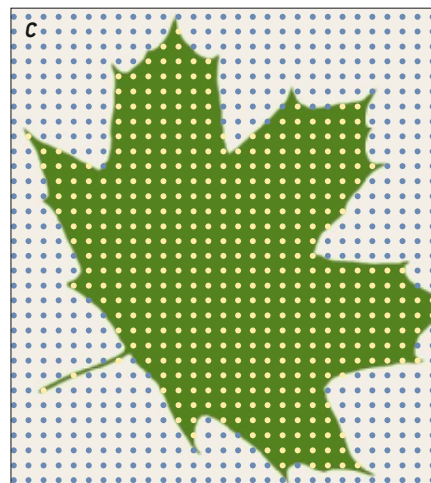
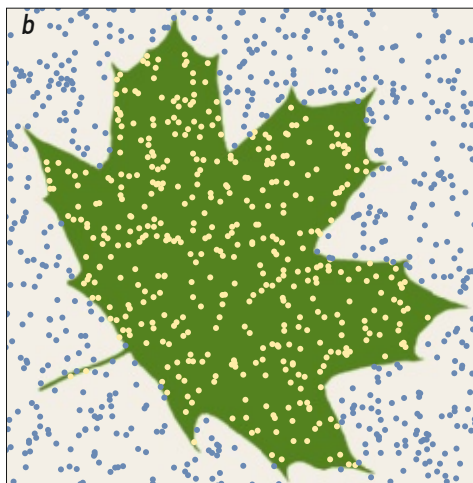
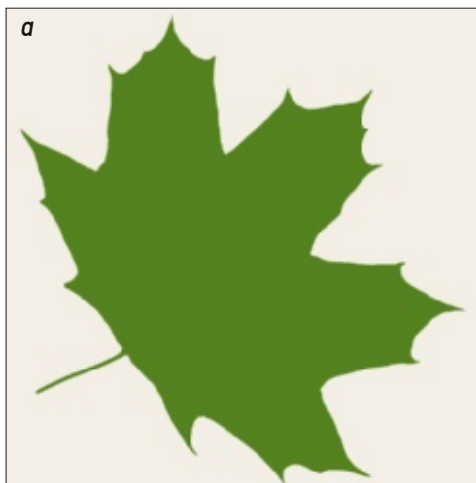




2. ESTIMER LA SUPERFICIE d'une forme complexe, ici une feuille d'érable, est une application classique de la méthode de Monte Carlo. L'image de la feuille est inscrite dans un carré dont on peut considérer que la superficie vaut 1 (*a*). Le principe de base est de saupoudrer des points dans le carré 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-

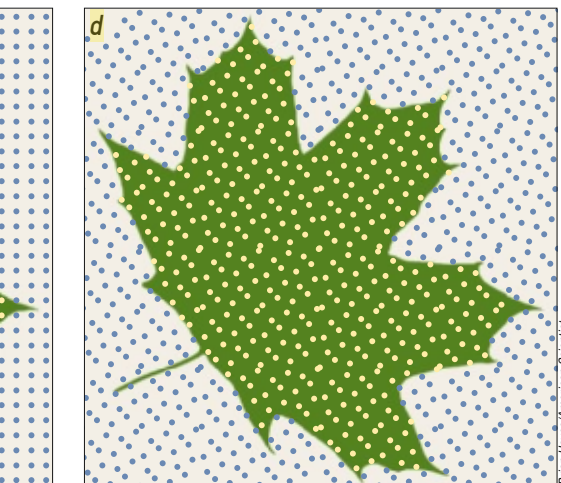
toires était 0,4141, ce qui donne une erreur de 1 pour cent.

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



Brian Hayes/American Scientist

à l'intérieur du petit cube. En d'autres termes, quand nous comptons les coups gagnants, il faut que nous en comptons au moins un. Pour un motif de quadrillage à 20 dimensions, cela signifie que nous avons besoin d'un million de points (ou, plus précisément, de $2^{20} = 1\,048\,576$ points). Avec l'échantillonnage aléatoire, l'exigence de taille est probabiliste et donc un peu floue, mais le nombre de points nécessaires pour nous attendre à un seul coup gagnant est là encore 2^{20} . L'analyse pour l'échantillonnage quasi-aléatoire donne le même résultat. En effet, si vous tâtonnez à l'aveuglette pour trouver un objet de volume $1/2^d$, peu importe l'agencement du motif de recherche; il vous faudra chercher à 2^d endroits.

Prévisibles, corrélés, mais équitables

Si les problèmes réels étaient aussi compliqués que celui-ci, les perspectives seraient assez sombres. Il n'y aurait pas le moindre espoir d'estimer une intégrale à 360 dimensions. Mais nous savons que les techniques de Monte Carlo viennent à bout de certains problèmes de cette taille; c'est une conjecture raisonnable de penser que les problèmes solubles ont une structure interne qui accélère la recherche. De plus, le choix du motif d'échantillonnage semble faire une différence, si bien qu'il y a une distinction significative à faire entre toutes les gradations allant de l'aléatoire vrai à l'ordonné, en passant par le pseudo-aléatoire et le quasi-aléatoire.

L'idée de hasard dans un ensemble de nombres a au moins trois aspects. Tout

d'abord, des nombres choisis au hasard sont imprévisibles: il n'existe pas de règle fixe présidant à leur sélection. Deuxièmement, ces nombres sont indépendants, ou non corrélés: le fait de connaître l'un des nombres ne vous aidera pas à en deviner un autre. Enfin, les nombres aléatoires sont non biaisés, ou uniformément distribués: peu importe comment vous découpez l'espace des valeurs possibles, chaque région peut s'attendre à en avoir sa part.

Ces concepts fournissent une clef utile pour distinguer entre les ensembles véritablement aléatoires, pseudo-aléatoires, quasi-aléatoires et ordonnés. Les nombres véritablement aléatoires présentent les trois caractéristiques énoncées: ils sont imprévisibles, non corrélés et non biaisés. Les nombres pseudo-aléatoires abandonnent l'imprévisibilité; ils sont engendrés par une règle arithmétique bien définie, et si vous connaissez la règle, vous pouvez reproduire la séquence entière. Mais les nombres pseudo-aléatoires restent non corrélés et non biaisés (au moins avec une bonne approximation).

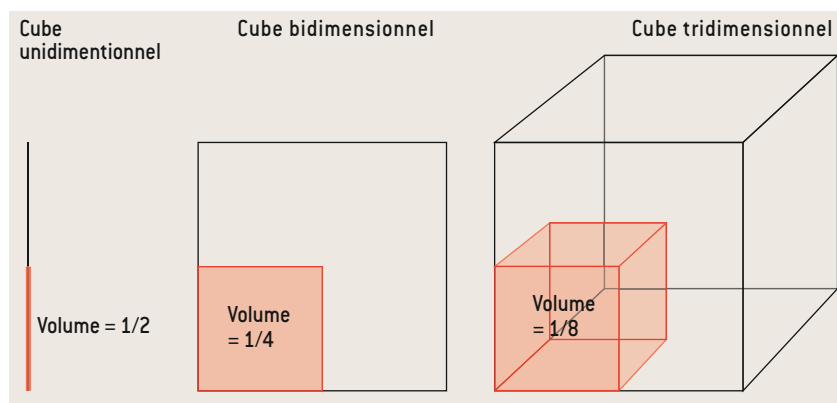
Les nombres quasi-aléatoires sont à la fois prévisibles et fortement corrélés. Il existe une règle bien définie pour les engendrer, et les motifs qu'ils forment, bien que pas aussi stricts que ceux d'un réseau cristallin, présentent une grande régularité. Par rapport au hasard, le seul élément que les nombres quasi-aléatoires préservent est la distribution uniforme ou équitable: ils sont distribués aussi équitablement et également que possible.

Un ensemble très ordonné, comme les points d'un réseau cubique, ne préserve aucune des propriétés du hasard. Il est évident que ces points échouent aux tests d'imprévisibilité et d'indépendance. Il n'est peut-être pas aussi clair que leur distribution n'est pas uniforme. Après tout, il est possible de découper un réseau de N points en N petits cubes, qui contiennent chacun exactement un point. Mais cela ne suffit pas pour qualifier le réseau d'équitabilité et également distribué. L'idéal de la distribution équitable exige un arrangement de N points qui fournit un point par région quand l'espace est divisé en n'importe quel ensemble de N régions identiques. Le réseau rectiligne échoue à ce test quand les régions sont des tranches parallèles aux axes.

Un léger désaccord: la discrédance

Ces trois aspects du hasard sont importants dans différents contextes. Dans le cas de l'autre Monte Carlo (le casino de Monaco), l'imprévisibilité est tout. Il en va de même pour les applications cryptographiques du hasard. Dans les deux cas, un adversaire essaie activement de détecter et d'exploiter toute ébauche de motif.

Certains types de simulations informatiques sont très sensibles aux corrélations entre des nombres aléatoires successifs, si bien que l'indépendance est importante. Mais pour les estimations de volume que nous discutons ici, c'est l'uniformité de la



3. L'ESTIMATION DES VOLUMES devient de plus en plus difficile à mesure que le nombre de dimensions de l'espace augmente, effet qualifié parfois de « malédiction de la dimension ». Supposons qu'un cube à d dimensions, dont les côtés ont pour longueur 1, renferme un cube de côtés $1/2$. Quand $d = 1$, le « cube » est en fait un segment de droite, et le volume est sa longueur; le petit « cube » [de longueur $1/2$] remplit la moitié du volume total. Pour $d = 2$ (un carré, pour lequel le volume correspond à la superficie), le volume du petit cube est de $1/4$, et, pour $d = 3$, il est égal à $1/8$. Ainsi, le volume fractionnaire diminue comme $1/2^d$, ce qui signifie que pour simplement détecter l'existence du petit cube, il faut déjà environ 2^d points d'échantillonnage.