis le grand mathématicien, astronome et

physicien français Pierre-Simon de Laplace l'a traité page 195 de sa *Théorie analytique des probabilités* de 1812. Dans ces solutions, les probabilités de sortie de chaque vignette sont égales.

En 1954, le mathématicien américain Herman von Schelling a étudié le cas plus difficile avec des probabilités inégales. Six ans plus tard, Donald Newman et Lawrence Shepp ont calculé le temps d'attente pour compléter deux collections de coupons dans le cas des probabilités égales. Plus récemment, d'autres chercheurs ont poursuivi le travail en résolvant de nouvelles questions: si les vignettes arrivent par paquets de k vignettes, si les collectionneurs opèrent des échanges, etc.

DES APPLICATIONS DANS DE NOMBREUX DOMAINES

Le problème du collectionneur de vignettes intervient dans de nombreux domaines, en particulier en biologie, où il est utilisé pour estimer le nombre d'espèces animales sachant qu'un grand nombre d'entre elles restent encore inconnues, et modéliser la diffusion d'infections; il sert aussi en chimie combinatoire, ou dans l'industrie pour les dispositifs de contrôle de qualité. En démographie, le souhait de certains couples d'avoir un enfant de chaque sexe (une forme de collection de vignettes!) a un effet mesurable prouvé récemment (*voir l'encadré 3*). Nous verrons aussi que les

LES COLLECTIONNEURS ET LES DÉCIMALES DE π

1

Quiconque se procure un par un au hasard des objets d'une série de N objets (timbres, images de footballeurs, cartes Pokémon comme ci-contre, etc.) tombe malheureusement plusieurs fois sur les mêmes, ce qui l'oblige à acheter plus que N objets pour obtenir la collection complète. Combien doit-il en acheter en moyenne ? Ce problème intéresse les mathématiciens depuis trois siècles. Si on suppose qu'à chaque nouvel achat la probabilité de chaque vignette est $1/N$ (cas équiprobable), la théorie nous indique que le temps d'attente moyen est donné par la formule :

$$T_N = N \times (1 + 1/2 + 1/3 + 1/4 + \dots + 1/N)$$

Ainsi : $T_2 = 3$; $T_3 = 5,5$; $T_4 = 8,333$; $T_5 = 11,42$; $T_6 = 14,7$; $T_7 = 18,15$; $T_8 = 21,74$; $T_9 = 25,46$; $T_{10} = 29,29$

D'autre part, on pense, sans savoir le démontrer, que les chiffres du nombre π dans n'importe quelle base de numération se comportent comme des chiffres tirés au hasard (donc comme des vignettes tirées au hasard). Si c'est le cas, le temps d'attente pour que chaque chiffre apparaisse au moins une fois ne doit pas différer sensiblement de ce que la théorie indique. Les résultats sont donnés dans le tableau ci-contre. Cette petite expérience est concluante, car les écarts avec la théorie sont limités. Ce résultat ne prouve pas que π est un nombre dit « normal », c'est-à-dire que dans sa suite infinie de décimales, les chiffres et les séquences finies de chiffres de ses décimales apparaissent avec la même fréquence, mais il en renforce l'hypothèse.



Base	Décimales de π	Attente	Théorie
2	11,00100100001111110101010100010001000010110100	3	3,0
3	10,01021101222201021100211111022122220111201212	4	5,5
4	3,0210033312222020201122030020310301030121202202	6	8,3
5	3,0323221430334324112412240414023142111430203100	7	11,4
6	3,0503300514151241052344140531253211023012144420	14	14,7
7	3,0663651432036134110263402244652226643520650241	10	18,1
8	3,1103755242102643021514230630505600670163211220	14	21,7
9	3,124188124074427886451776173103582851654535346	28	25,5
10	3,1415926535897932384626433832795028841971693993	32	29,3

calculateurs quantiques sont utiles pour diminuer le temps de calcul de certains algorithmes d'apprentissage directement liés au problème du collectionneur de vignettes.

LE CAS DE DEUX VIGNETTES

Commençons par le cas de deux vignettes en nous demandant combien de tirages en moyenne il faut faire pour avoir les deux vignettes de la collection. Nous allons voir que ce temps d'attente moyen est 3. Cela signifie que les parents qui veulent avoir au moins un fils et une fille doivent en moyenne attendre d'avoir trois enfants pour réussir, et que le coût moyen d'acquisition d'une collection de deux vignettes qui sont vendues au hasard équitablement 1 euro chacune, est de 3 euros.

Le raisonnement est assez simple dans le cas où les deux vignettes A et B sont équiprobables. Si on tire deux vignettes: une fois sur

deux, on en aura deux différentes car les quatre tirages possibles de deux vignettes, AA, AB, BA et BB, ont chacun une probabilité de $1/4$.

Lorsqu'on n'a pas réussi avec les deux premiers tirages, cas AA et BB, en tirant une vignette de plus on obtiendra ce qu'on attend dans un cas sur deux, car si on avait AA, on a une chance sur deux d'avoir le B qui manque, et si on avait BB, on a une chance sur deux d'avoir le A qui manque. Lorsqu'on n'a toujours pas réussi avec les trois premiers tirages, cas AAA et BBB, on a encore une chance sur deux de réussir au quatrième tirage, et c'est la même chose pour chaque tirage supplémentaire.

Le temps d'attente est donc 2 avec une probabilité de $1/2$; 3 avec une probabilité de $1/4$; 4 avec une probabilité de $1/8$, etc. Le temps d'attente moyen pour avoir deux vignettes différentes est la moyenne des temps d'attente en prenant les pondérations données par les

probabilités de chaque cas. Il est donc donné par la formule:

$$T = 2 \times (1/2) + 3 \times (1/4) + 4 \times (1/8) + 5 \times (1/16) + \dots + n \times (1/2^{n-1}) + \dots$$

On décompose chaque terme de la somme et on dispose les 1/2, 1/4, 1/8, etc., dans un tableau:

$$\begin{array}{r} 1/2 + 1/4 + 1/8 + 1/16 + 1/32 + 1/64 + \dots \\ 1/2 + 1/4 + 1/8 + 1/16 + 1/32 + 1/64 + \dots \\ 1/4 + 1/8 + 1/16 + 1/32 + 1/64 + \dots \\ 1/8 + 1/16 + 1/32 + 1/64 + \dots \\ 1/16 + 1/32 + 1/64 + \dots \\ 1/32 + 1/64 + \dots \\ \dots \end{array}$$

On remarque (ou on se souvient!) alors que $1/2 + 1/4 = 1 - 1/4$, $1/2 + 1/4 + 1/8 = 1 - 1/8$, ..., $1/2 + 1/4 + \dots + 1/2^n = 1 - 1/2^n$, ce qui, en faisant tendre n vers l'infini, signifie que la somme $1/2 + 1/4 + \dots + 1/2^n + \dots$ vaut exactement 1.

La somme des termes de la première ligne du tableau est donc 1, celle de la seconde est encore 1, celle de la troisième est 1/2, car c'est la ligne au-dessus privée de 1/2; celle de la quatrième est 1/4, etc. En tout, la somme des termes du tableau est donc $1 + 1 + 1/2 + 1/4 + 1/8 + \dots = 3$.

Une « preuve sans mots » de la formule est proposée dans l'encadré 3 où l'on précise le

problème des couples qui veulent absolument avoir un enfant de chaque sexe.

Dans le cas général avec N vignettes équiprobables, la théorie des probabilités indique que le temps d'attente moyen d'un collectionneur désirant une collection des N vignettes est $T_N = N(1 + 1/2 + 1/3 + 1/4 + \dots + 1/N)$. Cette valeur T_N est aussi le coût moyen de la collection lorsque chaque vignette achetée au hasard coûte 1 euro.

Il existe une formule approchée de T_N : $T_N \approx N \log(N) + \gamma N$ où γ est la constante d'Euler-Mascheroni dont la valeur est 0,5772156649... Bien que beaucoup moins populaire que π ou le nombre d'or, cette constante fait l'objet d'un livre entier du mathématicien britannique Julian Havil (*Gamma: Exploring Euler's Constant*, Princeton University Press, 2003). On trouvera par exemple dans cet ouvrage l'information que même s'il n'a pas été possible jusqu'à aujourd'hui de savoir si γ est ou non un quotient de deux nombres entiers, on a cependant démontré que si γ est le quotient de deux entiers a/b , alors b comporte plus de 242080 chiffres!

La série $1 + 1/2 + 1/3 + 1/4 + \dots + 1/N$ (dénommée « série harmonique ») tend vers l'infini.

2

L'ATTENTE POUR COMPLÉTER UNE COLLECTION

Supposons que chacune des N vignettes d'une collection qu'on cherche à constituer possède la même probabilité d'être obtenue à chaque fois qu'on s'en procure une de plus. Plus on a de vignettes, plus grande est la quantité de vignettes différentes qu'on possède.

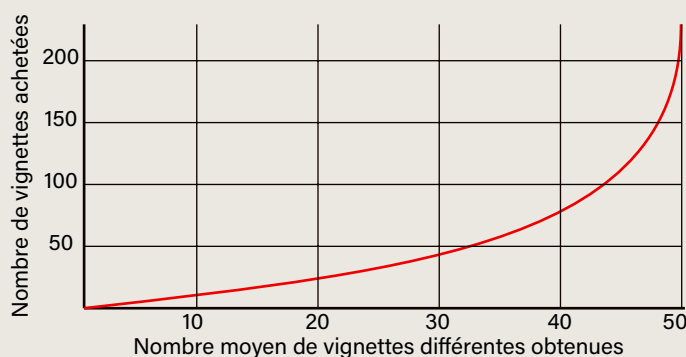
Cependant, plus on approche du but qui est d'avoir N vignettes différentes, plus il devient difficile d'en trouver une que l'on n'a pas déjà obtenue. C'est vrai, même si l'organisateur du jeu met en circulation des quantités égales de chaque sorte de vignette et qu'elles sont donc équiprobables à chaque nouvelle acquisition.

Le nombre de vignettes T qu'il faut acquérir pour en avoir M différentes, même quand $M < N$, augmente donc plus que linéairement en fonction de M .

C'est ce que montre la courbe ci-dessous calculée pour une collection de $N = 50$ vignettes.

On voit que pour réussir à avoir 20 vignettes différentes parmi les 50, il faut en moyenne en acquérir 25,3 ; que pour en avoir 30 différentes, il en faut en moyenne 45,2 ; que pour 40 différentes, il en faut en moyenne 78,7 ; que pour 45 différentes, il en faut en moyenne 110. Pour les avoir toutes, il faut en moyenne en avoir acquis 225.

Une autre question naturelle est : combien faut-il de vignettes en moyenne pour avoir la moitié de la collection (coût de la première moitié) et alors combien, en moyenne, en faudra-t-il en plus pour avoir la seconde moitié (coût de la seconde moitié) ? Le tableau donne les réponses, et indique aussi dans la dernière colonne le nombre de vignettes qu'il faut encore se procurer quand il n'en manque que trois.



Nombre de vignettes N	Coût collection	Coût collection première moitié	Coût collection seconde moitié	Coût des 3 dernières
2	3	1,00	2,00	/
4	8,34	2,34	6,00	7,34
6	14,69	3,70	10,99	10,99
8	21,74	5,07	16,67	14,66
10	29,26	6,45	22,81	18,32
12	37,20	7,84	29,36	21,96
14	45,47	9,22	36,25	25,62
16	54,07	10,61	43,46	29,33
18	63,89	12,00	51,89	33,00
20	71,89	13,39	58,50	36,58
100	518,07	68,79	449,28	184,62

En conséquence, pour obtenir la collection complète de vignettes, il faut, pour certaines collections, en moyenne acheter au moins deux fois plus de vignettes qu'il y en a de différentes (c'est vrai dès que N dépasse 3); pour certaines collections, il faut en moyenne acheter trois fois plus de vignettes qu'il y en a de différentes (c'est vrai dès que N dépasse 10), etc. Cette propriété explique sans doute pourquoi les collectionneurs ont souvent l'impression qu'ils ont beaucoup de mal à arriver au bout; ils soupçonnent, même quand c'est faux, que l'organisateur du jeu met volontairement en circulation moins d'exemplaires de certaines vignettes. Soyez rassuré, la série $1 + 1/2 + 1/3 + 1/4 + \dots + 1/N$ va vers l'infini assez lentement et ce n'est, par exemple, que pour les collections de plus de 12367 vignettes différentes qu'il faut en moyenne en acheter dix fois plus (donc 123670) pour compléter la collection.

La variance des valeurs qui donnent T_N est une mesure de la dispersion autour la moyenne. Elle a été calculée et elle est inférieure à $2N^2$.

La connaissance de T_N permet de répondre à certaines des questions posées au début de l'article. Quand on lance un dé (non truqué) à six faces, le temps d'attente moyen avant d'avoir vu chaque face au moins une fois est de 14,7 lancers. Sachant que $T_{10} = 29,3$, il n'est pas vraiment étonnant de devoir attendre la décimale 32 du nombre π pour avoir vu chaque chiffre au moins une fois (voir l'encadré 1). Le collectionneur des vignettes des 11 joueurs de son équipe de football préférée doit s'attendre à collecter en moyenne 33,2 vignettes avant d'avoir les 11 vignettes différentes de l'équipe au complet.

SUIS-JE BIENTÔT ARRIVÉ ?

Une fois qu'on a acquis la moitié des vignettes que l'on recherche, on est assez satisfait même si on se dit qu'on va mettre un peu plus de temps pour la seconde moitié que pour la première moitié. On soupçonne cependant rarement qu'en réalité cette seconde moitié sera vraiment beaucoup plus difficile à acquérir que la première. Le tableau proposé dans l'encadré 2 nous fait comprendre à quel point le surcoût de la seconde moitié est important comparé au coût de la première moitié. Par exemple, les 10 premières vignettes d'une collection de 20 ont un coût moyen de 13,4 vignettes alors que les 10 dernières coûtent 58,5 vignettes en moyenne. Pour une collection de 100 vignettes, on passe d'un coût moyen de 68 pour la première moitié à un coût moyen de 449 pour la seconde moitié, soit une multiplication par presque 7. Plus la collection est grande plus le surcoût de la seconde moitié par rapport à la première moitié est important !

3

L'EFFET SUR LES FAMILLES

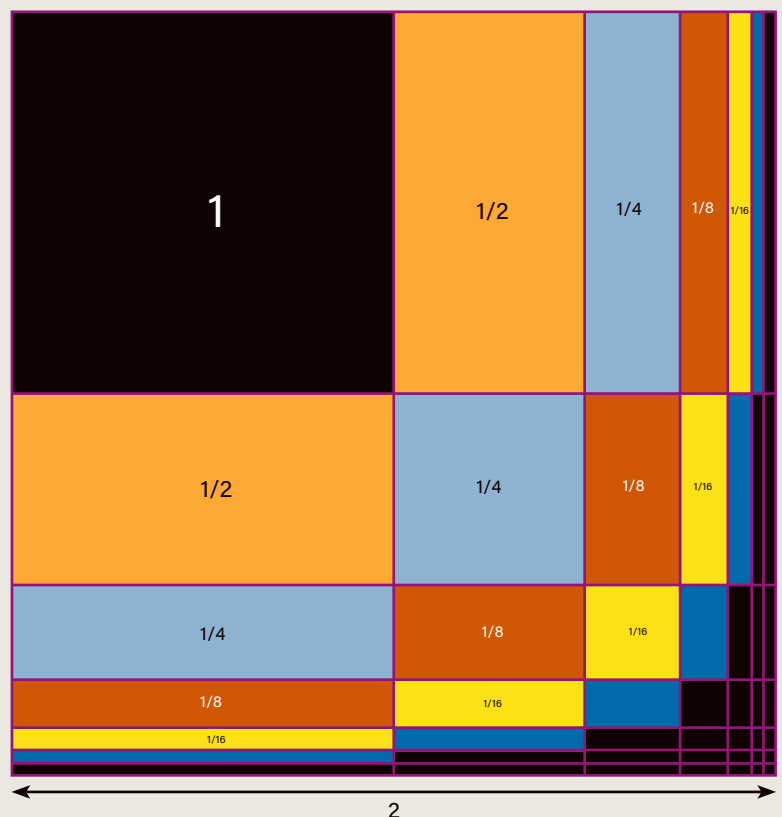
Les couples qui souhaitent avoir au moins un garçon et au moins une fille, et qui s'arrêtent dès qu'ils ont réussi, ont en moyenne trois enfants, car 3 est le temps d'attente moyen pour le collectionneur de vignettes quand la collection en comporte deux (voir l'encadré 1).

Un raisonnement direct, conduisant au nombre 3, est expliqué dans l'article principal. Il conduit à la formule qu'il faut alors simplifier :

$$T_2 = 2 \times (1/2) + 3 \times (1/4) + 4 \times (1/8) + 5 \times (1/16) + \dots + (k+1) \times (1/2^k) + \dots$$

On trouve aussi que $T_2 = 3$ en regardant le dessin ci-dessous qui constitue une « preuve sans mots » de l'égalité. Si vous avez besoin d'une petite aide : le grand carré a une aire de 4 ; quand on enlève le carré noir qui a une aire de 1, ce qui reste a une aire de 3. Or ce reste contient exactement 2 rectangles d'aire $1/2$, 3 rectangles d'aire $1/4$, 4 rectangles d'aire $1/8$, etc. CQFD.

Les démographes se sont posé la question : est-ce que le souhait de certains couples d'avoir un enfant de chaque sexe en continuant à en avoir tant qu'ils ne réussissent pas peut s'observer dans les statistiques ? La réponse est oui. Dans l'article « The coupon collection behavior in human reproduction » (*Current Biology*, 2020), Erping Long et Jianzhi Zhang, de l'université du Michigan, à Ann Harbor, aux États-Unis, ont étudié soigneusement la question à partir des fratries de 300 000 personnes. Ils ont observé qu'un nombre significativement plus élevé que prévu de familles ont tous les enfants du même sexe, à l'exception de l'enfant né en dernier. Il y a, par exemple, plus de familles du type garçon-garçon-fille que de famille garçon-fille-garçon. Cette tendance est plus prononcée dans les données correspondant à des enfants nés dans la décennie 1970 que pour ceux nés dans la décennie 1940, ce qui suggère que la volonté d'avoir au moins un enfant de chaque sexe devient de plus en plus fréquente.



4

LE CAS DE PROBABILITÉS INÉGALES

On cherche de nouveau une collection complète de N vignettes qu'on achète une à une, mais l'organisateur du jeu ne les rend pas disponibles avec la même probabilité. On a mené des simulations dans deux cas particuliers. Dans un cas, on a imaginé que l'une des vignettes est 10 fois moins fréquente que les autres qui sont équiprobables. Dans un second cas, on a supposé que l'une des vignettes est 10 fois plus fréquente que les autres. On observe sans trop de surprise qu'à chaque fois cette inégalité de fréquence augmente le nombre de vignettes qu'il faut se procurer en moyenne pour avoir la collection complète. Le cas d'une vignette 10 fois plus fréquente retarde toujours la constitution de la collection complète, mais il retarde moins le succès que la présence d'une vignette 10 fois plus rare que les autres.

Ces résultats sont conformes à ce qui a été démontré : toute inégalité dans la fréquence des vignettes ralentit l'aboutissement de la collection.

COÛT DE LA COLLECTION

Nombre de vignettes	Vignettes équiprobables	Une vignette est 10 fois moins fréquente	Une vignette est 10 fois plus fréquente
2	3	11,1	11,1
3	5,5	21,3	17,9
4	8,3	31,7	23,8
5	11,4	42,2	29,2
6	14,7	52,4	34,3
7	18,1	62,7	39,2
8	21,7	73,3	44,1
9	25,5	84,2	48,9
10	29,3	95,3	53,7
20	72	189,9	103,9
50	225	535,8	261,4
100	518,7	1022,8	555,7

Il se peut cependant que les probabilités de tomber sur les différentes vignettes soient inégales. Quelle que soit cette inégalité entre vignettes, on démontre qu'elle aura pour effet d'allonger le temps moyen nécessaire à l'acquisition de la collection complète. Cela répond à la question posée dans le premier paragraphe de l'article: quelle que soit la façon de truquer un dé, le nombre moyen de lancés nécessaire pour faire apparaître les six faces augmente et dépasse alors 14,7 lancés.

Un programme de calcul m'a permis de comparer deux répartitions inégales des probabilités entre vignettes. J'ai d'abord testé le cas où l'une des vignettes est 10 fois plus rare que les autres qui, elles, ont la même probabilité d'être tirées. J'ai ensuite testé l'idée que l'une des vignettes avait 10 fois plus de chance d'être tirée que les autres, et donc qu'un collectionneur serait encombré par cette vignette surreprésentée. Dans les deux cas, le temps moyen pour compléter la collection augmente, mais c'est le premier cas, une vignette rare, qui est le pire (voir l'encadré 4).

LA RARETÉ PÉNALISE DAVANTAGE

Si, par exemple, il y a 10 vignettes dont l'une est 10 fois plus rare que les autres, le coût (ou temps moyen pour avoir la collection) est de 95,3 (contre 29,3 pour 10 vignettes équiprobables), alors que si, parmi les 10 vignettes, il y en a une qui est 10 fois plus fréquente que les autres, le coût de la collection est de 53,7. Les vignettes rares sont plus pénalisantes que les vignettes fréquentes, mais les vignettes surreprésentées le sont aussi.

Une formule simple indique le nombre moyen de vignettes différentes qu'on obtient

après en avoir acheté m quand il y a N vignettes dont les probabilités de tirages sont respectivement $p(1), p(2), \dots, p(N)$.

$$\sum_{k=1}^N (1 - (1 - p(k))^m)$$

Cette formule est utile même quand les vignettes ont la même probabilité: elle se simplifie alors en $N(1 - (1 - 1/N)^m)$. Par exemple, au bout de 50 achats, quand il y a 100 vignettes équiprobables, on a en moyenne obtenu $50 \times (1 - (99/100)^{100}) = 39,5$ vignettes différentes.

LE COLLECTIONNEUR QUANTIQUE DE VIGNETTES

Depuis 2020, plusieurs chercheurs se sont intéressés à l'avantage que donneraient des calculateurs quantiques dans la situation où un collectionneur classique d'informations se confronte à un collectionneur quantique d'informations. Nous allons examiner ce que cela signifie.

L'un des acteurs clés de cette nouvelle recherche se nomme Srinivasan Arunachalam. Il travaille au centre de recherche IBM Thomas Watson, aux États-Unis. On remarquera qu'il porte le même prénom que le génie mathématique indien universellement admiré, Srinivasan Ramanuja (1887-1920).

Voyons d'abord le cas du collectionneur classique d'informations. On imagine une situation où est fixé un sous-ensemble S de k éléments pris dans un ensemble E de n éléments qu'on supposera être l'ensemble des entiers de 1 à n . Le collectionneur classique d'informations connaît k et n , et cherche à identifier S . On crée k vignettes portant chacune un élément de S . Elles sont mises dans un chapeau. Le collectionneur classique d'informations choisit au

hasard une vignette dans le chapeau. Il en prend connaissance et la remet en place. Il recommence un certain nombre de fois ce tirage au sort équiprobable, comme quand un collectionneur de vignettes en achète de nouvelles jusqu'à avoir la collection complète. Son but est de réussir à connaître quels sont les k éléments de l'ensemble S . Il cherche donc à tirer au moins une fois chacune des k vignettes du chapeau.

Imaginons que $n=10$, $E=\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $k=3$ et $S=\{3, 5, 10\}$. Le collectionneur classique d'informations tire d'abord, par exemple, le 3, puis le 10, puis le 10, puis le 10, puis le 5. Puisqu'il sait que $k=3$, et qu'il a trouvé 3 vignettes différentes, il connaît maintenant la «collection complète» correspondant à S . Il s'arrête.

Le temps moyen pour connaître l'ensemble S de k éléments est approximativement $k \cdot \log(k)$. En effet, le problème est équivalent à celui d'un collectionneur de vignettes qui les achète une à une, chacune ayant la probabilité $1/k$ d'être tirée à chaque nouvel achat, jusqu'à avoir la collection complète des k vignettes, et nous avons vu plus haut que, dans ce cas, le collectionneur réussit en moyenne au bout de $k \cdot \log(k)$ (en ordre de grandeur) vignettes. On exprime ce résultat en disant qu'apprendre S exige en moyenne $k \cdot \log(k)$ tirages pour le collectionneur classique d'informations.

Dans le cas du collectionneur quantique d'informations, un état quantique noté $|S\rangle$ est créé qui superpose chaque élément $|i\rangle$ avec le même poids, ce qui est l'équivalent quantique de mettre k vignettes avec les numéros correspondant à S dans un chapeau.

MANIPULER L'INFORMATION AVEC PRÉCAUTION

Le monde quantique ne permet pas à celui qui dispose de $|S\rangle$ de connaître chacune des composantes $|i\rangle$ de l'état superposé. En effet, en manipulant l'état superposé et en cherchant à en tirer des informations sur les $|i\rangle$ qui le composent, cela détruit $|S\rangle$, ce qui empêche alors d'accéder aux informations que l'on n'a pas réussi à tirer de la manipulation. Le collectionneur quantique d'informations peut néanmoins exploiter l'état superposé $|S\rangle$ de sorte à en extraire une certaine information, sans que cela lui donne complètement $|S\rangle$; la situation est parallèle à celle du cas classique où prendre une vignette dans le chapeau ne donne pas accès à la connaissance complète de S . En recommençant des manipulations à partir de nouvelles copies de l'état superposé $|S\rangle$, le collectionneur quantique d'informations en déduit une connaissance de plus en plus précise de $|S\rangle$. Là encore, la situation est analogue au cas classique; celui qui tire des vignettes du chapeau plusieurs fois et les remet connaît

progressivement de mieux en mieux $|S\rangle$ malgré les tirages en double.

La question est de savoir s'il existe des façons quantiques de procéder, c'est-à-dire de manipuler le vecteur $|S\rangle$, permettant au collectionneur quantique d'informations de connaître $|S\rangle$ en un nombre d'étapes plus petit que les $k \cdot \log(k)$ qui sont nécessaires en moyenne au collectionneur classique d'informations.

La réponse étonnante qui a été proposée est oui. Srinivasan Arunachalam et ses collègues ont conçu une façon quantique de manipuler $|S\rangle$ qui donne accès à la connaissance complète de $|S\rangle$ en opérant en moyenne un nombre de manipulations de l'ordre de $n \cdot \log(n-k+1)$. C'est tout à fait remarquable, car si $n-k$ est petit et fixé, $n \cdot \log(n-k+1)$ est, en ordre de grandeur, égal à $k \cdot \log(n-k)$, ce qui est beaucoup mieux que $k \cdot \log(k)$ puisque le facteur croissant $\log(k)$ est remplacé par le facteur constant $\log(n-k)$. À notre demande, le physicien Jean Dalibard a accepté d'expliquer la situation particulière $k=n-1$. Le document qu'il a rédigé et dont nous le remercions se trouve sous le lien suivant: bit.ly/PLS547_Dalibard

Grâce à l'algorithme quantique de Peter Shor découvert en 1994, on sait factoriser un entier n (c'est-à-dire trouver ses facteurs) en un temps polynomial en fonction de la taille de n , alors qu'on ne sait pas le faire avec un algorithme classique. Le résultat de Srinivasan Arunachalam montre que non seulement le monde quantique est plus performant que le monde classique pour la factorisation, mais qu'il l'est aussi pour le problème de l'acquisition d'informations.

En 2022, Min-Gang Zhou, de l'université de Nankin, en Chine, et une équipe autour de lui ont réalisé des expériences confirmant l'efficacité du protocole théorique de Srinivasan Arunachalam. Cela prouve qu'un collectionneur quantique d'informations peut apprendre plus rapidement des données qu'un collectionneur classique d'informations. C'est une nouvelle démonstration de l'intérêt qu'il y a à faire intervenir la mécanique quantique en informatique. Ces résultats signifient en particulier que les algorithmes d'apprentissage si importants en intelligence artificielle peuvent fonctionner plus efficacement en utilisant des dispositifs quantiques qu'avec nos bons vieux ordinateurs classiques. C'est assez troublant!

L'activité des collectionneurs n'est apparemment qu'un simple jeu de peu d'intérêt, pourtant, quand on l'examine avec le regard d'un mathématicien, elle suggère une quantité considérable de questions qui, plus de trois siècles après les premiers travaux d'Abraham de Moivre, semblent devoir se renouveler encore longtemps. ■

BIBLIOGRAPHIE

Min-Gang Zhou *et al.*, **Experimental quantum advantage with quantum coupon collector**, *Research*, 2022.

S. Arunachalam *et al.*, **Quantum coupon collector**, 15th Conference on the Theory of Quantum Computation, Communication and Cryptography, 2020.

S. Sardy et Y. Velenik, **Petite collection d'informations utiles pour les collectionneurs compulsifs**, *Images des mathématiques*, 2020.

P. Flajolet *et al.*, **Birthday paradox, coupon collectors, caching algorithms and self-organizing search**, *Discrete Appl. Math.*, 1992.

D. Newman et L. Shepp, **The double dixie cup problem**, *The American Mathematical Monthly*, 1960.

H. von Schelling, **Coupon collecting for unequal probabilities**, *The American Mathematical Monthly*, 1954.