

UNE SURPRISE DE L'INFINI MATHÉMATIQUE

L'infini mathématique produit parfois d'étonnantes propriétés dont on pourrait croire à tort qu'elles sont le fait de coïncidences. En voici un exemple : il est possible de trouver un multiple de votre année de naissance (ou de votre numéro de sécurité sociale, ou numéro de carte bancaire) qui ne contienne que des 7 et des 0, les 7 étant tous placés devant les 0. Exemple : le nombre $m = 1\,968$, multiplié par $k = 3\,952\,122\,854\,561\,875$, donne $n = 7\,777\,777\,777\,777\,770\,000$.

Ce qui est vrai pour 1 968 est vrai pour tout nombre entier ! Cela semble magique. Pourtant, c'est assez facile à démontrer. On pourra d'ailleurs aisément écrire un programme qui, pour tout m donné (pas trop grand !) trouve le k voulu.

Démonstration : Soit un entier quelconque m . Considérons tous les nombres $7, 77, 777, 7777, \dots$ et divisons-les par m , en notant les restes obtenus. – Deux de ces nombres (et même une infinité de couples) donneront le même reste, car les restes possibles d'une division par m sont $0, 1, \dots, m-1$ et sont donc en nombre fini. – Soustrayons le plus petit du plus grand : $n = 7777\dots777 - 777\dots77$. – Par construction, comme différence de deux nombres donnant le même reste quand ils sont divisés par m , le nombre n est un multiple de m . De plus, il est de la forme $777\dots777000\dots000$. CQFD.

2

- > trompe en croyant qu'une observation est significative et doit donc être expliquée, alors qu'elle devait se produire (elle ou une autre du même type) du fait du grand nombre d'hypothèses envisagées, parfois collectivement par l'ensemble des équipes de recherche, dans la quête de régularités statistiques.

DATA SNOOPING

Cette illusion est proche de celle dénommée *data snooping* (« furetage discret de données ») que l'on peut utiliser pour des tromperies délibérées. Pour en illustrer le fonctionnement, Halbert White, de l'université de Californie à San Diego, a suggéré la méthode suivante permettant à un journal financier d'augmenter le nombre de ses abonnés.

La première semaine, le journal envoie un exemplaire gratuit à 20 000 personnes; dans la moitié des numéros envoyés, il est écrit que l'indice boursier (par exemple le Dow Jones) va monter, dans l'autre moitié le journal affirme que l'indice va baisser. Selon que l'indice a effectivement monté ou baissé, le journal envoie la semaine suivante, aux 10 000 adresses qui ont reçu la bonne prévision, un second numéro gratuit, avec dans la moitié des exemplaires l'annonce pour la semaine suivante que l'indice va monter et dans l'autre moitié qu'il va baisser. La troisième semaine, le journal envoie 5 000 numéros gratuits à ceux qui ont reçu la bonne anticipation deux semaines de suite, etc.

À chaque fois, le journal insiste sur le fait qu'il a correctement prévu la tendance depuis plusieurs semaines et propose un bulletin d'abonnement. Au bout de 10 semaines, il ne restera qu'une vingtaine d'envois à faire, mais pour être persuadé que le journal sait prédire la bonne tendance de l'indice boursier, rares sont les lecteurs potentiels qui attendent une prévision exacte 10 fois de suite. Par conséquent, un bon nombre de nouveaux abonnements auront été souscrits à mesure du déroulement des envois.

Une autre situation de *data snooping* dans le traitement des données financières conduit à une navrante désillusion. On cherche des règles du type « Si aujourd'hui le cours de l'action A monte et que le cours de l'action B baisse et que..., alors le cours de l'action X montera demain ». On en écrit un grand nombre, voire on écrit toutes les règles de ce type comportant 10 éléments dans leurs prémisses. On sélectionne ensuite, à l'aide d'une série de données provenant du passé, les meilleures règles. On élimine toutes les règles qui se sont trompées avec les données passées, et on retient seulement celles qui ont toujours fourni une bonne prédiction, ou celles qui ont eu raison le plus souvent. On disposera alors inévitablement d'une série réduite de règles qui, si elles avaient été appliquées sur les données du test, auraient permis de gagner beaucoup d'argent... et qui pourtant ne rapporteront rien dans les semaines qui suivront leur mise en œuvre pour faire des achats et des ventes.

Bien sûr, le piège est connu des chercheurs et des méthodes ont été mises au point pour ne pas être victime de l'illusion. On cherchera par exemple à évaluer, avant de mener la recherche des bonnes règles, la probabilité que lorsque les données sont tirées au hasard l'une des règles de la liste envisagée fonctionne avec ce hasard, et on s'assurera que cette probabilité est proche de 0.

LOTÉRIE TRUQUÉE?

Qui peut considérer comme normal qu'à quelques jours de distance, la même série de 6 numéros sorte d'un tirage du Loto? C'est pourtant ce qui se produisit le 10 septembre 2009 pour le Loto bulgare. La série 4, 15, 23, 24, 35, 42 n'a rien d'extraordinaire, sauf qu'elle avait déjà été tirée 6 jours auparavant, le 4 septembre, par le même Loto bulgare. Une coïncidence incroyable, au point que le ministre des Sports Svilen Neïkov demanda l'ouverture d'une enquête. Aucune tricherie n'a été détectée, et d'ailleurs les deux tirages avaient eu lieu devant les caméras de la télévision. La chose fut considérée si extraordinaire que la presse dans le monde entier rapporta l'événement.