> à 916,34 euros, ce qui correspond à une augmentation de 152%. Pourtant, à nouveau personne ne ment. Comment est-ce possible?

Le tableau répond: le salaire hebdomadaire des ouvriers, passant de 200 euros à 180 euros, a baissé de 10%. Celui des cadres, passant de 2000 à 1800 euros, a lui aussi diminué de 10%. Les syndicats ont donc raison d'affirmer que les salaires ont baissé de 10%.

De son côté, le patron ne triche pas! Le salaire versé aux 1100 employés de l'entreprise en 2006 était chaque semaine de 400000 euros ($1000 \times 200 + 100 \times 2000$), soit 363,64 euros par employé. En 2007, le salaire versé aux 1100 employés (l'effectif global est inchangé) s'est élevé à 1008000 euros ($180 \times 600 + 1800 \times 500$), soit 916,34 euros par employé. Le salaire moyen a donc bien augmenté de 152%.

L'arnaque, car il y en a une, est que les effectifs par catégorie ont changé entre 2016 et 2017. Il en résulte que la baisse du salaire dans chaque catégorie est compensée par l'augmentation du nombre des cadres qui sont mieux rétribués. Une baisse de 10% du salaire de chaque catégorie de personnel est parfaitement compatible avec une augmentation du salaire moyen des employés.

Cette situation n'est pas tant fictive que cela, car chaque année l'État français insiste sur les chiffres de la masse salariale des fonctionnaires et sur le salaire moyen d'un fonctionnaire qui évoluent plus favorablement que le point d'indice utilisé pour payer les salaires des fonctionnaires. Ce point d'indice détermine, à grade fixé, le salaire d'un fonctionnaire et bien sûr c'est à lui que les syndicats préfèrent se fier. L'augmentation de l'âge moyen des fonctionnaires – due en particulier à un fort recrutement dans l'enseignement dans les années 1970 et 1980 – entraîne mécaniquement une augmentation du grade moyen (comme dans l'entreprise fictive de l'exemple indiqué plus haut) et conduit donc à une divergence entre les chiffres portant sur le salaire moyen d'un fonctionnaire et ceux portant sur le point d'indice, que ministres et syndicats utilisent de la façon qui les arrange le mieux.

Toujours à propos des fonctionnaires, voyons un autre paradoxe apparent. On peut affirmer sans se contredire que: (*a*) le salaire moyen dans la fonction publique est supérieur à celui dans le privé; (*b*) la majorité des fonctionnaires gagneraient plus s'ils partaient dans le privé. L'explication réside à nouveau dans la répartition des emplois du secteur public et du secteur privé: dans le premier, le nombre d'emplois qualifiés est plus important, ce qui a pour effet d'augmenter le salaire moyen du secteur public, sans pour autant contredire qu'à diplôme égal on gagne moins dans le public que dans le privé.

La statistique utilise des indices aux définitions précises qui résument en un seul nombre une masse parfois considérable de chiffres. Ces indices sont inévitables: sans eux, on ne pourrait synthétiser des tableaux de données. Cependant les indices des statisticiens tendent des pièges et peuvent engendrer des résultats totalement absurdes. Soit que celui qui les examine les comprenne imparfaitement, soit qu'il prenne mal en compte des situations particulières.

L'espérance de vie est un indice piège (*voir l'encadré page précédente*). Le seuil de pauvreté tel que le définit l'Insee est tout aussi troublant. Par définition, le seuil de pauvreté dans un pays est la moitié du revenu médian, c'est-à-dire la moitié du revenu X tel qu'il y a autant de gens gagnant plus de X que de gens gagnant moins dans le pays. L'indice de pauvreté est le pourcentage de gens vivant sous le seuil de pauvreté.

Si maintenant quelqu'un vous annonce que dans tel pays 60% des gens vivent sous le seuil de pauvreté, c'est une ânerie. Par définition, ce n'est pas possible: la moitié (à quelques unités près) des gens ont un revenu inférieur au revenu médian, et nécessairement moins de la moitié ont un revenu inférieur à la moitié du revenu médian.

La conséquence dramatique est que même dans un pays où tout le monde mourrait de famine, le nombre de pauvres resterait inférieur à 50%, alors qu'à l'opposé, dans un pays composé uniquement de milliardaires, il pourrait tout à fait y avoir 45% de gens vivant sous le seuil de pauvreté. La pauvreté et la richesse sont des concepts relatifs, certes, mais tout de même!

ÊTRE PAUVRE EN COCAGNE

Nicolas Gauvrit imagine la petite histoire suivante qui montre par l'absurde combien l'utilisation de l'indice de pauvreté est dangereuse. Au pays de Cocagne, une pomme coûte un millième de sol. Un logement en coûte 5. On mange bien pour 0,2 sol. On vit dans le confort pour 100 sols par mois, et comme un nabab pour 200 sols par mois. Il y a dans ce pays deux types de travailleurs: les ouvriers qui ont un revenu de 1000 sols par mois, et les penseurs qui ont un revenu mensuel de 3000 sols. Il y a autant d'ouvriers que de penseurs, exactement. Tous vivent en harmonie et dans une opulence que les pays voisins jalousent. Seule exception à la règle des deux salaires: le président du pays touche, pour sa part, une rémunération de 2800 sols. Le revenu médian est alors de 2800 sols, comme vous le confirmerait tout statisticien. La moitié du revenu médian est donc de 1400 sols, et la moitié de la population (sans compter le président) touche une rétribution inférieure. L'indice de pauvreté est donc de 50% et c'est la valeur la plus élevée possible de cet indice.

Mais un beau jour de mai 2068, le président accepte de baisser ses indemnités, car, dit-il «il n'y a pas de raison pour que je réclame plus qu'un ouvrier! Je suis un travailleur comme les autres». Tenant compte de ses responsabilités, il s'accorde toutefois une petite rallonge et se déclare satisfait de son nouveau revenu de 1400 sols mensuels.

CITATIONS

«Le loto, c'est un impôt sur les gens qui ne comprennent pas les statistiques.»

Anonyme

«Il est statistiquement prouvé que sur dix personnes atteintes de bronchite, une seule va chez son médecin et les neuf autres dans une salle de spectacle.»

Anonyme

«En moyenne chaque personne possède un testicule.»

Anonyme

«À la question: "Faites-vous encore confiance aux sondages?", 64% des Français répondent OUI et 59% répondent NON.»

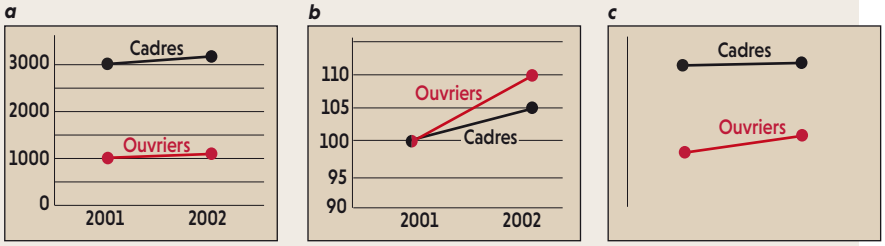
Philippe Geluck

L'ART DU DESSIN

Il y a mille façons de projeter graphiquement des données, chacune permettant de mettre en avant un fait, d'en masquer un autre, ou de donner à croire autre chose que ce que les données indiquent. L'exemple suivant a d'abord été proposé dans la revue *The Economist*. Entre 2001 et 2002, les salaires des cadres et des ouvriers ont évolué ainsi :

	SALAIRES MENSUELS MOYENS	
	2001	2002
OUVRIERS	1000 €	1100 €
CADRES	3000 €	3150 €

En moyenne, les cadres et les ouvriers gagnent plus en 2002 qu'en 2001. En valeurs brutes, l'augmentation des cadres (150 euros) est supérieure à celle des ouvriers (100 euros). Cependant, la situation est inversée en valeur relative : 5 % pour les cadres contre 10 % pour les ouvriers. Ces points de vue différents ont leurs équivalents graphiques, respectivement en *a* et en *b*.
Le commentaire naturel de la représentation *a* est : « Les ouvriers et les cadres ont été augmentés. Les cadres recevaient plus, et reçoivent encore plus que les ouvriers. L'écart entre les salaires empire. »
La représentation *b* s'attache aux augmentations relatives. Elle montre que les cadres ont été moins augmentés en pourcentage que les ouvriers, mais elle ne renseigne pas sur les éventuelles modifications de rémunération en valeur absolue. Cette courbe donne l'impression que les cadres sont moins gâtés que les ouvriers…
Peut-on représenter les évolutions relatives et les différences entre les professions ? Oui, en ayant recours à une échelle logarithmique en ordonnée *(c)*. Avec cette graduation, une même différence de hauteur correspond à un même coefficient multiplicateur.
La lecture de ce graphique suggère cette fois : « Les cadres recevaient et reçoivent encore plus que les ouvriers, mais les inégalités se réduisent. »
Ces trois représentations étant toutes parfaitement honnêtes, on peut commenter l'évolution des salaires selon sa préférence.



Le revenu médian en Cocagne passe alors illico à 1400 sols, et le demi-revenu médian à 700 sols, et par un miracle statistique à couper le souffle, l'indice de pauvreté passe aussitôt, en mai 2068, à 0, le minimum possible. »

Créé suite à une polémique sur l'indice de pauvreté, le BIP 40 ou baromètre des inégalités et de la pauvreté est un indicateur synthétique des inégalités et de la pauvreté. Cet indice a été proposé par le Réseau d'alerte sur les inégalités en 2002 et il élimine certains des problèmes mentionnés ci-dessus. Toutefois, il est assez complexe, si bien qu'il est délicat d'en comprendre le sens et qu'on ne peut garantir qu'il évitera toutes les absurdités des indices plus simples.

UNE AFFAIRE DE FAMILLE

Parmi tous les pièges que détaille le livre de Nicolas Gauvrit, l'un d'eux est assez subtil et mérite une attention particulière, nous le nommerons le paradoxe du nombre moyen d'enfants. Une enquête exhaustive menée dans une ville lointaine indique que les familles ayant des enfants de moins de 18 ans se répartissent de la manière suivante : 10 % de familles à 1 enfant, 50 % à 2 enfants, 30 % à 3 enfants, 10 % à 4 enfants. Le nombre moyen d'enfants par famille (parmi celles qui ont des enfants) est donc de $(10 + 100 + 90 + 40)/100 = 2{,}4$.

Pour contrôler cette statistique, les autorités administratives procèdent à un sondage. On interroge 1000 enfants de moins de 18 ans soigneusement pris au hasard et on leur demande combien il y a d'enfants dans leur famille, eux compris. En faisant la moyenne des réponses, on obtient… 2,68 ! Cela semble absurde. On recommence donc le sondage en interrogeant cette fois 10000 enfants, on trouve maintenant 2,67. Un troisième sondage sur 100000 enfants donne 2,668 à nouveau. Pourquoi cet écart si important avec les 2,4 de la statistique qui prenait en compte toutes les familles ayant des enfants ?

La réponse tient dans le fait qu'en interrogeant des enfants au hasard, vous interrogerez 4 fois plus d'enfants des familles à 4 enfants que vous n'en interrogerez dans les familles à 1 enfant, ce qui fausse la moyenne. S'il y a 1000 familles, il y aura 100 enfants uniques, 1000 enfants appartenant à une famille de 2 enfants, 900 enfants appartenant à une famille de 3 enfants, 400 enfants appartenant à une famille de 4 enfants. Au total, les réponses données par ces 2400 enfants conduiront au résultat de 2,666… enfants par famille.

Les sondages opérés n'évaluent pas le nombre moyen d'enfants d'une famille prise au hasard, mais le nombre moyen d'enfants qu'on trouve dans la famille d'un enfant pris au hasard. « Prendre une famille au hasard » et « Prendre un enfant au hasard » n'est pas la même chose.

Nous le savons depuis Condorcet pour les votes, il est bien difficile de synthétiser un ensemble de nombres en un seul. Nous espérons que les pièges de la statistique, des représentations graphiques, des indices synthétiques, des sondages présentés ici – et ceux que vous trouverez dans le livre de Nicolas Gauvrit – vous aideront à mieux comprendre les vérités cachées derrière l'inquiétant et fluctuant monde des nombres. ■

BIBLIOGRAPHIE

N. GAUVRIT, Statistiques. Méfiez-vous, *Éditions Ellipses*, 2014.

I. EKELAND, Statistiques incroyables, *Pour la Science* n° 334, p. 6, août 2005.

P. BICKEL *ET AL*, Sex Bias in Graduate Admissions: Data from Berkeley, *Science*, vol. 187, n° 4175, pp. 398-404, 1975.

G. YULE, Notes on the Theory of Association of Attributes in Statistics, *Biometrika*, vol. 2, n° 2, pp. 121-134, 1903.

L'ESSENTIEL

- La loi des grands nombres et le théorème de la limite centrale décrivent le comportement de la moyenne d'un grand nombre de petites contributions indépendantes.
- Ils échouent cependant lorsque des événements rares ou extrêmes dominent.
- D'autres outils statistiques – lois de Lévy ou des valeurs extrêmes – sont alors plus indiqués.
- Mais quand le risque n'est pas quantifiable, ces outils se révèlent à leur tour impuissants.

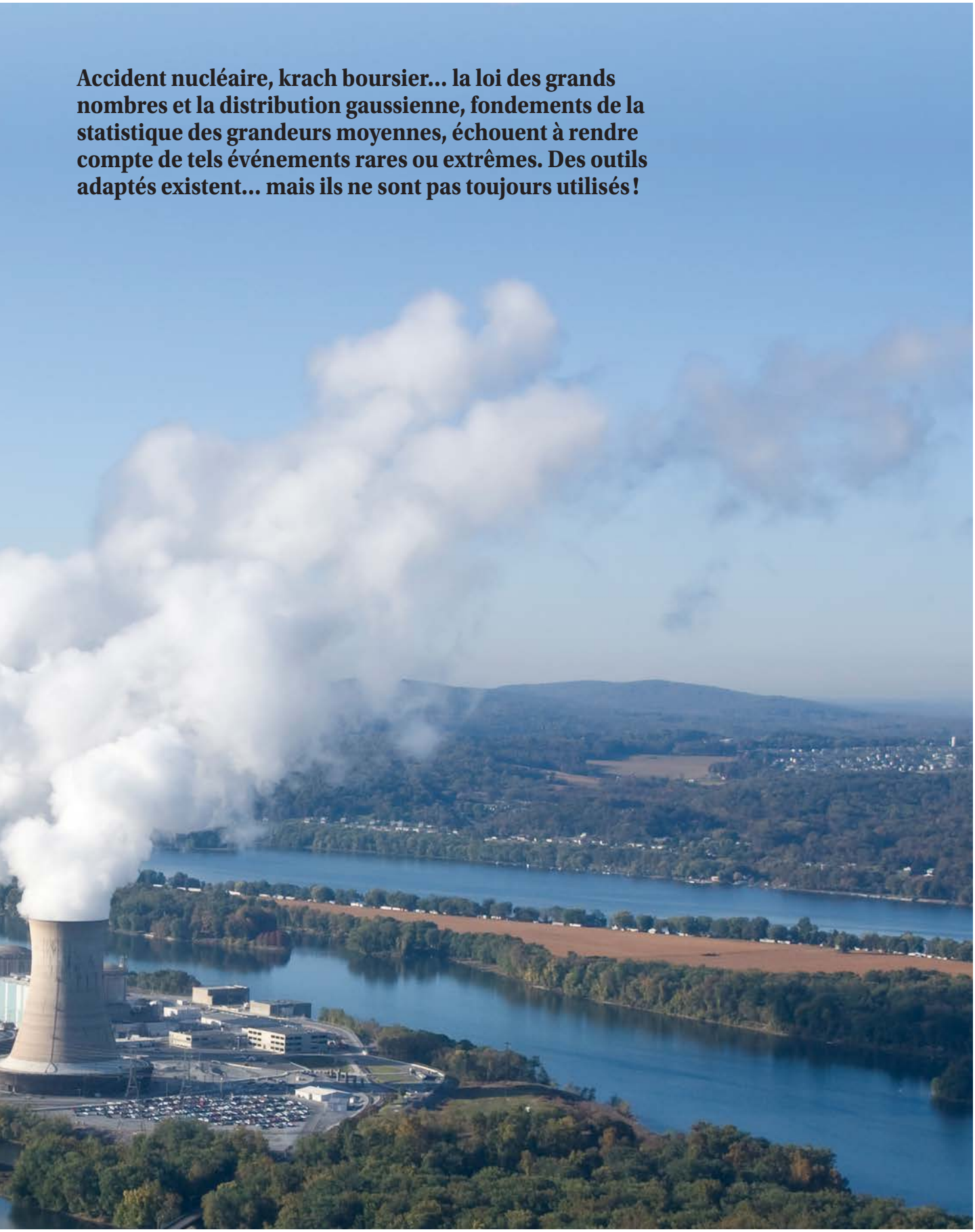
L'AUTEUR

RAMA CONT
est directeur de recherche CNRS au Laboratoire de probabilités et modèles aléatoires, à Paris. Ses travaux portent sur les processus aléatoires et la modélisation mathématique des risques financiers.

Prévoir l'IMPROBABLE

L'accident de la centrale nucléaire de Three Mile Island, en 1979, résulte d'un enchaînement improbable de petites erreurs – dont une humaine –, chacune étant difficilement quantifiable. Était-il prévisible ?

Accident nucléaire, krach boursier… la loi des grands nombres et la distribution gaussienne, fondements de la statistique des grandeurs moyennes, échouent à rendre compte de tels événements rares ou extrêmes. Des outils adaptés existent… mais ils ne sont pas toujours utilisés !



Le 15 septembre 2008, la banque américaine Lehman Brothers se déclare en faillite. Cet événement, consécutif de la crise des *subprimes* l'année précédente, ébranle le système financier mondial et affole les bourses de la planète. Les gouvernements et les banques centrales enchaînent les plans de relance, mais rien n'y fait, le monde plonge dans la crise financière la plus grave depuis celle de 1929. Dans quelle mesure aurait-on pu prévoir ce cataclysme ?

De tels événements rares et catastrophiques (aux krachs boursiers, ajoutons les tremblements de terre, les inondations, les accidents nucléaires…) fascinent par leur caractère imprévisible et l'importance des enjeux sociaux et scientifiques qu'ils représentent. Ils sont aussi des défis à la modélisation scientifique. Exceptionnels presque par définition, ou parfois mis de côté comme « aberrations statistiques », les événements rares sont néanmoins accessibles à la théorie statistique… pour peu qu'elle sollicite des outils mieux adaptés que ceux du quotidien.

Dans beaucoup d'applications, les statistiques se résument au calcul de moyennes et d'écarts-types, et la distribution des observations se répartit sagement selon la fameuse « courbe en cloche » ou distribution gaussienne.

LES STATS POUR LES NULS

Le statisticien représente souvent un échantillon d'observations (températures, précipitations, indices boursiers…) comme une suite de tirages indépendants d'une « variable aléatoire » notée X et représentant la grandeur en question. On s'intéresse alors à la moyenne, ou espérance mathématique, de cette variable aléatoire, à son écart-type (la dispersion des valeurs autour de la moyenne) et, plus généralement, à sa distribution de probabilité.

La loi des grands nombres, établie pour la première fois dans sa version la plus simple au XVIIIe siècle par Jacques Bernoulli, assure que, lorsque la taille de l'échantillon est assez grande, la moyenne empirique de l'échantillon tend vers la moyenne théorique de la variable aléatoire qu'il représente. De même, l'écart-type empirique converge vers l'écart-type de la variable aléatoire. En garantissant qu'un échantillon fournit une bonne estimation d'une variable aléatoire réelle, la loi des grands nombres s'impose comme un des piliers de la statistique.

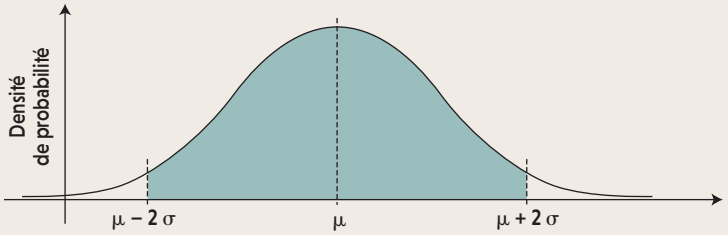
La loi des grands nombres est une loi asymptotique : la moyenne d'un échantillon ne converge vers l'espérance mathématique que si la taille N de l'échantillon est grande. La différence entre les deux valeurs – l'erreur statistique – est régie par le « théorème de la limite centrale » (*voir l'encadré page ci-contre*). Ce théorème, second pilier des statistiques, est connu sous des formes diverses depuis le XVIIIe siècle, mais sa forme définitive, due au mathématicien français Paul Lévy, date du début du XXe siècle. Il affirme que, si la variable aléatoire X a un écart-type fini, alors pour un échantillon de N observations indépendantes de X, la différence entre la moyenne de l'échantillon et la moyenne de X est bien représentée par une distribution gaussienne dont l'écart-type est proportionnel à $1/\sqrt{N}$.

La distribution gaussienne apparaît donc comme une description statistique de toute quantité que l'on peut représenter comme la somme d'un grand nombre de petites contributions indépendantes.

Ces deux principes ont une grande portée. La loi des grands nombres est le fondement des sondages, car elle montre qu'il suffit de sonder un échantillon assez grand pour approcher les caractéristiques statistiques de la population réelle. Le théorème de la limite centrale permet, pour sa part, de contrôler l'erreur ainsi commise : l'écart entre la moyenne empirique et la moyenne théorique suit une loi gaussienne.

La moyenne et l'écart-type n'ont plus forcément de sens pour les distributions inégalitaires

LOI DES GRANDS NOMBRES ET THÉORÈME DE LIMITE CENTRALE



Le théorème de la limite centrale permet de quantifier la probabilité que la moyenne empirique approche la moyenne réelle avec une précision donnée.

La loi des grands nombres énonce que lorsqu'on échantillonne une variable aléatoire, sous certaines conditions, lorsque la taille de l'échantillon est assez grande, la moyenne empirique de cet échantillon tend vers la moyenne théorique de la variable aléatoire échantillonnée.
C'est sur cette loi que reposent notamment les sondages. Considérons un échantillon d'observations $x_1, x_2, \ldots x_n$ d'une variable aléatoire X, d'espérance μ et d'écart-type σ finis. La loi forte des grands nombres affirme que quand n tend vers l'infini, la moyenne empirique $M_n = (x_1 + x_2 + \ldots + x_n)/n$ converge presque sûrement vers μ, c'est-à-dire que la probabilité que la moyenne empirique converge vers la moyenne théorique est égale à 1.

$$\mathbb{P}\left(\lim_{n\to+\infty} M_n = \mu\right) = 1$$

En d'autres termes, la moyenne d'un grand nombre d'observations aléatoires indépendantes n'est plus aléatoire ! La loi des grands nombres assure ainsi que la moyenne empirique est un estimateur convergent (ou « consistant ») de l'espérance mathématique. Cela reste vrai tant que la contribution de chaque terme x_i à la moyenne reste négligeable : aucune observation n'est assez grande pour dominer la somme. Cette situation est qualifiée de hasard « sage ».
Le théorème de la limite centrale est l'autre pilier de l'échantillonnage. Considérons des variables aléatoires X_i, indépendantes et de même distribution (d'espérance μ et d'écart-type σ). Alors la somme centrée $(X_1 + X_2 + \ldots + X_n - n\mu)/\sqrt{n}$ tend vers une variable aléatoire normale (ou gaussienne) d'espérance nulle et d'écart-type σ quand n tend vers l'infini. La densité de probabilité d'une variable gaussienne de moyenne m et écart-type σ est définie par la fonction

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

et sa représentation graphique est la célèbre courbe en cloche.
Le théorème de la limite centrale reste valable même si les distributions individuelles sont différentes, pour autant que l'écart-type de chacun des termes soit négligeable vis-à-vis de l'écart-type de la somme.
Ce théorème implique que la différence entre la moyenne empirique M_n de l'échantillon et l'espérance μ de la variable aléatoire X est représentée par une distribution gaussienne dont l'écart-type est proportionnel à $1/\sqrt{n}$. Cela permet de contrôler l'erreur faite en approximant la moyenne théorique μ par la moyenne empirique M_n. La probabilité que M_n se trouve dans l'intervalle $[\mu - t\sigma, \mu + t\sigma]$ est représentée, pour n assez grand, par l'aire sous la courbe en cloche comprise entre les abscisses $\mu - t\sigma$ et $\mu + t\sigma$. Il y a par exemple 95 % de chances que l'écart entre la moyenne empirique M_n et l'espérance μ soit inférieur à 2σ.

Par exemple, si un sondage électoral porte sur 10 000 personnes et que 46 % d'entre elles se déclarent favorables au candidat A, le théorème de la limite centrale indique qu'il y a 95 % de chance que lors du vote, il recueille entre 45 et 47 % des voix. Concrètement, il n'est pas nécessaire de sonder un très grand échantillon de la population pour avoir une estimation fiable.

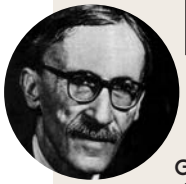
Le théorème de la limite centrale est invoqué pour justifier la représentation par une distribution gaussienne dans des domaines allant des sciences sociales à la physique, en passant par l'économie, la biologie ou la finance. La courbe en cloche est mise à toutes les sauces. On l'utilise même pour « normaliser » la distribution des notes dans les concours…

Cependant, l'omniprésence de la distribution gaussienne en statistique (d'où son qualificatif de « loi normale ») est peut-être due à la trop grande confiance des statisticiens plutôt qu'à la réalité : tout phénomène ne peut pas être ramené à une somme de petites contributions indépendantes. De fait, en sciences économiques et sociales, la loi normale paraît plutôt être l'exception que la règle.

Dès le début du XX^{e} siècle, l'économiste Vilfredo Pareto s'est par exemple intéressé à la richesse nationale italienne, et a constaté que les 20 % les plus riches de la population détenaient 80 % de la richesse totale du pays. Cette situation d'inégalité n'est manifestement pas bien représentée par une distribution gaussienne des richesses, où la richesse individuelle serait concentrée autour de la richesse moyenne et où les individus très riches ou très pauvres seraient très rares.

Pareto a proposé une distribution plus réaliste qui porte aujourd'hui son nom : la part d'individus ayant une richesse supérieure à un niveau u est proportionnelle à $1/u^{\alpha}$. Plus l'exposant α de cette loi de puissance est petit, plus >

LES LOIS STABLES DE LÉVY

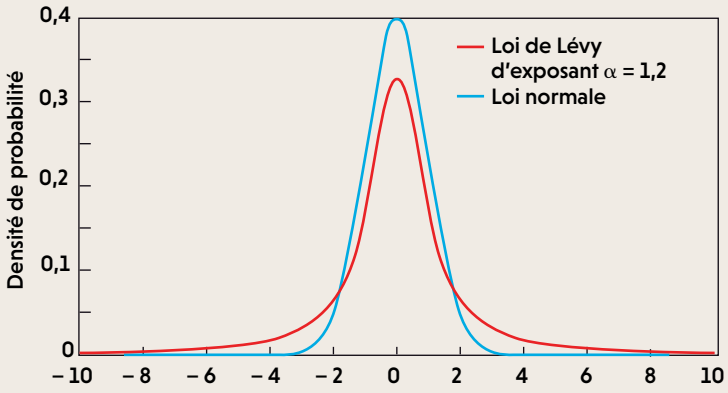


Paul Lévy était un pionnier de la théorie moderne des processus aléatoires. Il découvrit dans les années 1920, en même temps que Boris Gnedenko en Union soviétique, les distributions « stables » qui portent aujourd'hui son nom et étendit le théorème de la limite centrale au cas du « hasard sauvage ».

Soit une suite $X_1, X_2, \ldots, X_n$ de variables aléatoires indépendantes centrées, dont la queue de distribution se comporte comme une loi de Pareto $1/x^{\alpha}$ avec $0 < \alpha < 2$ (l'écart-type est alors infini) ; alors la somme $X_1 + X_2 + \ldots + X_n$ n'obéit plus au théorème de la limite centrale.

La distribution de $(X_1 + X_2 + \ldots + X_n)/n^{1/\alpha}$ se comporte, pour n assez grand, non comme une loi gaussienne, mais comme une « loi stable de Lévy » d'exposant. À la différence du théorème de la limite centrale, le facteur de normalisation est ici $n^{1/\alpha}$ au lieu de n.

Les distributions stables n'ont pas de formules analytiques explicites, sauf dans le cas $\alpha = 1$ (loi de Cauchy) et $\alpha = 1/2$. Leur queue de distribution se comporte comme $1/|x|^{\alpha}$ pour $|x|$ assez grand : la somme a la même queue de distribution que chacun de ses termes.



> cette proportion est grande et plus la distribution des richesses est inégalitaire. Face à de telles distributions, les notions de moyenne et d'écart-type se révèlent inadaptées et offrent une description trompeuse.

LES MATHS DES INÉGALITÉS

En effet, la richesse moyenne et l'écart-type peuvent par exemple rester identiques alors que la distribution des richesses se modifie (ce qui est impossible dans le cadre d'une distribution gaussienne).

Pire, si l'exposant de la loi de Pareto est inférieur à deux, alors l'écart-type de la distribution est infini. S'il est inférieur à un, c'est aussi le cas pour la moyenne. Ces indicateurs sont donc dénués de sens pour les lois de Pareto.

Ce fait a conduit les économistes à définir d'autres indicateurs pour mieux caractériser les distributions inégalitaires. Le coefficient de Gini, proposé par le statisticien italien Corrado Gini en 1912, est ainsi un nombre compris entre 0 et 1 qui mesure le degré d'inégalité de la distribution des richesses. Plus il est proche de 1, plus la répartition est inégalitaire. Dans le cas d'une loi de Pareto d'exposant α, le coefficient de Gini vaut $1/(2\alpha - 1)$. Ces indicateurs se focalisent sur les plus grands événements dans un échantillon, événements d'occurrence faible, mais d'importance prépondérante. Ces « événements rares » forment des queues de distributions.

Dès lors que la moyenne et l'écart-type n'ont plus forcément de sens pour les distributions inégalitaires, qu'advient-il de la loi des grands nombres et du théorème de la limite centrale ? Dans les années 1920, Lévy fut l'un des premiers à remarquer que les hypothèses qui les sous-tendent peuvent cesser d'être valables dans le cas où des événements rares influent de façon prépondérante sur la moyenne de l'échantillon.

Lévy étendit le théorème de la limite centrale à ces situations : il montra que pour un échantillon de variables centrées dont la queue de distribution se comporte comme une loi de Pareto avec un exposant compris entre 0 et 2, la moyenne empirique se comporte non plus comme une variable gaussienne, mais selon une distribution que Lévy appela loi stable, dont la queue se comporte comme une loi de Pareto de même exposant. En d'autres termes, la loi limite est une loi de Lévy, et non une loi normale.

Pour ces distributions de Lévy, variance et écart-type n'ont plus de sens : ils sont infinis. Benoît Mandelbrot, qui fut élève de Lévy à l'École polytechnique, montra en 1963 que, loin d'être une curiosité mathématique, les situations décrites par les lois de Lévy étaient omniprésentes en économie, en finance et en assurance : les rendements des matières premières, la richesse des individus ou la taille des entreprises suivent des lois de Pareto.

Mandelbrot qualifia ces situations, où le comportement de l'échantillon n'est plus correctement caractérisé par la moyenne et l'écart-type, de hasard sauvage, par opposition au hasard sage de la loi gaussienne.

DOMPTER LE HASARD SAUVAGE

Ce comportement sauvage est manifeste dans les rendements d'indices boursiers (*voir l'encadré page 27*). Alors que l'amplitude des fluctuations d'un échantillon gaussien reste de l'ordre de son écart-type, certains indices boursiers présentent des variations d'amplitude erratiques atteignant plusieurs dizaines de fois l'écart-type ! De nombreuses autres situations se conformant aux lois de Lévy ont été rapidement mises en évidence en physique statistique.

L'épithète « sauvage » suggère l'incontrôlabilité du hasard, mais une bonne compréhension de l'aléa aide parfois à l'apprivoiser. C'est le cas dans certains systèmes physiques, ainsi que dans des situations, où,

indépendamment du fait que l'aléa sous-jacent soit sage ou sauvage, la moyenne ou l'écart-type d'un échantillon ne sont tout simplement pas les indicateurs pertinents, même si on peut les estimer avec précision.

Il en est ainsi des études de fiabilité, où, pour déterminer la probabilité de défaillance d'un système, on cherche à estimer celle de son « maillon le plus faible ». Par exemple, pour choisir la dimension et les caractéristiques d'un barrage, il faut avoir une idée précise de la pression maximale que l'ouvrage pourra être amené à supporter, et non de la pression moyenne qu'il subira en conditions normales. Pour cela, il faut caractériser directement les valeurs extrêmes dans l'échantillon d'observations. Les notions d'écart-type, de moyenne et le théorème de la limite centrale ne sont alors pas d'un grand secours...

Parallèlement aux travaux de Lévy sur le théorème de la limite centrale, un groupe de statisticiens, dont sir Ronald Fisher et Leonard Tippett en Grande-Bretagne, Boris Gnedenko en Union soviétique, Maurice Fréchet en France et Ludwig von Mises en Autriche, ainsi qu'Emil Julius Gumbel, mathématicien allemand réfugié en France, développaient une nouvelle branche des statistiques dont le but est d'étudier non plus les valeurs moyennes des échantillons, mais leurs valeurs extrêmes, c'est-à-dire maximales ou minimales.

De même que le théorème de la limite centrale décrit le comportement de la moyenne d'un grand échantillon, cette théorie des valeurs extrêmes décrit le comportement du maximum d'un échantillon. Ces statisticiens ont montré que, selon la nature de la variable aléatoire étudiée, la valeur maximale de l'échantillon ne peut obéir qu'à l'une parmi trois distributions différentes : les distributions de Fréchet, de Gumbel ou de Weibull (*voir l'encadré ci-contre*).

Comme la loi normale et la loi de Lévy, qui sont des candidats naturels pour modéliser les sommes de variables indépendantes, les lois de valeurs extrêmes sont des candidats privilégiés pour la modélisation des valeurs maximales d'un échantillon. Gumbel présenta en 1941 une application de ces concepts à l'analyse statistique des crues fluviales. Cette théorie allait être mise en application dans un contexte dramatique.

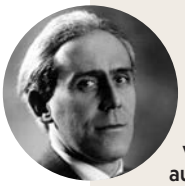
APPRIVOISER LA RARETÉ... ET LA MER

Les Pays-Bas ont toujours vécu sous la menace de la montée des eaux. Les nombreuses digues qui parsemaient le pays ont longtemps suffi à le protéger. Mais dans la nuit du 31 janvier au 1er février 1953, une montée des eaux due à une tempête eut raison de leur résistance et submergea le Sud-Ouest du pays, tuant plus de 1800 personnes et des milliers de têtes de bétail.

Sous le choc, le gouvernement néerlandais lança peu après un vaste projet de construction de barrages. La commission chargée de revoir les normes de sécurité exigea de fixer la hauteur des digues de façon à ce que la probabilité que la marée les dépasse soit inférieure à 1/10000. Une équipe de scientifiques, armée de la récente théorie des valeurs extrêmes, se mit alors à étudier marées et crues passées afin d'estimer la distribution de probabilité de la hauteur maximale des inondations. Ils montrèrent qu'elle suit une loi de Gumbel, et que la hauteur requise dépassait 5 mètres, une nette augmentation par rapport aux digues de l'époque. Sur la base de ces estimations, d'énormes barrages furent construits (*voir la photo page 26*) ; ces édifices protègent encore le pays aujourd'hui.

Dans une ambiance de foi grandissante dans les méthodes quantitatives, la théorie des valeurs extrêmes fut par la suite appliquée aux >

LES LOIS DE VALEURS EXTRÊMES

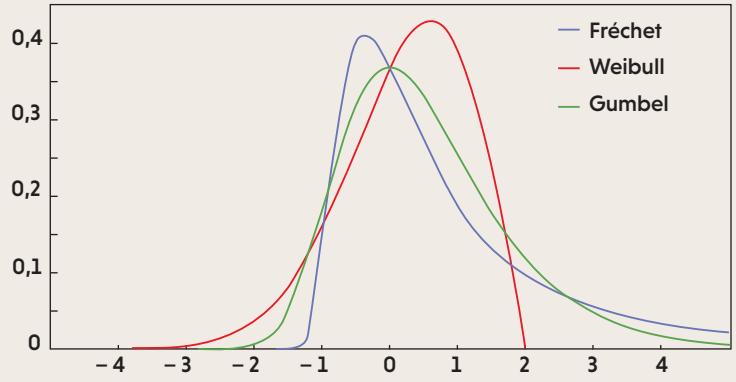


Considérons un échantillon $x_1, x_2, \ldots, x_n$ de tirages indépendants d'une distribution de probabilité G. La théorie des valeurs extrêmes s'intéresse au comportement de la valeur maximale $U_n = \max(x_1, x_2, \ldots, x_n)$ d'un échantillon de grande taille (n grand). Un des résultats fondamentaux de la théorie, le théorème de Fisher-Tippett-Gnedenko, montre que s'il existe des constantes de normalisation a_n et b_n telles que la distribution de $(U_n - a_n)/b_n$ ait une limite quand n tend vers l'infini, alors la distribution limite de la suite normalisée de maxima des variables aléatoires est nécessairement de la forme :

$$F(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

définie pour $1+\xi(x-\mu)/\sigma > 0$, et $\sigma > 0$.

On distingue alors trois cas.
Le premier, quand $\xi < 0$, nommé distribution de Weibull, décrit le cas où les variables x_i sont bornées.
Le second, lorsque $\xi = 0$, correspond à la distribution de Gumbel. Elle est au maximum d'un échantillon ce que la loi normale est à la moyenne : elle représente le cas où le hasard décrit par la distribution G est sage.
Enfin, si les variables x_i suivent une loi de Pareto d'exposant α, alors $\xi = 1/\alpha > 0$. La distribution, dite de Fréchet, décrit le maximum d'un échantillon qui obéit au hasard sauvage. Plus ξ est grand, plus la queue de distribution est importante, c'est-à-dire plus les événements extrêmes sont fréquents.
Ces trois lois de probabilité sont adaptées aux distributions de valeurs extrêmes.



> assurances, à la gestion des risques financiers et celle des risques environnementaux. Les assureurs l'utilisent par exemple pour estimer les plus grosses pertes qu'ils peuvent subir en vendant des assurances contre les catastrophes naturelles.

Le scénario catastrophe par excellence étant l'explosion d'une centrale nucléaire, toutes les idées sur la mesure des risques extrêmes ont trouvé tôt ou tard un terrain d'application dans le domaine de la sûreté nucléaire. Dans les années 1970, sont apparues les études probabilistes de sûreté (EPS), une méthodologie d'estimation de la probabilité des scénarios d'accident dans les installations nucléaires. La méthode consiste à identifier les séquences d'événements pouvant mener à un résultat catastrophique, telle la fusion du cœur de réacteur ; et assigner des probabilités à ces séquences à partir de principes physiques ou de l'expérience acquise. L'objectif, comme pour les digues hollandaises, est de définir les normes à satisfaire pour garantir une probabilité d'occurrence de l'événement catastrophique inférieure à un seuil donné.

Mais en 1979, l'accident de la centrale nucléaire de Three Mile Island, en Pennsylvanie, aux États-Unis, rappela brutalement que, dans un système complexe formé de multiples composantes interdépendantes comme une centrale nucléaire, une petite erreur – humaine en l'occurrence – peut provoquer une catastrophe (*voir la figure pages 20-21*).

Or l'occurrence de telles erreurs, qui ne sont pas liées à un facteur physique obéissant à une loi bien identifiée, est difficile, voire impossible, à quantifier. Dès lors, quelle foi accorder aux « certitudes quantitatives » produites par les études de sûreté et sur les certifications techniques qui en résultent ?

Frank Knight, économiste américain et théoricien du risque, distinguait déjà en 1921 dans *Risk, Uncertainty and Profit* le risque probabilisable, correspondant à des événements dont on peut évaluer la probabilité avec une certitude raisonnable, et l'incertitude, qui recouvre des événements dont on ignore jusqu'aux probabilités d'occurrence. Lorsqu'on s'éloigne des systèmes physiques pour aller vers des domaines où les facteurs humains sont plus importants, la part de l'incertitude augmente.

Données d'observation incomplètes, hypothèses abusivement simplificatrices (telle l'indépendance des facteurs de risques, fréquemment postulée dans les modèles statistiques), ou encore omission de certains facteurs de risque, les sources d'incertitudes dans les modèles de risque sont nombreuses. Elles sont autant de limitations qui invitent à rester modestes dans nos attentes envers les modèles statistiques lorsque ceux-ci ne s'appuient pas sur des principes physiques bien ancrés.

LA CHASSE AU CYGNE NOIR

Que faire alors ? Tout au plus peut-on compléter les résultats des modèles par l'analyse au cas par cas de scénarios extrêmes. C'est l'objet de la méthode de *stress test*, qui apporte un complément utile aux analyses statistiques de risque. Le CEA et EDF ont retenu cette option en complément des études EPS pour définir la politique de sûreté nucléaire.

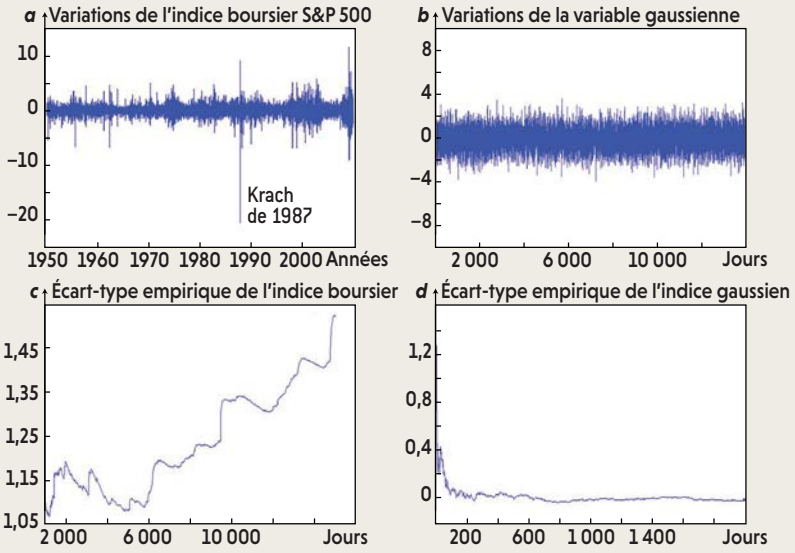
Les statistiques montrent là leurs limites. Faut-il pour autant abandonner l'ambition de quantifier les événements rares en dehors des sciences physiques, où les principes physiques permettent de calculer leurs probabilités ?

C'est ce que certains semblent suggérer, tels Nassim Taleb ou Nicolas Bouleau, de l'École nationale des ponts et chaussées, à la

Le barrage d'Oosterschelde, aux Pays-Bas, a été conçu pour que la probabilité qu'une marée le dépasse soit inférieure à 1/10 000. La théorie des valeurs extrêmes indiqua que la hauteur des digues devait être de 5 mètres au moins.

HASARD SAUVAGE CONTRE HASARD SAGE

Benoît Mandelbrot *(ci-contre)* fut le premier à mettre en évidence dans les années 1960 les manifestations du hasard sauvage en économie et en finance. On peut illustrer la différence entre hasard sauvage et hasard sage en comparant les rendements journaliers d'un indice boursier réel (l'indice S&P 500 de la Bourse de New York) de janvier 1950 à septembre 2009, et un échantillon de 14 000 observations simulées d'une variable aléatoire gaussienne de même moyenne (0) et même écart-type (1 %) que les données réelles. Sur le graphique de l'indice boursier *(a)*, on voit que la variabilité est grande et que l'indice connaît des bouffées de volatilité. En octobre 1987, l'indice a perdu 20 % de sa valeur en une journée, soit 20 fois son écart-type ! *A contrario*, les variations de la série gaussienne *(b)* restent concentrées autour de la moyenne : la plus grande perte est de 4 %, soit seulement 4 fois l'écart-type journalier. Les écarts-types empiriques des deux séries évoluent aussi de façons très différentes en fonction du nombre d'observations. Alors que l'écart-type de l'échantillon gaussien *(d)* converge doucement vers sa valeur théorique, comme le prévoit la loi des grands nombres, l'écart-type empirique des rendements de l'indice boursier *(c)* continue à osciller à mesure que de nouvelles observations sont ajoutées à l'échantillon, sans donner aucun signe de convergence. Ces observations suggèrent que les aléas financiers relèvent du hasard sauvage, domaine où les événements extrêmes jouent un rôle déterminant. Mandelbrot proposa ainsi de modéliser les mouvements boursiers à l'aide des distributions stables de Lévy.



L'indice boursier S&P 500 obéit à un hasard sauvage, alors qu'une variable aléatoire gaussienne suit un hasard sage.

lumière notamment de l'apparent échec des méthodes quantitatives de gestion de risques financiers lors de la crise financière dont nous avons parlé. Selon Nassim Taleb, il est en effet impossible de quantifier le risque d'événements rares catastrophiques dont nous n'avons jamais connu d'exemple par le passé. Ces « cygnes noirs », tels qu'il les qualifie, invalideraient les approches statistiques de modélisation du risque.

ÉVITER LA CRISE ?

Cependant, la récente crise financière est-elle un cygne noir au même titre que l'accident de Three Mile Island, où, malgré les précautions, l'erreur humaine, non quantifiable, mit en échec le système de sûreté ? Ou ressemble-t-elle plutôt à la tempête qui ravagea les Pays-Bas en 1953, événement rare, mais dont on pouvait mesurer la probabilité ?

Je penche en faveur de la seconde option : les institutions financières sont loin d'avoir mis en pratique des méthodes adéquates pour mesurer les risques de leurs portefeuilles. Plus de quarante ans après les travaux de Mandelbrot, nombre d'établissements bancaires, par facilité, utilisent encore la loi normale dans leurs calculs de risques : difficile alors de blâmer les modèles statistiques… qui ne sont pas utilisés !

Sans aller jusqu'à affirmer, comme le faisait récemment un journaliste du *Nouvel Observateur*, que si la banque d'affaires Lehman Brothers avait calculé ses risques avec les distributions de Lévy, elle aurait évité la faillite, il est certain que l'utilisation de modèles plus réalistes ne pourra qu'améliorer la gestion des risques financiers.

Mais Nicolas Bouleau soulève un autre point, d'ordre sociologique. Les énormes enjeux sociaux de certains événements catastrophiques – météorologiques, géologiques ou économiques – ont pour conséquence d'engendrer une pression importante des décideurs publics sur les « experts » pour, sinon prévoir ces événements rares, au moins quantifier leur risque d'occurrence. La tentation est alors grande de préférer des indicateurs statistiques inadaptés plutôt que de reconnaître que les connaissances scientifiques actuelles n'apportent par de réponse à certaines questions.

C'est sans doute un aspect de la défaillance des systèmes de gestion des risques financiers lors de la crise récente : indépendamment de leur précision, ils ont entretenu l'illusion d'une sécurité aux yeux de leurs utilisateurs, qui ont de ce fait baissé leur garde. Ne rejetons pas les méthodes statistiques de modélisation des risques, utilisons-les avec discernement ! ■

BIBLIOGRAPHIE

P. DEHEUVELS, *La Probabilité, le hasard et la certitude*, « Que Sais-Je ? », Presses Universitaires de France, 2008.

B. MANDELBROT ET R. HUDSON, *Une approche fractale des marchés. Risquer, perdre et gagner*, Odile Jacob, 2005.

F. BARDOU, J.-PH. BOUCHAUD, A. ASPECT ET C. COHEN-TANNOUDJI, *Lévy Statistics and Laser Cooling: How Rare Events Bring Atoms to Rest*, Cambridge University Press, 2002.

N. BOULEAU, Splendeurs et misères des lois de valeurs extrêmes, *Risques*, vol. 3, pp. 85-92, 1991.

L. DE HAAN, Fighting the arch-enemy with mathematics, *Statistica Neerlandica*, vol. 44(2), pp. 45-68, 1990.

L'ESSENTIEL

- Certains résultats fracassants en matière de statistiques sont parfois fondés sur de petits effets difficiles à interpréter.
- La pertinence statistique de tels résultats requiert une estimation précise des incertitudes et une méthode d'analyse adaptée.
- Les analyses fréquentistes et bayésiennes sont parmi les méthodes les plus fiables.
- Autre précaution nécessaire: une estimation des petits effets n'a de sens que si la taille des échantillons est grande.

LES AUTEURS

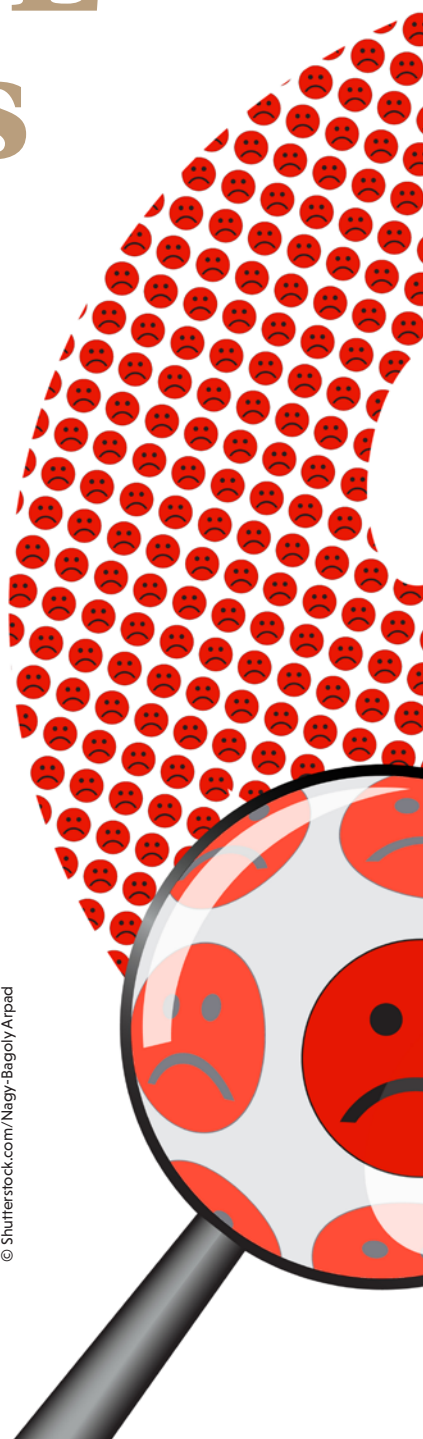
ANDREW GELMAN
professeur de statistiques et de sciences politiques, dirige le Centre de statistiques appliquées de l'université Columbia, à New York.

DAVID WEAKLIEM
est professeur de sociologie à l'université du Connecticut.

Nous remercions la revue *American Scientist* de nous avoir autorisés à publier cet article.

LE CASSE-TÊTE des petits effets

On accorde parfois à de petites différences un sens qu'elles n'ont pas. Distinguer les petits effets statistiquement pertinents de ceux relevant du hasard est un défi !



© Shutterstock.com/Nagy-Bagoly Arpad

« **Les ingénieurs ont plus de garçons, les infirmières plus de filles », «Les hommes violents ont plus de garçons», «Les individus séduisants ont plus de filles»...** On doit ces phrases définitives à Satoshi Kanazawa, psychologue évolutionniste à l'École d'économie de Londres. L'auteur est certes controversé, mais il a néanmoins publié ces «conclusions» pour le moins insolites dans des revues scientifiques sérieuses. L'un de ses derniers articles, parus en 2017, est intitulé «Les individus sont d'autant moins séduisants que leur père était âgé au moment de la conception». Qu'en pensez-vous?

Les spécialistes ont rendu leur avis: les analyses statistiques sur lesquelles ces articles étaient fondés ont été invalidées, par exemple à cause de biais d'échantillonnage. Ainsi, ces «découvertes» ne sont pas significatives et peuvent n'être que le fruit du hasard: elles n'auraient jamais été publiées si leur pertinence statistique avait été correctement évaluée.

Faut-il pour autant rejeter en bloc ces travaux? Non, et d'ailleurs, certaines hypothèses de Satoshi Kanazawa s'appuient sur des théories reconnues par la communauté scientifique. Cependant, le bulldozer de la médiatisation a écrasé toute nuance et précaution pourtant indispensables à l'interprétation des petites différences statistiques relevées par le psychologue. Elles relèvent de ce que l'on nomme les petits effets. Dans quelle mesure ces derniers ont-ils un sens statistique? Comment interpréter des résultats non significatifs (et que l'on cherche malgré tout à interpréter)? Quelles peuvent être les conséquences d'une interprétation erronée de ces petits effets?