

l'expérience a-t-il été une surprise, que les mathématiciens ont essayé d'expliquer tant bien que mal. Une question clef est de savoir si la même approche fonctionnera pour d'autres problèmes.

Tout cela met en lumière le rôle curieusement ambivalent du hasard en informatique. Les algorithmes sont par définition strictement déterministes, et pourtant beaucoup d'entre eux exploitent une petite dose de hasard : la possibilité, de temps à autre, de faire un choix en tirant à pile ou face.

Pseudo-aléatoire ou quasi-aléatoire ?

En pratique, cependant, les nombres aléatoires fournis aux programmes informatiques ne sont jamais vraiment aléatoires. Ils sont pseudo-aléatoires – ce sont des faux astucieux, qui semblent aléatoires et qui passent les tests statistiques, mais qui sont issus d'une source déterministe. Chose curieuse, les faux nombres aléatoires semblent fonctionner très bien, en tout cas pour la plupart des tâches.

Les nombres quasi-aléatoires vont un pas plus loin. Ils ne font même pas l'effort

de la feuille et celle du panneau. Si l'on convient que l'aire du panneau est égale à 1, l'aire estimée de la feuille sera donc égale à n/N .

Lancer des fléchettes au hasard est non seulement difficile, mais aussi dangereux. Mieux vaut remplacer ces projectiles par des points et laisser l'ordinateur se charger de les parsemer au hasard. La mesure est alors facile, et l'expérience fonctionne remarquablement bien.

J'ai cueilli une feuille d'érable, je l'ai photographiée et j'ai placé son image verte sur un champ carré de 1024×1024 pixels blancs. J'ai alors écrit un programme qui jette 1024 points aléatoires (en fait des points pseudo-aléatoires) sur l'image numérisée. La première fois que j'ai fait tourner le programme, 429 des 1024 points sont tombés sur la feuille, d'où une superficie estimée de 0,4189. La superficie réelle, définie par comptage des pixels verts, était de 0,4185. L'approximation des points aléatoires était meilleure que ce à quoi je m'attendais (un coup de chance, mais rien d'extraordinaire cependant). Quand j'ai répété l'expérience un millier de fois, l'estimation moyenne de la superficie de la feuille était

-aléatoires

de se faire passer pour aléatoires. Et pourtant, ils semblent eux aussi très efficaces dans de nombreux cas où l'on fait habituellement appel au hasard. Ils pourraient même faire mieux que leurs cousins pseudo-aléatoires dans certaines circonstances.

Un petit problème amusant aide à préciser les différences entre pseudo-aléatoire et quasi-aléatoire. Supposons que l'on veuille estimer la superficie d'un objet de forme compliquée, une feuille d'érable par exemple (voir les figures 1 et 2). Une astuce bien connue permet de résoudre ce problème avec l'aide d'un peu de hasard. Placez la feuille sur un panneau de superficie connue, et lancez des fléchettes dessus au hasard, sans viser. Si un total de N fléchettes touche le panneau, et que n d'entre elles atterrissent sur la feuille, alors le rapport n/N est approximativement égal au rapport entre la super-

ficie de la feuille et celle du panneau. Si l'on convient que l'aire du panneau est égale à 1, l'aire estimée de la feuille sera donc égale à n/N .

L'idée de base des calculs de Monte Carlo est de reformuler un problème mathématique (tel que calculer la surface d'une feuille) comme un jeu de hasard, où les gains attendus d'un joueur sont les réponses au problème. Dans les jeux simples, on peut calculer la probabilité exacte de chaque résultat, et les gains attendus peuvent donc également être déterminés exactement. Là où de tels calculs sont infaisables, l'alternative consiste à se lancer et jouer au jeu, et voir ce qui ressort. C'est la stratégie d'une simulation de Monte Carlo. L'ordinateur joue au jeu de nombreuses fois, et prend comme estimation de la valeur attendue réelle le résultat moyen.

Pour le problème de la feuille sur le carré, la valeur attendue est le rapport de la superficie de la feuille à celle du carré. En choisissant N points aléatoires et en

L'ESSENTIEL

✓ De nombreux calculs numériques sont réalisés à l'aide d'un échantillonnage aléatoire de points : c'est le principe des méthodes dites de Monte Carlo.

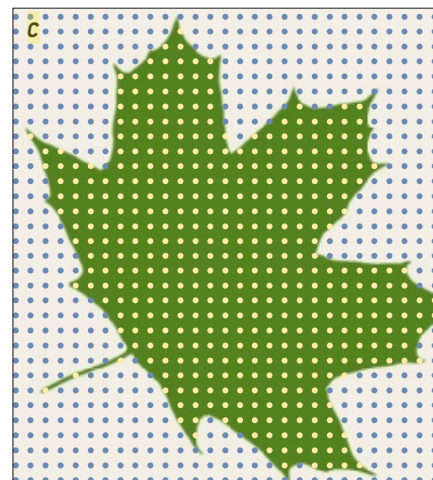
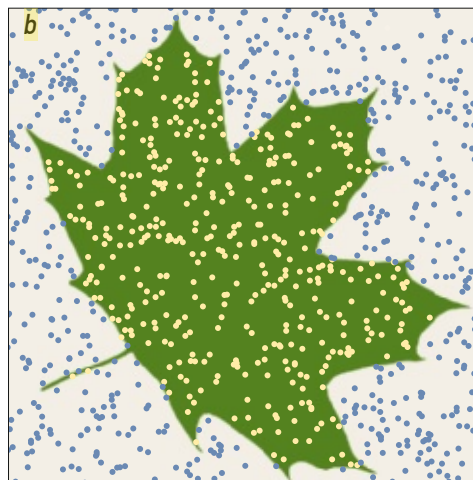
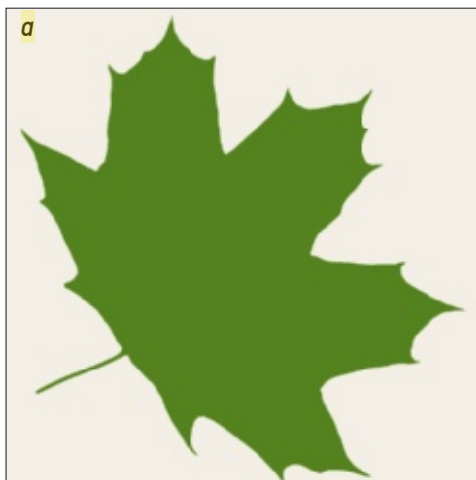
✓ Dans certaines circonstances qui restent à préciser, un échantillonnage plus régulier, dit quasi-aléatoire, conduit à un calcul plus efficace.

✓ Les méthodes de calcul correspondantes, dites méthodes de quasi-Monte Carlo, connaissent aujourd'hui un regain d'intérêt.

© Shutterstock/Kundra : Pour la Science

Brian Hayes

1. POUR CALCULER LE VOLUME
d'une forme irrégulière (en rouge), le principe de la méthode de Monte Carlo est de prendre un échantillon aléatoire de points uniformément distribués dans un domaine plus grand et de volume connu (ici un cube), puis de mesurer la proportion des points se trouvant à l'intérieur du volume recherché. Cette proportion est une approximation du volume relatif, d'autant plus précise que le nombre de points est grand.



2. ESTIMER LA SUPERFICIE d'une forme complexe, ici une feuille d'érable, est une application classique de la méthode de Monte Carlo. L'image de la feuille est inscrite dans un carré dont on peut considérer que la superficie vaut 1 (a). Le principe de base est de saupoudrer des points dans le carré et de compter combien tombent sur la feuille (en jaune) et en dehors (en bleu). Le rapport du nombre de points gagnants sur le nombre total de points donne une approximation de la superficie de la feuille. La technique de Monte Carlo classique (b) éparpille les points au hasard sur le carré. Une alternative totalement déterministe (c) contraint l'agencement des points en les plaçant sur une grille régulière. La méthode de quasi-Monte Carlo (d) utilise un échantillonnage caractéristique qui vise à atteindre une densité uniforme de points, sans être ni aléatoire ni régulière. Avec un échantillonnage de 1 024 points, les trois méthodes donnent respectivement des estimations de la surface de 0,4189, 0,4209 et 0,4141, qui sont toutes à moins de un pour cent de la valeur réelle 0,4185.

comptant le nombre de « coups gagnants », on a une approximation du rapport des superficies, et l'approximation est d'autant meilleure que N est élevé. Quand N tend vers l'infini, la mesure devient exacte. Ce dernier point n'est pas juste une observation empirique, mais aussi un théorème, qui résulte de la loi des grands nombres. C'est ce même principe qui garantit qu'une pièce non truquée tombera sur face dans la moitié des cas si on la lance un grand nombre de fois (échantillon grand).

Le hasard a un rôle clair dans cette description de la méthode de Monte Carlo. En particulier, l'invocation de la loi des grands nombres suppose que les points de l'échantillon soient choisis au hasard. Et il est facile de voir (juste en regardant une image) pourquoi l'échantillonnage au hasard marche bien : il éparpille les points partout. Ce qui n'est pas si facile à voir, c'est pourquoi d'autres types d'arrangement des points-échantillons ne feraient pas aussi bien l'affaire. Après tout, on pourrait mesurer la superficie de la feuille en posant une grille régulière de points sur le carré et en comptant les coups gagnants. J'ai tenté cette expérience avec mon image de feuille, en plaçant 1 024 points sur une grille 32×32 . J'ai obtenu une estimation de 0,4209, pas tout à fait aussi bonne qu'avec mon tirage aléatoire si chanceux, mais qui s'approche tout de même à 0,6 pour cent près de la véritable superficie.

J'ai également essayé de mesurer la feuille avec 1 024 points quasi-aléatoires, dont l'arrangement est, d'une certaine manière, intermédiaire entre le chaos total et l'ordre total (voir l'encadré page 59). L'estimation obtenue à partir du comptage de coups gagnants avec les points quasi-aléa-

toires était 0,4141, ce qui donne une erreur de 1 pour cent.

Chacune de ces trois procédures donne des résultats très corrects. Cela signifie-t-il qu'elles sont également puissantes ? Je pense que non : cela signifie simplement que mesurer la superficie d'une feuille en deux dimensions est un problème facile.

La tâche devient beaucoup plus difficile en dimensions supérieures. Afin de comprendre pourquoi, un petit exercice de raisonnement multidimensionnel s'impose. Imaginez un « cube » à d dimensions, avec des côtés de longueur unité, et à l'intérieur un cube plus petit, dont les côtés ont pour longueur $1/2$. À une dimension, soit $d = 1$, un cube est juste un segment de droite, et le volume est équivalent à la longueur ; ainsi, le plus petit des deux cubes a la moitié du volume du plus gros. Pour $d = 2$, le cube est un carré, et le volume est l'aire ; le petit cube a un volume de $1/4$. Le cas $d = 3$ correspond au cube ordinaire, et le volume rempli par le petit cube vaut maintenant $1/8$.

La malédiction de la dimension

La progression continue. Quand nous atteignons $d = 20$, le petit cube (toujours avec un côté de longueur $1/2$ dans chaque dimension) n'occupe qu'un millionième du volume total. Le mathématicien américain Richard Bellman a qualifié ce phénomène de « malédiction de la dimension ». Si nous voulons mesurer le volume du petit cube en dimension $d = 20$ (ou même simplement détecter sa présence), nous avons besoin de suffisamment de points d'échantillonnage pour obtenir au moins un point