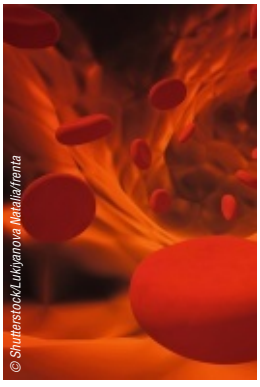
INCOMPATIBILITÉ SANGUINE ET GROSSESSE

L'incompatibilité sanguine se manifeste pendant une grossesse, quand les globules rouges de la mère ne portent pas les mêmes antigènes que ceux de son fœtus. Les antigènes de globules rouges sont au nombre de trois, A, B et D (le marqueur Rhésus).

Prenons l'exemple de l'incompatibilité Rhésus, la plus grave. Si une mère a un Rhésus négatif (15 pour cent des individus), ses globules rouges ne portent pas l'antigène D. Or si elle a un enfant avec un père de Rhésus positif, le fœtus peut avoir le même Rhésus que le père. Alors les globules rouges du fœtus – portant l'antigène D – sont considérés comme étrangers par l'organisme de la mère, qui peut produire des anticorps contre les globules rouges de l'enfant. Heureusement, peu de globules rouges fœtaux passent dans le sang de la mère au niveau du placenta. Toutefois, pendant le travail de l'accouchement et la délivrance, des globules rouges fœtaux peuvent circuler dans le sang de la mère, qui s'immunise donc contre les antigènes D et fabrique des anticorps anti-antigène D.

Lors d'une deuxième grossesse, si le fœtus est aussi Rhésus positif, le système immunitaire de la mère reconnaît le sang du fœtus comme étranger et peut détruire ses globules rouges. Ce qui engendre des pathologies graves.


© Shutterstock/Lukjanova Natalia/frenta

Mais en général, dans les 72 heures qui ont suivi le premier accouchement, on a vérifié le sang de la mère et on lui a injecté des anticorps anti-antigène D ; ces derniers détruisent les globules rouges fœtaux qui seraient passés dans son sang. Ainsi, la mère n'a pas eu le temps de développer une réaction immunitaire, et sa deuxième grossesse se déroule comme la première. L'incompatibilité est semblable avec les autres antigènes sanguins, par exemple quand une mère de groupe O (ses globules rouges ne portent ni l'antigène A ni l'antigène B) attend un enfant de groupe A ou B.

de cellules provenant d'un autre organisme. La grossesse est la principale circonstance où apparaît un microchimérisme. Un double trafic s'installe : de la mère vers le fœtus et du fœtus vers la mère. Et cela se produit sans que le système immunitaire de l'un ne réagisse aux cellules de l'autre.

Pour l'étudier, nous avons choisi une pathologie de la femme enceinte, les éruptions polymorphes de la grossesse (qui concernent environ quatre pour cent des femmes enceintes). Des dermatoses, c'est-à-dire des lésions de la peau, se développent sur l'abdomen autour du nombril durant le troisième trimestre de la grossesse, période où le microchimérisme fœtal est le plus important. On ignore les mécanismes de ces éruptions.

En 1998, nous avons trouvé des cellules fœtales dans des prélèvements de lésions cutanées de femmes enceintes. Puis nous avons cherché l'expression des molécules HLA-G dans la peau de ces patientes, dans celle de femmes enceintes non atteintes, et dans la peau de femmes non enceintes. La peau des femmes ne contient des HLA-G que pendant la grossesse.

Des cellules fœtales exprimant les antigènes HLA-G sont donc présentes dans la peau de femmes enceintes et il y en a sans doute ailleurs dans l'organisme de la mère. On sait même que ces cellules existent encore dans la peau de la mère 27 ans après l'accouchement. La mère conserve toute sa vie des cellules de ses enfants.

Qui plus est, des acteurs autres que les HLA-G participeraient à la tolérance immunitaire qui s'établit entre la mère et le fœtus. Par exemple, des lymphocytes *T* régulateurs apparaissent chez le fœtus quelques semaines après la fécondation ; ces derniers inhibent des cellules tueuses fœtales qui détruiraient les cellules de la mère circulant dans le fœtus. Toutefois, ces facteurs n'apparaissent qu'à une certaine période de la grossesse et pendant un temps limité ; ils ne sont pas indispensables à la grossesse, car leur absence ne provoque pas forcément d'avortement. Seuls les HLA-G sont nécessaires au bon déroulement d'une grossesse pendant neuf mois.

Ainsi, la communauté internationale s'accorde sur le rôle clé des HLA-G dans la tolérance fœto-maternelle. Ils mettent en veille certaines réactions immunitaires – les réactions de rejet – et permettent ainsi à la mère de tolérer l'enfant qu'elle porte. En plus de leurs propriétés inhibitrices vis-à-vis des cellules immunitaires, les HLA-G participent à la production de lymphocytes *T* régulateurs qui mettent en place un environnement de tolérance pendant la grossesse.

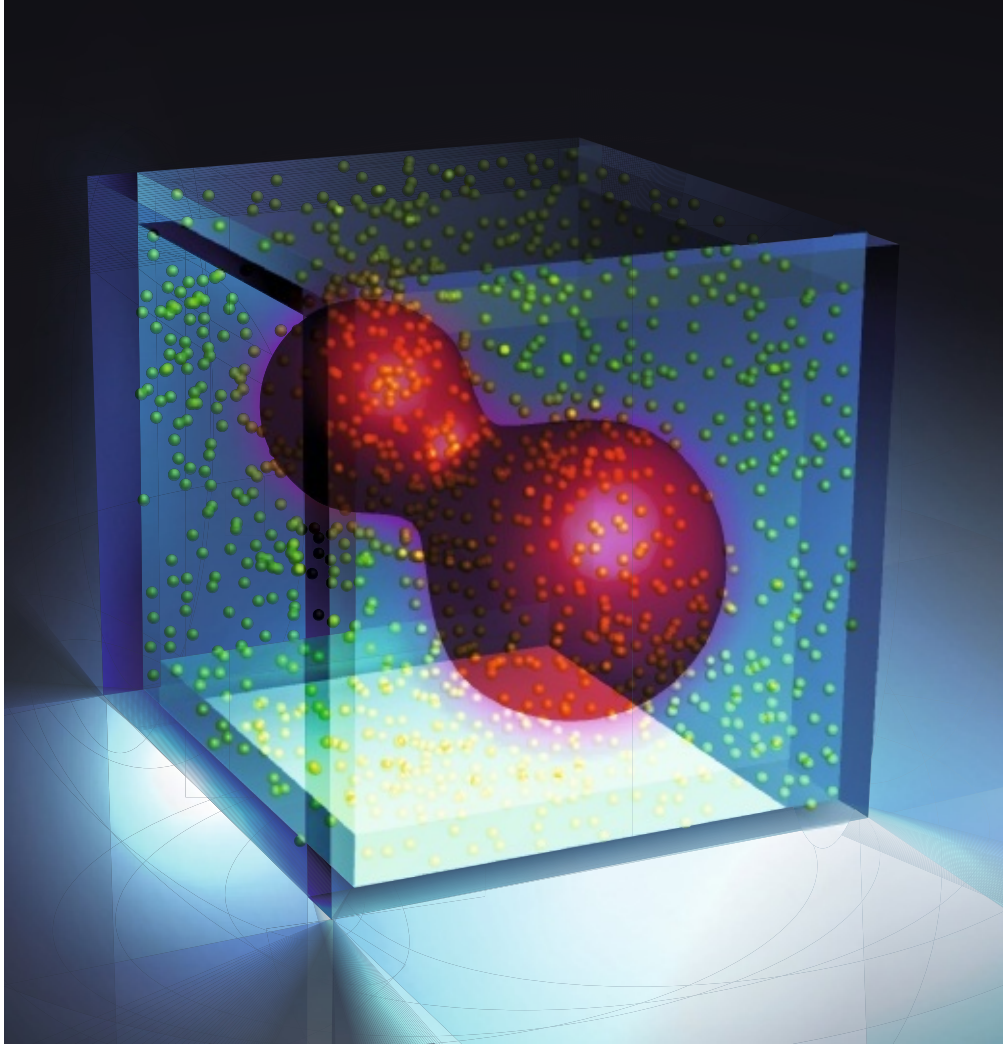
Depuis que nous avons mis en évidence le rôle des HLA-G dans la tolérance fœto-maternelle, plusieurs applications en médecine ont vu le jour. Lorsqu'une fécondation *in vitro* est réalisée, les médecins peuvent prédire la réussite de l'implantation de l'œuf selon la quantité de HLA-G solubles qu'il produit. Plus l'œuf sécrète de HLA-G, plus ses chances d'implantation dans l'utérus sont élevées. À chaque intervention, les médecins peuvent aussi ajuster le nombre d'œufs insérés dans l'utérus selon ce paramètre. En outre, on espère développer des traitements qui augmentent la sécrétion des HLA-G. Par exemple, certaines molécules immunitaires solubles, les interleukines et les cytokines, favorisent la production des HLA-G dans des cellules ne les exprimant pas ou peu. Elles pourraient améliorer les fécondations *in vitro*.

Thérapies et HLA-G

Par ailleurs, bien que l'expression des protéines HLA-G soit éteinte dans la plupart des tissus de l'organisme après la naissance, ces antigènes favoriseraient la transplantation d'organes. En 2000, en collaboration avec le chirurgien cardiaque Alain Carpentier, nous avons mis en évidence l'expression de la protéine HLA-G dans un cœur transplanté. Et nous avons montré que cette expression diminue considérablement le nombre de rejets. En outre, en 2010, lors d'une greffe de rein, nous avons constaté que la survie du greffon était meilleure quand nous stimulions la sécrétion des molécules HLA-G.

Malheureusement, les antigènes HLA-G sont aussi exprimés par les cellules tumorales : ils les protègent du système de défense de l'hôte. Nous savons d'ailleurs que presque toutes les cellules cancéreuses expriment des molécules HLA-G, ainsi que les cellules de leur environnement proche (les tissus du malade). Le dosage des HLA-G permet désormais de suivre l'évolution tumorale. De plus, on étudie aujourd'hui la possibilité de bloquer l'expression de ces molécules dans les cellules tumorales avec des anticorps.

Les antigènes HLA-G sont donc des molécules de tolérance dans le groupe des HLA. Bien qu'ils soient surtout produits par des cellules fœtales, ils peuvent s'exprimer chez l'adulte dans certaines pathologies, inflammatoires ou tumorales par exemple. ■



Les méthodes de Monte Carlo font intervenir des nombres pseudo-aléatoires pour effectuer de lourds calculs numériques. Mais des nombres dits quasi-aléatoires, où le hasard a moins de place, se révèlent parfois plus efficaces.

Excursions quasi-

Au début des années 1990, Spassimir Paskov, étudiant en thèse à l'Université Columbia, commençait à analyser un instrument financier exotique nommé CMO (*Collateralized Mortgage Obligation*, obligation garantie par une hypothèque) introduit par la banque d'investissements *Goldman Sachs*. Il cherchait à estimer la valeur actuelle d'un CMO, d'après les possibles encaissements à venir correspondant à des milliers d'emprunts-logement sur 30 ans. Il ne s'agissait pas simplement d'appliquer la formule usuelle pour les intérêts composés. Beaucoup d'emprunts-logement sont liquidés en avance quand le logement est vendu ou refinancé ; certains emprunts ne sont pas honorés ; les taux d'intérêt montent et baissent. Ainsi, la valeur actuelle d'un CMO sur 30 ans dépend de 360 (30×12) flux de trésorerie mensuels incertains et interdépendants. Il s'agissait donc plutôt d'évaluer une intégrale dans un espace à 360 dimensions.

Il n'y avait aucun espoir de faire une évaluation exacte. S. Paskov et son directeur de thèse, Joseph Traub, ont décidé d'essayer une technique d'approximation quelque peu obscure, la méthode de « quasi-Monte Carlo ». Un calcul de Monte Carlo ordinaire prend des échantillons aléatoires dans l'ensemble de toutes les solutions possibles. La variante « quasi » effectue un échantillonnage d'un type différent, non aléatoire, mais pas tout à fait régulier non plus. S. Paskov et J. Traub ont trouvé que certains de leurs programmes de quasi-Monte Carlo tournent mieux et plus vite que la technique classique. Leur découverte permettrait à un banquier ou un investisseur d'estimer la valeur d'un CMO en quelques minutes de calcul seulement, au lieu de plusieurs heures.

Cela ferait une belle histoire si l'on pouvait alors raconter que la période « d'exubérance irrationnelle » qui a suivi sur les marchés financiers (la frénésie des transactions de produits dérivés complexes, et le triste cortège de crise, effondrement, récession, chômage) avait pour origine une innovation mathématique dans l'évaluation des intégrales de dimension élevée. Mais ce n'est pas le cas ; il y avait d'autres causes à cette folie.

Cependant, les travaux de S. Paskov et J. Traub ont bien eu un effet : ils ont suscité un extraordinaire regain d'intérêt pour la méthode de quasi-Monte Carlo. Des résultats théoriques antérieurs avaient suggéré que les modèles de quasi-Monte Carlo commenceraient à s'essouffler quand le nombre de dimensions dépasserait une dizaine ou une vingtaine, en tout cas longtemps avant d'atteindre 360. Aussi le succès de

-aléatoires

Brian Hayes

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.

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 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 superficie 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.

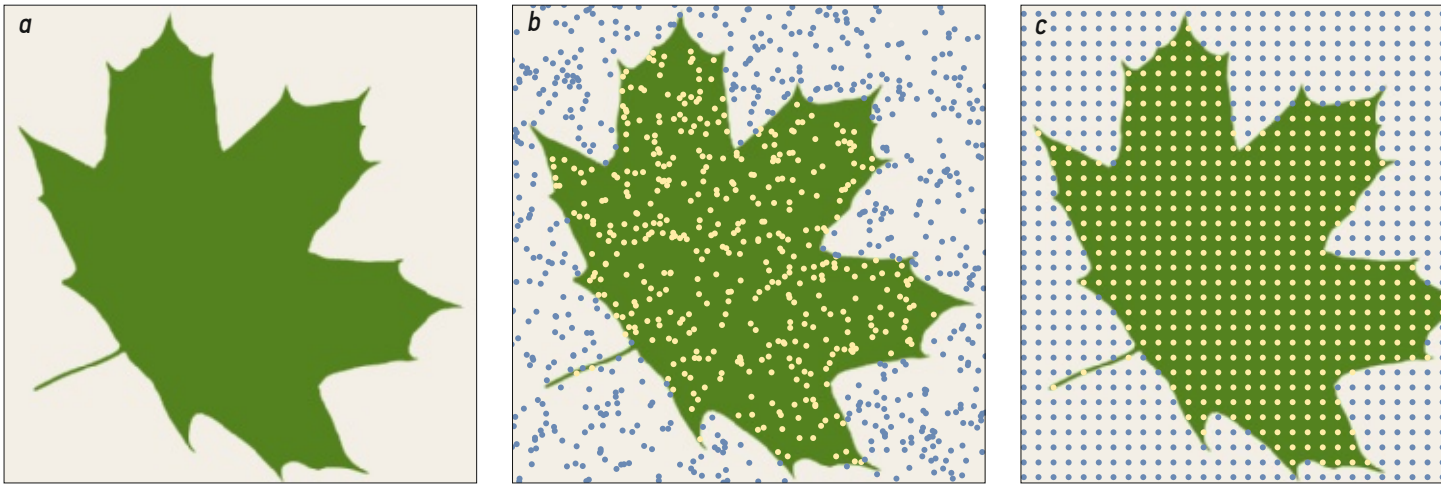
Lancer des fléchettes au hasard est non seulement difficile, mais aussi dangereux. Mieux vaut remplacer ces projectiles par des points et laisser l'ordinateur se charger de les parsemer au hasard. La mesure est alors facile, et l'expérience fonctionne remarquablement bien.

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 1 024 points aléatoires (en fait des points pseudo-aléatoires) sur l'image numérisée. La première fois que j'ai fait tourner le programme, 429 des 1 024 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 de 0,4183, avec un écart-type de 0,0153.

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

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.

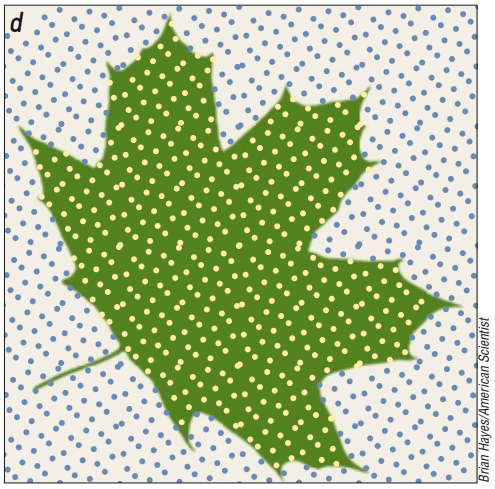
J'ai également essayé de mesurer la feuille avec 1 024 points quasi-aléatoires, dont l'arrangement est, d'une certaine manière, intermédiaire entre le chaos total et l'ordre total *(voir l'encadré page 59)*. L'estimation obtenue à partir du comptage de coups gagnants avec les points quasi-aléatoires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

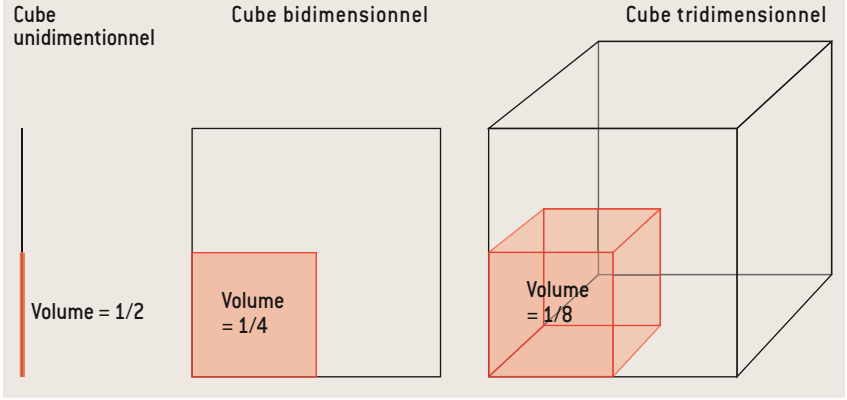
Les nombres quasi-aléatoires sont à la fois prévisibles et fortement corrélés. Il existe une règle bien définie pour les engendrer, et les motifs qu'ils forment, bien que pas aussi stricts que ceux d'un réseau cristallin, présentent une grande régularité. Par rapport au hasard, le seul élément que les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

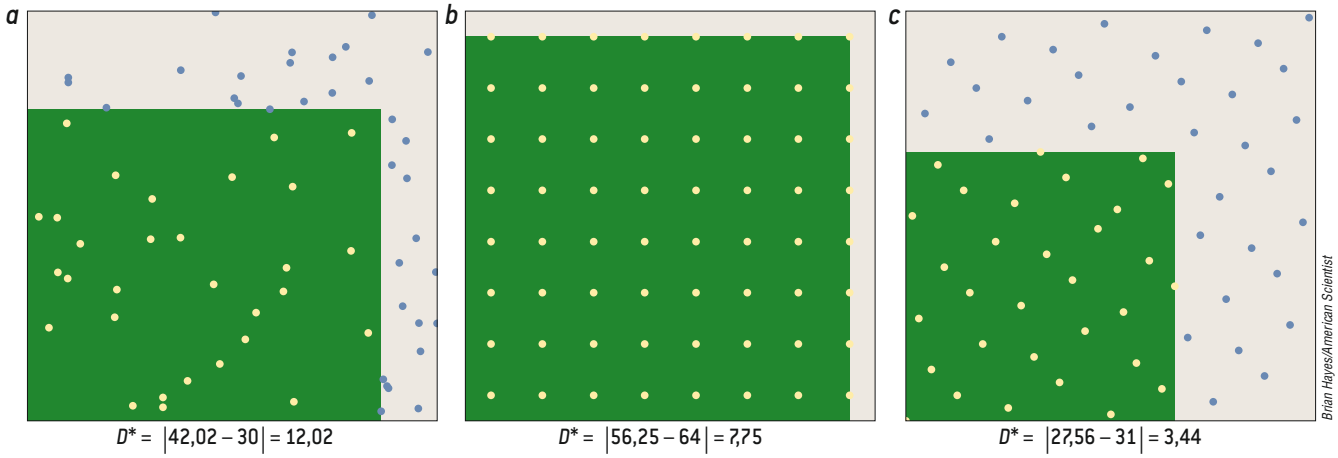
Un léger désaccord : la discrépance

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.



4. LA DISCRÉPANCE mesure de combien un ensemble de points s'écarte d'une distribution spatiale uniforme. La mesure illustrée ici, appelée discrépance étoile et notée D^*, est définie en termes de rectangles *(en vert)* ayant un sommet ancré dans le coin inférieur gauche. Si la distribution des points était parfaitement uniforme, le nombre de points que renferme l'un quelconque de ces rectangles serait proportionnel à la superficie du rectangle ; la différence entre ce nombre attendu de points et le nombre réel est la discrépance associée au rectangle. La discrépance étoile pour l'ensemble du motif est donnée par le rectangle pour lequel cette différence est maximale. La discrépance étoile est donnée ici pour trois motifs de 64 points : un réseau pseudo-aléatoire *(a)*, une réseau régulier *(b)* et un réseau quasi-aléatoire *(c)*. La configuration quasi-aléatoire est, par définition, une configuration conçue pour minimiser la discrépance.

distribution qui importe le plus. Et les nombres quasi-aléatoires, en abandonnant toute prétention à l'imprévisibilité ou à l'absence de corrélation, parviennent à atteindre une plus grande uniformité.

L'uniformité de la distribution est mesurée en termes de son opposé, la « discrépance ». Pour les points d'un carré bidimensionnel, la discrépance est mesurée comme suit. Considérez tous les rectangles aux côtés parallèles aux axes des coordonnées qu'il est possible de dessiner à l'intérieur du carré. Pour chacun de ces rectangles, comptez le nombre de points qu'il contient, et calculez aussi le nombre de points qui seraient contenus, en fonction de la superficie du rectangle, si la distribution était parfaitement uniforme. L'écart maximal entre ces nombres, calculés pour tous les rectangles possibles, est une mesure de la discrépance.

Une autre mesure de la discrépance, nommée discrépance étoile (parce qu'on la note D^*), ne considère que le sous-ensemble de rectangles parallèles aux axes qui ont un coin ancré à l'origine du carré unité. Et il n'y a aucune raison de ne considérer que les rectangles ; on peut définir d'autres versions de la discrépance avec des cercles, des triangles ou d'autres figures. Les mesures ne sont pas toutes équivalentes ; les résultats varient selon la forme. Il est intéressant de noter que le processus pour mesurer la discrépance a une grande ressemblance avec le processus Monte Carlo lui-même.

Les grilles et les réseaux ne sont pas très performants lorsqu'il est question de discrépance, parce que les points sont disposés en longues rangées et colonnes. Un rectangle infiniment étroit, qui ne devrait contenir aucun point, pourrait en englober toute une rangée. À cause de ces rectangles extrêmes, la discrépance d'un réseau carré de N points vaut $\sqrt{N}$. Il est intéressant de noter que la discrépance d'un réseau aléatoire ou pseudo-aléatoire vaut encore à peu près $\sqrt{N}$. En d'autres termes, dans un réseau d'un million de points aléatoires, il existe vraisemblablement au moins un rectangle qui a soit 1 000 points en trop, soit 1 000 points en moins.

Quasi-Monte Carlo en dimension élevée ?

Les motifs quasi-aléatoires sont délibérément conçus pour éliminer toutes les possibilités de dessiner de tels rectangles à discrépance élevée. Pour les points quasi-aléatoires, la discrépance peut ne pas dépasser le logarithme de N, ce qui est beaucoup plus petit que $\sqrt{N}$. Par exemple, pour $N = 10^6$, $\sqrt{N} = 1\,000$, mais $\log N = 20$ (avec les logarithmes de base 2).

La discrépance est un facteur clef pour évaluer les performances des simulations de Monte Carlo et quasi-Monte Carlo. Elle détermine le degré d'erreur ou d'imprécision statistique à attendre d'une taille d'échantillon donnée. Pour le Monte Carlo classique, avec un échantillonnage aléatoire, l'erreur attendue diminue en $1/\sqrt{N}$. Pour le quasi-Monte Carlo, la « vitesse de convergence » correspondante vaut $(\log N)^d/N$, où d est la dimension de l'espace (divers détails et facteurs constants étant négligés ici, ces vitesses de convergence sont approximatives).

La vitesse de convergence en $1/\sqrt{N}$ pour l'échantillonnage aléatoire peut être extrêmement lente. Pour gagner une décimale de précision, il faut augmenter N d'un facteur 100, ce qui signifie qu'un calcul d'une heure devient un calcul de quatre jours. La raison de cette lenteur provient des agrégats présents dans une distribution aléatoire de points. En effet, lorsque deux points sont très proches, les ressources de calcul sont gaspillées, puisque la même région est échantillonnée deux fois ; inversement, les grands vides entre les points laissent certaines zones non échantillonnées et contribuent donc à alourdir le poids des erreurs.

En compensation de cet inconvénient, l'échantillonnage aléatoire présente un avantage important : la vitesse de convergence ne dépend pas de la dimension de l'espace. En première approximation, le programme tourne aussi vite en dimension 100 qu'en dimension 2.

La méthode de quasi-Monte Carlo est différente à cet égard. Si seulement nous pouvions ignorer le facteur $(\log N)^d$, les performances de l'échantillonnage quasi-aléatoire seraient radicalement meilleures que celles de l'échantillonnage aléatoire. Nous aurions une convergence

en $1/N$, ce qui multiplie seulement par 10 l'effort pour chaque décimale de précision supplémentaire. Mais nous ne pouvons pas ignorer le terme $(\log N)^d$. Ce facteur croît exponentiellement avec la dimension d. L'effet est petit en dimension 1 ou 2, mais il finit par devenir énorme. En dimension 10, par exemple, $(\log N)^d/N$ reste supérieur à $1/\sqrt{N}$ tant que N est inférieur à 10^{43}.

C'est en raison de la lenteur connue de cette convergence en dimension élevée que S. Paskov et J. Traub ont créé la surprise avec le succès des calculs financiers pour $d = 360$. La seule explication plausible est que la « dimension effective » du problème serait en fait très inférieure à 360. En d'autres termes, le volume mesuré ne s'étendrait pas vraiment à toutes les dimensions de l'espace. De façon similaire, une feuille de papier occupe trois dimensions, mais on ne perd pas grand-chose à considérer qu'elle n'a pas d'épaisseur.

Cette explication pourrait sembler condescendante : le calcul a été un succès non pas parce que l'outil était plus puissant, mais parce que le problème était plus simple qu'il n'y paraissait. Mais il faut noter que cette réduction effective de la dimension fonctionne même quand nous ignorons lesquelles des 360 dimensions peuvent être négligées sans risque. C'est presque magique.

Dans quelle mesure ce phénomène est-il fréquent ? Est-ce juste un coup de chance, ou bien est-il confiné à une classe étroite de problèmes ? La réponse n'est pas encore totalement claire, mais une notion portant le nom de « concentration de mesure » donne des raisons d'être optimistes. Elle suggère que les univers de grande dimensionnalité sont essentiellement des mondes assez plats et lisses, analogues à une planète où régnerait une forte gravité et où la formation de paysages escarpés et accidentés serait trop coûteuse en énergie.

Regain d'intérêt pour une idée née à Los Alamos

La méthode de Monte Carlo n'est pas une idée nouvelle, et la variante de quasi-Monte Carlo non plus. Des simulations fondées sur un échantillonnage aléatoire ont été tentées il y a plus d'un siècle par Francis Galton, lord Kelvin et d'autres. À l'époque, ces scientifiques travaillaient

L'AUTEUR

Brian HAYES tient depuis 1993 la chronique *Computing Science* dans la revue *American Scientist*.

Article publié avec l'aimable autorisation de *American Scientist*.

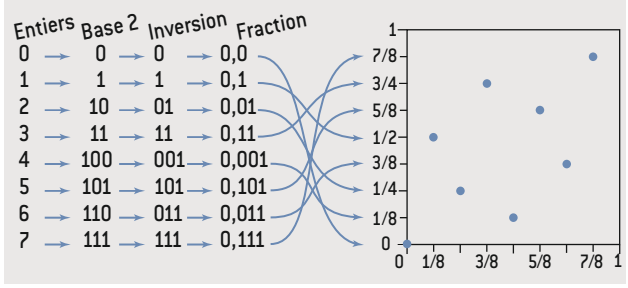
UNE RECETTE DE NOMBRES QUASI-ALÉATOIRES

Les nombres nécessaires pour former des motifs de discrépance faible ont été étudiés pour l'intérêt qu'ils présentaient en eux-mêmes longtemps avant l'invention de la méthode de quasi-Monte Carlo. Dans les années 1930, le mathématicien néerlandais Johannes van der Corput s'est demandé si des nombres distincts en séquence infinie pouvaient être placés un par un sur l'intervalle unité [0, 1] de telle sorte que la discrépance mesurée à chaque étape ne dépasse jamais une borne finie donnée. La réponse, prouvée une décennie plus tard par Tatyana van Aardenne-Ehrenfest, une compatriote de van der Corput, est négative : la discrépance croît sans limite. Néanmoins, les travaux de van der Corput ont mis en lumière toute une famille d'ensembles et de suites quasi-aléatoires à discrépance faible.

La procédure pour construire une séquence de van der Corput est particulière en cela qu'elle mélange des nombres (des entités mathématiques) avec des numéraux (la représentation des nombres sous forme de listes de chiffres). L'idée de base consiste à prendre un entier, d'inverser ses chiffres, de mettre une virgule devant, et enfin de traiter le résultat comme une fraction comprise entre 0 et 1. Les opérations peuvent être effectuées en n'importe quelle base. En base 2, par exemple, le nombre 100 (égal à 4 en base 10) devient 001 puis 0,001, qui est la représentation binaire de la fraction 1/8.

Pour créer ainsi un motif quasi-aléatoire de N points en dimension 2, on part de la séquence de nombres entiers $i = 0, 1, 2, \ldots, N - 1$. Puis, pour chaque point, on fixe la coordonnée x égale à i/N, tandis que la coordonnée y correspondante est donnée par le processus d'inversion des chiffres de van der Corput pour i. Le résultat pour $N = 8$ points est présenté ci-contre. Le motif quasi-aléatoire de la figure 2*d* a été créé de la même façon avec $N = 1\,024$. De nombreux autres algorithmes pour suites à discrépance faible ont été développés, mais la plupart d'entre eux reposent aussi sur le mélange des chiffres d'un nombre. Certaines de ces suites se généralisent facilement à des dimensions arbitrairement grandes.

Les motifs créés par de tels mécanismes maintiennent un équilibre délicat entre ordre et désordre. L'espacement entre les points est raisonnablement uniforme, si bien que n'apparaissent pas de vides ou de regroupements, qui augmentent la discrépance des ensembles aléatoires. Mais la distribution uniforme doit être obtenue sans laisser les points former des rangées ou autres structures régulières qui augmenteraient elles aussi la discrépance. Trouver le juste équilibre entre ces objectifs un peu contradictoires n'est pas difficile en deux dimensions, mais les compromis sont inévitables en dimension supérieure.



✔ BIBLIOGRAPHIE

H. Wozniakowski et L. Plaskota (éds.), **Monte Carlo and Quasi-Monte Carlo Methods 2010**, Springer, à paraître.

B. Tuffin, **La simulation de Monte Carlo**, Hermès Sciences, 2010.

J. Dick et F. Pillichshammer, **Digital Nets and Sequences. Discrepancy Theory and Quasi-Monte Carlo Integration**, Cambridge University Press, 2010.

P. L'Ecuyer, **Quasi-Monte Carlo methods with applications in finance**, *Finance and Stochastics*, vol. 13(3), pp. 307-349, 2009 (www.springerlink.com/content/384785458w558462/fulltext.pdf).

F. Y. Kuo et I. H. Sloan, **Lifting the curse of dimensionality**, *Notices of the American Mathematical Society*, vol. 52, pp. 1320-1328, 2005.

B. Chazelle, **The Discrepancy Method : Randomness and Complexity**, Cambridge University Press, 2000.

S. H. Paskov et J. E. Traub, **Faster valuation of financial derivatives**, *Journal of Portfolio Management*, vol. 22(1), pp. 113-121, 1995.

avec de véritables nombres aléatoires (et s'amusaient beaucoup à les engendrer !).

La méthode de Monte Carlo en elle-même a été inventée et a reçu son nom au Laboratoire de Los Alamos, dans les années qui ont suivi la Seconde Guerre mondiale. Il n'est pas étonnant que l'idée ait émergé en ce lieu : les chercheurs de ce laboratoire se confrontaient à des problèmes difficiles (encore plus effrayants que ceux des CMO) ; ils avaient accès aux premiers ordinateurs (ENIAC, MANIAC) ; et ils comptaient parmi eux des scientifiques très créatifs tels que Stanislaw Ulam, John von Neumann, Nicholas Metropolis et Marshall Rosenbluth.

Dès le départ, le groupe de Los Alamos a utilisé les nombres pseudo-aléatoires. Lors de la première conférence consacrée aux méthodes de Monte Carlo, en 1949, von Neumann a lancé sa célèbre boutade : « Quiconque envisage les méthodes arithmétiques de production de nombres aléatoires est, évidemment, dans le péché. » Suite à quoi, il suivit la voie du péché.

Le quasi-Monte Carlo n'était pas loin derrière. La première publication sur le sujet, en 1951, était un rapport de Robert Richtmyer, alors directeur du Département théorique de Los Alamos. L'article est une entrée en matière impressionnante. Il expose les motivations de l'échantillonnage quasi-aléatoire, introduit une bonne partie de la terminologie et explique les mathématiques associées. Mais il était également présenté sous forme de compte-rendu d'une expérience ratée. Richtmyer avait voulu obtenir un temps de convergence amélioré pour les calculs de quasi-Monte Carlo, mais ses résultats étaient négatifs, ce qui a pu freiner les études ultérieures.

En 1968, Stanislaw Zaremba, à l'Université du Wisconsin à Madison, écrivit une vibrante apologie de l'échantillonnage quasi-aléatoire (doublée d'une diatribe contre les nombres pseudo-aléatoires). À ma connaissance, il ne fit pas beaucoup d'émules.

Les travaux sur les mathématiques des suites à discrépance faible se sont poursuivis au fil des décennies (en particulier grâce au Britannique Klaus Friedrich Roth, au Russe Ilya Sobol, à l'Autrichien Harald Niederreiter, à l'Australien Ian Sloan). On assiste actuellement à un regain d'intérêt pour les applications, et pas seulement parmi les analystes quantitatifs de Wall Street. La physique et d'autres sciences s'y mettent aussi. Le lancer de rayon, technique de calcul des éclairages en infographie, est un autre domaine prometteur.

Les fortunes changeantes des méthodes pseudo-aléatoires et quasi-aléatoires s'inscrivent dans le contexte de tendances plus larges. Au XIXe siècle, le hasard de toute sorte était un intrus malvenu, toléré à contrecœur seulement quand cela était indispensable (en thermodynamique, dans la théorie de l'évolution darwinienne).

En revanche, le XXe siècle s'est entiché de tout ce qui est aléatoire. Les méthodes de Monte Carlo faisaient partie de ce mouvement. La théorie quantique en était un morceau plus conséquent, qui a insisté sur les jeux de dés divins. Le mathématicien d'origine hongroise Paul Erdös a introduit le choix aléatoire dans la méthodologie des démonstrations mathématiques, ce qui est peut-être l'endroit où l'on s'attendrait le moins à ce qu'il joue un rôle. En calcul informatique, les algorithmes probabilistes sont devenus un sujet de recherche majeur. Le concept a filtré jusque dans les arts, avec la musique aléatoire et les tableaux d'éclaboussures du peintre américain Jackson Pollock. Puis il y eut la théorie du chaos. En 1967, un essai de l'informaticien américain Alfred Bork (1926-2007) qualifiait l'aléatoire de « caractéristique capitale de notre siècle ».

Du hasard à dose homéopathique

Aujourd'hui, toutefois, on est peut-être en train d'assister à un retour de balancier, surtout en informatique. Les algorithmes probabilistes conservent une grande importance en pratique, mais l'excitation intellectuelle est du côté de la « déprobabilisation », en montrant que la même tâche peut être effectuée par un programme déterministe. Une question ouverte est de savoir si tous les algorithmes peuvent être déprobabilisés. Certains spécialistes pensent que la réponse sera positive. S'ils ont raison, l'aléatoire ne confère pas d'avantage essentiel en calcul, bien qu'il puisse être bien commode.

Le quasi-aléatoire semble nous emmener dans la même direction, avec une préférence pour une administration du hasard à toutes petites doses. C'est peut-être une doctrine homéopathique qui nous attend, où plus la dilution de l'agent actif est grande, plus il est puissant. C'est absurde en médecine, mais qui sait si cela ne marche pas en mathématiques et en algorithmique ? ■

ACTUELLEMENT EN KIOSQUE

Dossier Pour la Science

Découvrir la chimie qui régule les écosystèmes ou la vie autour des sources hydrothermales des abysses ; exploiter de façon raisonnée la richesse des fonds marins (microalgues, ressources minérales, méthane…) ; découvrir des médicaments nouveaux ; protéger les mers contre les marées noires ou les déchets plastiques… Aujourd'hui, la chimie est au centre d'une alliance durable entre les hommes et les océans.

Numéro 73, octobre-décembre 2011 – Prix : 6,95 €

Format classique ou pocket

L'Essentiel Cerveau & Psycho

Aujourd'hui, d'innombrables résultats de recherche confirment que les enfants – mais aussi les adolescents et les adultes – soumis à la violence (images, jeux vidéo, etc.) sont eux-mêmes plus violents, moins sensibles aux violences faites à autrui. Sociologues, psychologues, psychiatres cherchent à identifier les causes de la violence pour mieux la combattre.

Numéro 8, novembre 2011-janvier 2012 – Prix : 6,95 €

Format classique ou pocket

Cerveau & Psycho

Nous nous sentons stressés par notre rythme de vie, notre travail, etc. Et si le stress n'était pas forcément mauvais ? Certains y trouvent un stimulant et, dans certaines conditions, il renforce la mémoire. Mais, comme en toute chose, il faut éviter l'excès !
À lire aussi : la méthode Coué validée par les psychologues, les enfants bilingues avantagés, les méfaits du bruit sur la concentration, etc.

Numéro 48, novembre-décembre 2011 – Prix : 6,95 €

POUR LA SCIENCE