par J. -G. Dumas et D. Trystram

Les anniversaires des briseurs de codes

Les factorisations efficaces des nombres entiers font partie des armes des briseurs de codes. Une idée est d'utiliser le paradoxe des anniversaires.

Parmi 50 personnes, deux ont probablement le même anniversaire.

Dans une classe, quelle est la probabilité qu'au moins deux personnes aient le même anniversaire ? Si la classe compte k personnes, il semble logique que ce nombre soit égal à $k / 365$ environ, ce qui fait 1/16 pour une classe de 23 élèves. Cette intuition est totalement fausse ! Pour une classe de 23 élèves, la probabilité que deux élèves aient le même anniversaire est de 1/2, pour 53 ou plus, elle est de 99% !

Origine d'un paradoxe

D'où vient ce paradoxe ? L'explication tient dans une analyse plus fine des probabilités. Considérons l'événement contraire c'est-à-dire celui où toutes les personnes sont nées des jours différents. Pour déterminer une classe où tous les anniversaires sont différents, nous pouvons choisir l'anniversaire du premier ce qui fait 365 possibilités (366 les années bissextiles). Pour le second, nous n'avons plus que 364 possibilités, pour le troisième 363 et ainsi de suite. En ne tenant compte que des anniversaires, nous pouvons donc former $365 \times 364 \times \ldots \times (365 - k + 1)$ classes de k personnes ayant tous leurs anniversaires distincts. Comme il est possible de fabriquer 365^k classes distinctes, la probabilité que toutes les personnes soient nées des jours différents est donc égale à :

$$\frac{365 \times 364 \times \ldots \times (365 - k + 1)}{365^k}$$

donc la probabilité qu'au moins deux personnes aient le même anniversaire est :

$$1 - \frac{365 \times 364 \times \ldots \times (365 - k + 1)}{365^k}.$$

Pour $k = 23$ et $k = 53$, nous trouvons bien les probabilités annoncées.

Plus généralement, prenons n possibilités et k tirages aléatoires indépendants successifs. Le même calcul montre qu'il suffit que k soit égal à environ racine de $\sqrt{n}$ pour avoir une chance sur deux que deux tirages soient identiques (voir l'encadré *Un calcul asymptotique pour le détail*).

Ed Kienholz, *The Birthday*, 1964

Méthode de Monte-Carlo

Soit n un nombre non premier, p son plus petit facteur et q le facteur complémentaire ($n = pq$). Une méthode pour factoriser n est de tirer au hasard des nombres entre 0 et $n - 1$. Parmi ces nombres, q sont multiples de p ($0, p, 2p, \ldots, (q-1)p$), on a donc une chance sur p de tomber sur un multiple de p et donc de trouver p en calculant le plus grand commun diviseur de ce nombre et de n. En moyenne, cet algorithme nécessite donc p tirages pour trouver p. Comme le plus petit facteur premier de n est plus petit que sa racine carrée, en moyenne, cette méthode nécessite un nombre d'étapes de l'ordre de la racine carrée de n pour déterminer le plus petit facteur premier de n. Cette méthode est dite de Monte-Carlo car, de même que la fortune des casinos, elle est fondée sur l'exploitation du hasard.

Un calcul asymptotique

De même que dans le calcul des anniversaires, la probabilité que deux tirages de k nombres parmi n soient identiques est égale à 1 moins :

$$\frac{n\,(n-1)\,\ldots\,(n-k+1)}{n^k} = \frac{n\,!}{n^k\,(n-k)}.$$

Cette quantité peut être simplifiée grâce à la formule de Stirling selon laquelle :

$$n\,! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Pour $k = \sqrt{n}$, nous obtenons après implifications :

$\left(1-\frac{1}{\sqrt{n}}\right)^{\sqrt{n}-n-\frac{1}{2}} e^{-\sqrt{n}}$ dont le logarithme est égal à : $\left(\sqrt{n}-n-\frac{1}{2}\right)\ln\left(1-\frac{1}{\sqrt{n}}\right)-\sqrt{n}$.

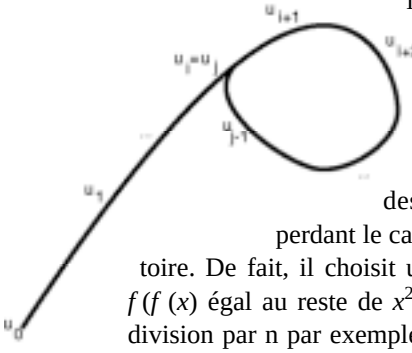
En utilisant un développement limité du logarithme, on en déduit que sa limite est égale à $-\frac{1}{2}$ donc celle de $\frac{n\,!}{n^k\,(n-k)}$ à $e^{\frac{1}{2}}$. Pour n assez grand, la probabilité que deux tirages soient identiques est donc égale à $1 - e^{\frac{1}{2}}$ qui est inférieur à $\frac{1}{2}$.

Amélioration par les anniversaires

Le paradoxe des anniversaires permet d'améliorer cette méthode de Monte-Carlo. Tirons toujours des nombres au hasard entre 0 et $n-1$, mais en les conservant et en les comparant deux à deux. En racine de p tirages (en moyenne), on obtient deux nombres u et v ayant le même « anniversaire » c'est-à-dire tels que $u-v$ soit un multiple de p ! Il suffit de considérer le plus grand commun diviseur de n et de $u-v$ pour factoriser n. En racine quatrième de n tirages en moyenne, on a donc réussi à factoriser n. Le seul problème restant est le stockage en mémoire de tous ces u. Comme ce sont des très grands nombres, c'est impossible.

Algorithme de Pollard

L'idée Pollard est de ne stocker que certains d'entre eux, et de faire en sorte que, si un « anniversaire » se produit, il se répète à intervalles réguliers. Pour cela, il modifie la règle de formation des u en en perdant le caractère aléatoire. De fait, il choisit une fonction f (f (x) égal au reste de x^2+1 dans la division par n par exemple) et un premier u de façon aléatoire, soit u_0. Les autres sont donnés en appliquant f successivement : $u_1 = f(u_0)$, $u_2 = f(u_1)$, etc. Si f est bien choisie, cette suite se comporte comme une suite aléatoire. Comme les u sont des nombres entre 0 et $n-1$, ils ne peuvent être tous distincts. Cette répétition forme alors un cycle et est représentée par un schéma en forme de ρ d'où son nom.



Le ρ de Pollard : il existe i et j tels que $u_i = u_j$. La règle de formation de la suite implique que les termes se reproduisent ensuite : $u_{i+1} = u_{j+1}$, etc.

F. Waldmüller, *The Birthday Table*, 1840

En pratique, cet algorithme factorise en quelques secondes les nombres d'une vingtaine de chiffres (les facteurs de 12 ou 13 chiffres nécessitent environ dix millions d'itérations !), mais il devient très rapidement inutilisable pour des facteurs plus grands.

Casser le code RSA

Le code RSA (voir l'article *Le code RSA*) est le système de codage à clef publique le plus connu. C'est aussi le plus utilisé, par exemple lors de transactions sécurisées sur Internet (pour la confidentialité du courrier ou l'authentification des utilisateurs). La clef publique correspond à un nombre $n = pq$. Pour le casser, il suffit de connaître p et q d'où l'intérêt de savoir factoriser de grands nombres. Le ρ de Pollard suffit pour factoriser des nombres de quelques dizaines de chiffres mais est insuffisant pour casser le code utilisé par la carte bleue mais il peut être utilisé pour des recherches de collisions dans les fonctions de hachage (voir l'article *Signature et hachage*).

J.-G.D. & D.T.