

**Zeitschrift:** Horizons : le magazine suisse de la recherche scientifique  
**Herausgeber:** Fonds National Suisse de la Recherche Scientifique  
**Band:** 25 (2013)  
**Heft:** 98

**Artikel:** Point fort statistique : penser en chiffres  
**Autor:** Amrhein, Valentin / Hafner, Urs / Koechlin, Simon  
**DOI:** <https://doi.org/10.5169/seals-554030>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

### **Terms of use**

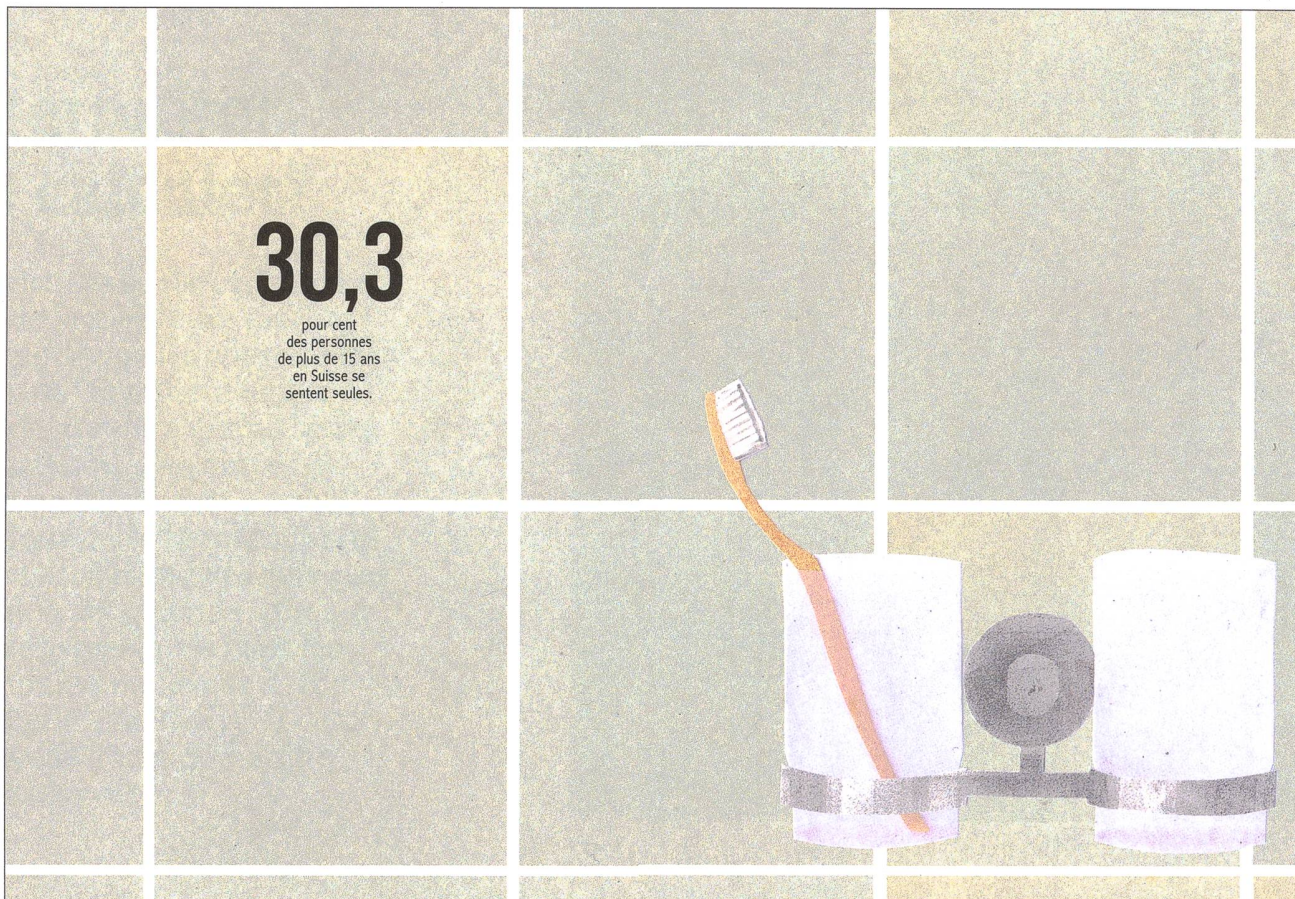
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

**Download PDF:** 08.02.2025

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

**30,3**

pour cent  
des personnes  
de plus de 15 ans  
en Suisse se  
sentent seules.



**1776**

est la dernière  
année au cours  
de laquelle  
aucune nation  
n'était en guerre  
contre une autre.



Des chiffres pertinents? Johanna Schaible, Haute école de Lucerne, design & art.

# 80 000

personnes ont  
renoncé à suivre  
Justin Bieber  
sur Twitter  
en raison de sa  
nouvelle coiffure.



## Penser en chiffres

La statistique s'est imposée dans le monde scientifique et dans la vie quotidienne. En traitant d'énormes quantités de données, elle est censée explorer la réalité de manière approfondie et permettre des pronostics plausibles. Elle fournit toutefois souvent des images biaisées.

**Le monde sous forme de diagrammes.** Les infographies simplifient des thèmes complexes et permettent de visualiser des proportions au moyen de gâteaux, de bâtons ou de colonnes. Dans les médias et les articles scientifiques, nous pouvons ainsi rapidement saisir de quoi il retourne. Les graphiques ne sont toutefois ni neutres ni objectifs. Ils déterminent la manière dont nous appréhendons la réalité, souvent sans divulguer leurs prémisses. C'est ce que mettent en évidence les illustrations sur cette page et sur les suivantes. Elles nous font croire que le monde sous forme de diagrammes est simple et compréhensible.



D'immenses quantités de données attendent d'être intelligemment analysées par des statisticiens. Une profession en plein essor. La statistique continue pourtant d'être considérée comme un mal nécessaire.

Par Valentin Amrhein

## L'art de conjecturer

«Le métier de statisticien sera regardé comme le plus sexy au cours des dix prochaines années.» Ainsi parlait Hal Varian, professeur de technologies de l'information et économiste en chef chez Google, en 2008. Car la montagne de données qui nous fait face est d'une ampleur inimaginable. Aux Etats-Unis, la Bibliothèque du Congrès sauvegarde toutes les informations publiques diffusées sur Twitter. Objectif: constituer une base pour des études sur le comportement social humain. Le stock comporterait actuellement 170 milliards de tweets.







La National Security Agency (NSA), Google, Migros et Coop, tous ces acteurs n'ont jamais eu aussi facilement accès aux informations. Car nous présentons de bon gré nos cartes clients, nous laissons nos smartphones communiquer. Et sauvegarder automatiquement des données coûte de moins en moins cher. Dans les laboratoires de recherche aussi, c'est désormais moins l'art d'obtenir des données qui compte que celui de bien rassembler les bonnes, de les analyser judicieusement, de procéder à une interprétation fiable des résultats et de les représenter de manière intelligible. Autant de tâches qui nécessitent des compétences de statisticien. Le cabinet de conseil McKinsey estime que rien qu'aux Etats-Unis, d'ici 2018, il manquera près de 200 000 experts dans ce domaine.

«Malheureusement, bon nombre de gens considèrent la statistique davantage comme un mal nécessaire que comme un outil génial», déplore Beat Hulliger, profes-

seur de recherche économique et sociale à la Haute école spécialisée de la Suisse du Nord-Ouest. Selon lui, on peine à aborder des phénomènes où incertitude et probabilités jouent un rôle de premier plan. Alors que les êtres humains sont en réalité des machines à prédire: nous devons sans cesse estimer quand l'eau se mettra à bouillir, de combien de temps nous aurons besoin pour faire nos courses, et si nos clients aimeront encore tel produit au cas où il renchérit. Pour toutes ces décisions, nous mobilisons des données issues de notre savoir théorique, de notre expérience pratique et des conditions actuelles, pour en déduire une probabilité quant à la survenance de tel fait. Un processus purement statistique. «Aux Etats-Unis, le Business Analytics, c'est-à-dire l'analyse des données opérationnelles d'une entreprise au moyen de méthodes statistiques avancées revêt un grande importance, relève Beat Hulliger. Et les résultats sont intégrés à la planification des affaires.» En Suisse, pour de nombreuses entreprises, tout cela est encore très lointain.

Pourtant, la statistique moderne a ses origines en Suisse. C'est à la fin du XVII<sup>e</sup> siècle que le mathématicien et physicien bâlois Jakob Bernoulli a fondé la théorie des probabilités. A l'occasion du 300<sup>e</sup> anniversaire de la publication de son œuvre phare intitulée *Ars Conjectandi*, l'art de conjecturer, la Swiss Statistical Society organise à la mi-octobre un congrès à Bâle. 2013 a même été déclarée Année internationale de la statistique ([www.statistics2013.org](http://www.statistics2013.org)).

Vue aérienne fictive montrant comment le canton de Berne utilise ses surfaces agricoles. Montage de Tobias Gutmann, Haute école des arts de Berne.

	Céréales (26 607 ha)
	Pommes de terre, betteraves (8523 ha)
	Oléagineux (2288 ha)
	Surfaces herbagères (140 119 ha)
	Cultures pérennes et autres surfaces agricoles utiles (12 785 ha)
	1 bâtiment correspond à 10 exploitations bio (1196 exploitations)

# La Terre est ronde ( $p < 5\%$ )

A quoi servent les statistiques et que veut dire «significativité»? Alors que la statistique du XXe siècle touche à ses limites, les idées d'un prêtre anglais sont de nouveau d'actualité, malgré leurs 250 ans d'âge. *Par Valentin Amrhein*

Comment un médecin s'y prend-il pour poser un diagnostic? Evaluer le risque de développer telle maladie ou le déroulement d'une affection? Il se base sur les données d'anciens patients. «Tous les processus décisionnels cliniques sont fondés sur la statistique», déclare Andreas Papassotiropoulos, chef du département des neurosciences moléculaires à l'Université de Bâle. Avant d'ajouter aussitôt: «Mais chez les médecins, comme chez les chercheurs en biologie et en médecine, les connaissances dans ce domaine recèlent toujours d'effrayantes lacunes.» A l'image de ce chercheur qui présentait des mesures réalisées sur une souris génétiquement modifiée et sur une souris normale. Lorsqu'on lui a demandé sur combien de souris il avait effectué ses mesures, le chercheur a répondu: «Juste sur ces deux-là. On voit bien qu'il y a une différence.»

Pourquoi aurait-il besoin de se former à la statistique? Prenons un exemple: nous aimerions savoir si, en Suisse, les hommes sont plus grands que les femmes. Le moyen le plus simple d'avoir la réponse, comme dans le cas des souris, consiste à considérer un individu de chaque sexe au sein de la population suisse. Mais il se pourrait que, par hasard, l'homme choisi soit plus petit que la femme sélectionnée. De notre observation, nous concluons à tort qu'en général les hommes sont plus petits que les femmes. C'est pourquoi, d'ordinaire, les chercheurs se penchent sur des échantillons plus importants, et mesurent, par exemple, la taille de 50 individus de chaque sexe considérés au hasard. Mais comment s'y prendre pour comparer les données issues de ces mesures? Les comparer une à une n'a guère de sens. Nous devons les simplifier, notamment en mettant en parallèle la moyenne des hommes et celle des femmes.

## Une moyenne peu significative

La moyenne constitue dès lors un modèle statistique qui ne correspond pas à la réalité. Car aucun individu, en Suisse, n'atteint la taille moyenne au nanomètre près. «On ne peut donc pas dire avec certitude à un patient comment il ira demain, en se basant sur l'évolution moyenne de cas similaires», souligne Andreas Papassotiropoulos.

Gérer ce fait implique une certaine conscience statistique, tant du côté du médecin que du patient. La moyenne est un bon modèle, mais les médecins doivent toujours le compléter avec les données individuelles de chaque patient. Et avec l'avènement de la médecine personnalisée, ces données sont toujours plus nombreuses. La statistique est la compagne permanente de la recherche et de la pratique médicale, et il faut être prêt à se former toute sa vie. Ainsi, le chercheur bâlois estime que dans les hautes écoles et les universités, la formation à la statistique devrait se poursuivre pendant toutes les études.

## Statistique inductive

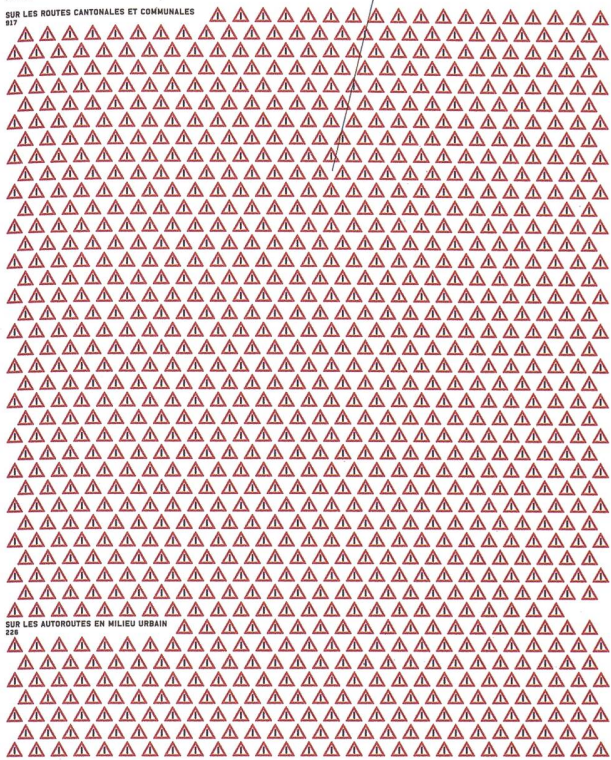
Après tout, on peut même trouver du plaisir à en faire. Un constat qui, dans un projet de recherche, s'impose au plus tard lorsqu'une analyse statistique devient inévitable. Car la statistique est bien davantage que le calcul de statures moyennes et l'art de présenter clairement des données. A côté de cette statistique, dite descriptive, il existe une statistique inductive, qui permet de vérifier des hypothèses scientifiques et de quantifier la «significativité» d'un résultat. Mais que veut dire «significativité statistique»?

Revenons à l'exemple de la taille des gens. Le problème fondamental des échantillons réside dans ce qu'on appelle l'erreur d'échantillonnage. Il se pourrait ainsi que tout à fait par hasard nous n'ayons sélectionné que des hommes de grande taille et des femmes de petite taille. Un simple test statistique permet de voir si la différence entre les valeurs moyennes des 50 femmes et celles des 50 hommes est suffisamment fiable pour qu'il soit effectivement possible de tirer des conclusions sur l'ensemble des personnes en Suisse.

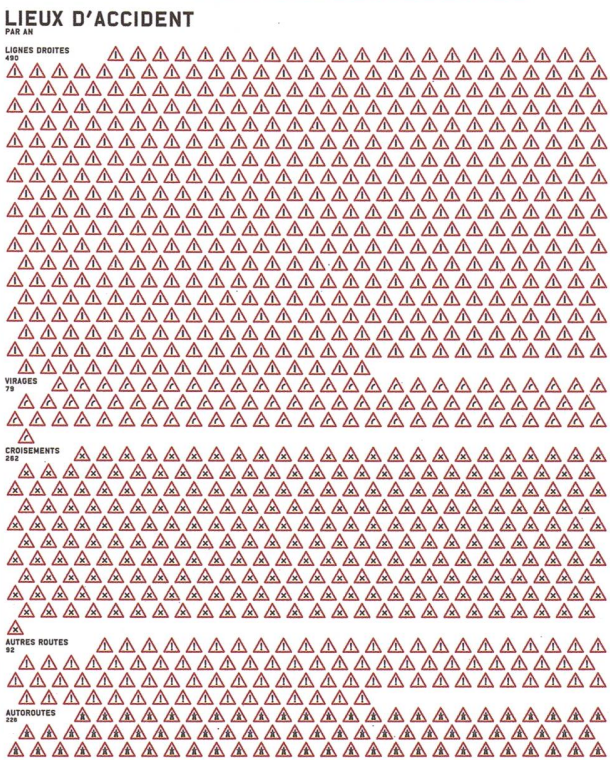
Comprendre le concept de significativité implique un effort intellectuel. Une analyse statistique part de ce qu'on appelle l'hypothèse zéro. Dans notre cas, cette hypothèse zéro serait: les hommes et les

Statistique des accidents de la ville de Berne, 2008. Kaspar Allenbach, Haute école des arts de Berne.

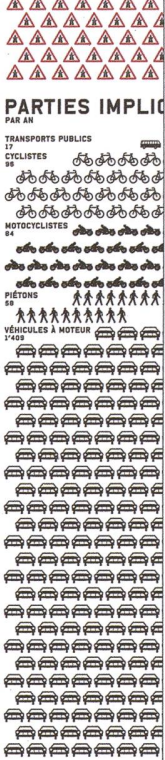
# NOMBRE D'ACCIDENTS



# LIEUX D'ACCIDENT

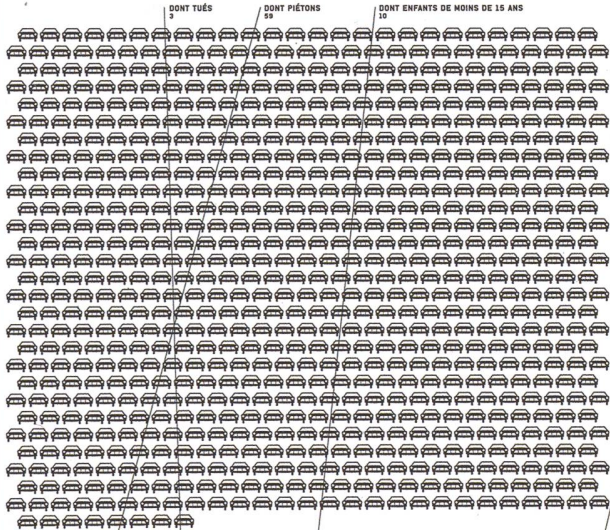
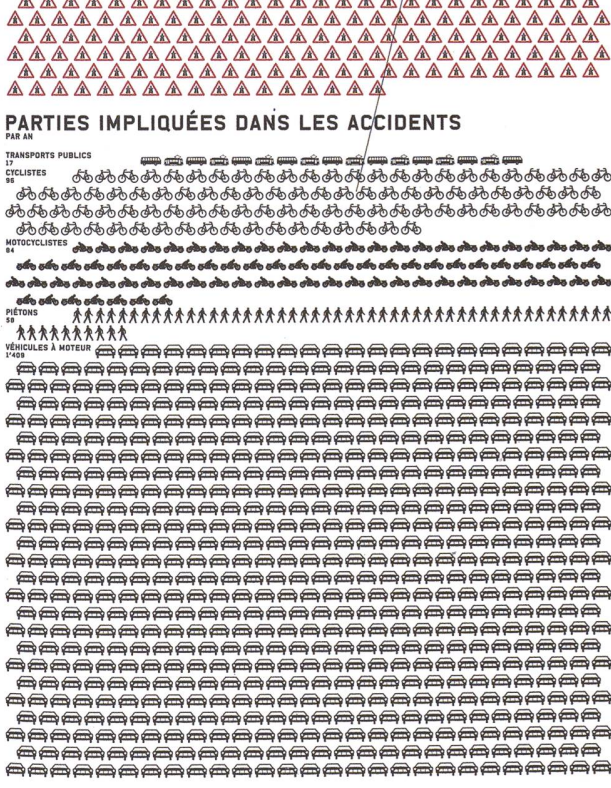


# PARTIES IMPLIQUÉES

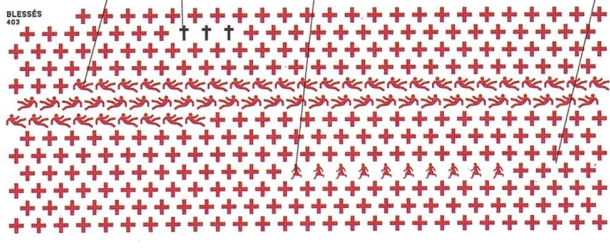


ROUTES CANTONALES ET COMMUNALES 86.65M

# PARTIES IMPLIQUÉES DANS LES ACCIDENTS



# BLESSÉS



femmes ont exactement la même taille. Un modèle statistique permet alors de simuler quelle serait la différence qui apparaîtrait entre les valeurs moyennes des tailles des hommes et des femmes si nous prenions à volonté des échantillons de 50 hommes et de 50 femmes. Si la différence de taille que nous avons effectivement déterminée apparaît dans moins de 5% des cas, on dit que le résultat est significatif: la différence de taille que nous avons trouvée serait très invraisemblable au cas où, dans la réalité, les femmes et les hommes étaient bel et bien de la même taille.

La plupart des études en sciences naturelles, sociales ou économiques, en psychologie ou en médecine n'ont la chance d'être publiées que si les résultats sont significatifs. C'est la raison pour laquelle de nombreux chercheurs considèrent la statistique comme nécessaire. Il n'y a qu'un problème: un résultat significatif ne répond malheureusement pas à la question posée par la recherche. Car avec un test de significativité, nous avons seulement vérifié la probabilité des différences entre les valeurs moyennes, si l'hypothèse zéro est exacte (dans la réalité, les femmes et les hommes ont la même taille). Mais comme pour la plupart des hypothèses zéro, il était d'emblée clair, pour celle-là aussi, qu'elle ne se vérifierait pas: où trouve-t-on en effet deux groupes d'êtres vivants ayant exactement la même taille? D'autre part, ce n'était pas l'hypothèse zéro qui nous intéressait, mais celle qui dit que les hommes sont plus grands que les femmes. Or, malheureusement, la statistique classique de l'hypothèse zéro ne nous fournit aucune information sur la probabilité que cette hypothèse soit correcte.

#### La valeur de $p$

Dans la plupart des cas, la statistique déductive ne fait donc pas ce qu'elle devrait. Pourtant, chercheurs et profanes, journalistes et lecteurs sont particulièrement soumis au mot «significatif» et à ce qu'on appelle la valeur de  $p$ , qui doit être inférieure à 5% pour être significative ( $p$  désigne la probabilité de notre résultat, ou d'un résultat encore plus extrême si l'hypothèse zéro est correcte). La critique du test de l'hypothèse zéro est presque aussi ancienne que la valeur de  $p$ , inventée au début du XXe siècle. Jacob Cohen, grand

psychologue et statisticien, a fort brillamment pointé son absurdité et la tendance des scientifiques à lui accorder trop d'importance, dans sa publication critique «The earth is round ( $p < 5\%$ )», parue en 1994 dans la revue *American Psychologist*.

#### Trois grands avantages

La statistique classique s'éloigne donc du test de l'hypothèse zéro et se concentre plutôt sur l'identification de schémas au cœur des données pour comprendre les processus qui les gouvernent. Depuis 250 ans environ, il existe une alternative au test de significativité. Elle est l'œuvre de Thomas Bayes, mathématicien et ecclésiastique anglais. Cette forme de statistique présente trois grands avantages. Première-

**«La statistique bayésienne permet de calculer des modèles beaucoup plus complexes.»**

Penelope Vounatsou

ment, elle rend possible l'intégration des connaissances préliminaires. Il n'est donc pas nécessaire, à chaque saisie de données, de faire comme si personne n'avait jamais étudié le sujet. Deuxièmement, elle permet d'obtenir ce que l'on veut vraiment, à savoir une indication de la probabilité que notre hypothèse soit exacte.

«Le troisième avantage, explique Penelope Vounatsou, statisticienne à l'Institut tropical et de santé publique suisse à Bâle, c'est que la statistique bayésienne permet de calculer des modèles beaucoup plus complexes que les méthodes statistiques classiques.» Les modèles statistiques de l'Institut tropical calculent, par exemple, pour 10 000 villages, combien de personnes contracteront probablement telle maladie à tel moment dans chacun d'entre eux. Une donne face à laquelle les méthodes statistiques classiques seraient dépassées. «Dans les disciplines comme la protéomique ou la génomique, la statistique bayésienne représente l'avenir», poursuit la statisticienne. Ces domaines analysent en effet

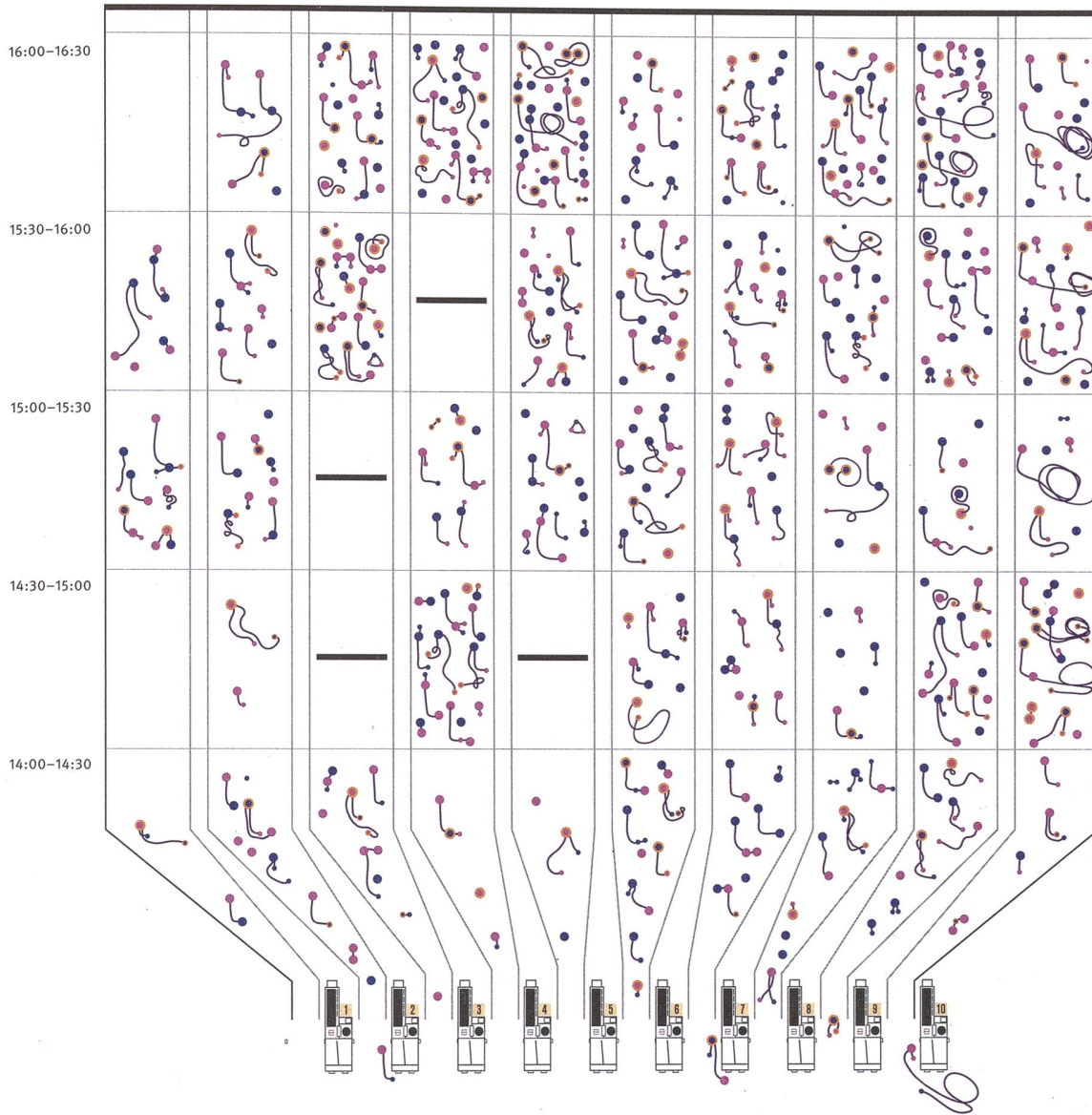
les fonctions et la synergie de l'ensemble des protéines ou des gènes d'une cellule ou d'un organisme.

Mais pourquoi la statistique bayésienne ne s'est-elle pas imposée voilà 250 ans? Parce que ses modèles sont souvent mathématiquement insolubles en raison de la complexité des intégrales. Ce n'est qu'avec le développement des technologies de simulation, dans les années 1970, couplé à l'avènement d'ordinateurs performants, que les méthodes sont devenues applicables. Certaines d'entre elles font encore débat, comme l'intégration correcte des connaissances préalables dans les calculs. Souvent, il s'agit de méthodes peu standardisées. Il n'existe donc pas de logiciel statistique rendant possible la conduite d'un test bayésien en quelques clics. La question est de savoir si un tel logiciel verra jamais le jour, car pour de nombreuses analyses, des spécialistes sont indispensables. Les chercheurs emmenés par Penelope Vounatsou programment donc eux-mêmes les fonctions informatiques dont ils ont besoin, et ce dans différents langages informatiques.

Andreas Papassotiropoulos a lui aussi engagé un mathématicien pour la conduite d'analyses bayésiennes au Département des neurosciences moléculaires. Il conclut néanmoins en ces termes son plaidoyer en faveur d'une formation continue à vie en statistique, pour une frange aussi large que possible de la population: «Même si, à l'avenir, des spécialistes se chargeront des analyses, cela ne dispense pas les chercheurs, les médecins et autres mandataires de comprendre les méthodes. Car au bout du compte, ceux qui interprètent les résultats et doivent prendre des décisions sur cette base, c'est nous.»

Files d'attente à la caisse du supermarché, le samedi entre 14 et 16 h 30 (2009). Timo de Wit, Haute école des arts de Berne.



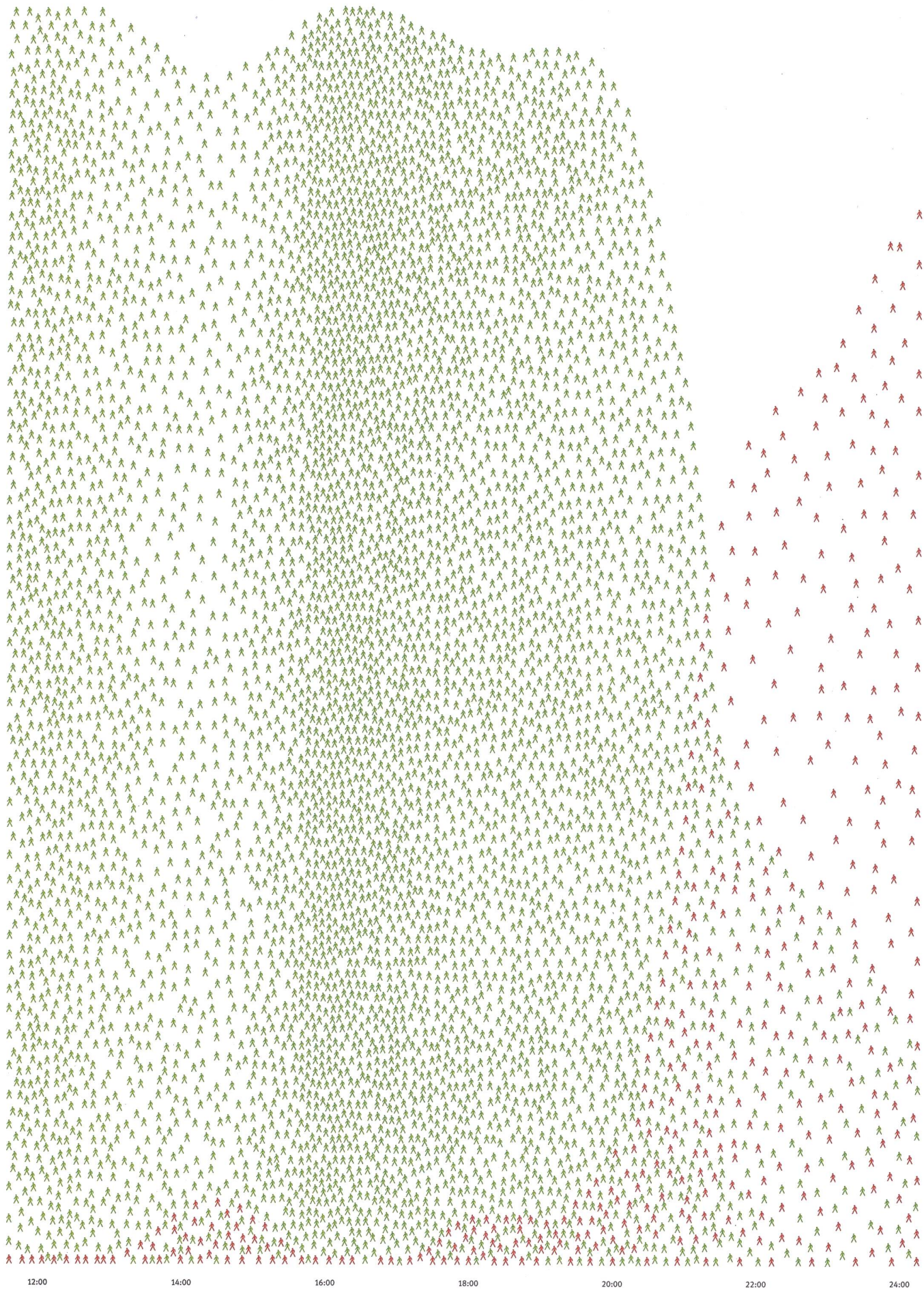


Légende

- Masculin
- Féminin
- Nervosité et excitation
- ⤵ Lien & interaction
- Age 0-18
- Age 18+

Exemples choisis

- Une femme fâchée admoneste son mari
- Père et fils payent, mère et fille emballent les achats
- Deux amies avec des poussettes
- L'enfant fait des glissades, la mère est à la caisse
- Couple fonctionnant bien
- Le père paye, mère et fille se baladent
- Un jeune joue le médiateur
- Un garçon impressionne deux fillettes (achète une glace)
- Père préoccupé par ses enfants turbulents
- Les parents se bécotent, la fille emballe les achats



Inventée par des savants du début des Temps modernes, la statistique a connu un succès foudroyant et sans précédent. Fort utile pour opérer des radiographies socioculturelles, la pensée statistique occulte cependant souvent une partie de la réalité.

Par Urs Hafner

## Par-delà l'aspect qualitatif

Le quotidien de boulevard alémanique *Blick* a rapporté récemment les résultats d'une nouvelle étude selon laquelle les lauréats de prix Nobel viendraient plus souvent de pays où l'on consomme beaucoup de produits laitiers. La Suède arrive en tête de ce classement, alors que la Chine ferme la marche. L'hypothèse non dénuée d'humour du chercheur: le lait contient beaucoup de vitamine D, et il est souvent consommé avec du chocolat dont la forte teneur en flavonoïdes améliore aussi les performances intellectuelles.

Alors que cette étude, publiée dans une revue scientifique renommée, aurait pu être interprétée comme une critique du recours candide aux données statistiques, pour certains médias, elle indiquait que la relation de cause à effet suggérée existait vraiment. Et donc, qu'en moyenne, les Chinois étaient moins intelligents que les Européens de l'Ouest. La plausibilité de ces résultats tenait à l'autorité des chiffres. L'étude venait étayer une image de la réalité, tout en témoignant de la puissance de cette image: pour être réel, un phénomène doit se fonder sur des chiffres, correspondre à une moyenne et être considéré comme représentatif.

### Les Roumains en queue de liste

Le fait qu'une analyse statistique ne reproduise pas certains clichés mais critique une réalité sociale ne garantit pas qu'elle soit porteuse d'une réflexion. La *Frankfurter Allgemeine Zeitung* a récemment rapporté les résultats d'une étude de l'Unicef selon laquelle les petits Allemands seraient «de plus en plus malheureux» malgré leur aisance matérielle. En termes de niveau de vie moyen, de possibilités en matière d'éducation, de santé et d'environnement, ils arrivent au septième rang européen. Les petits Néerlandais sont en tête, et les jeunes Roumains ferment la marche, au 29e rang. Mais lorsqu'on leur demande d'évaluer leur situation, les petits Allemands se placent en 22e position alors que les petits Roumains restent à la dernière. L'étude montre deux choses: malgré un haut niveau de vie, les enfants peuvent se sentir mal, et dans

les pays d'Europe de l'Est, les conditions d'existence des jeunes sont bien plus difficiles que dans les pays d'Europe du Nord.

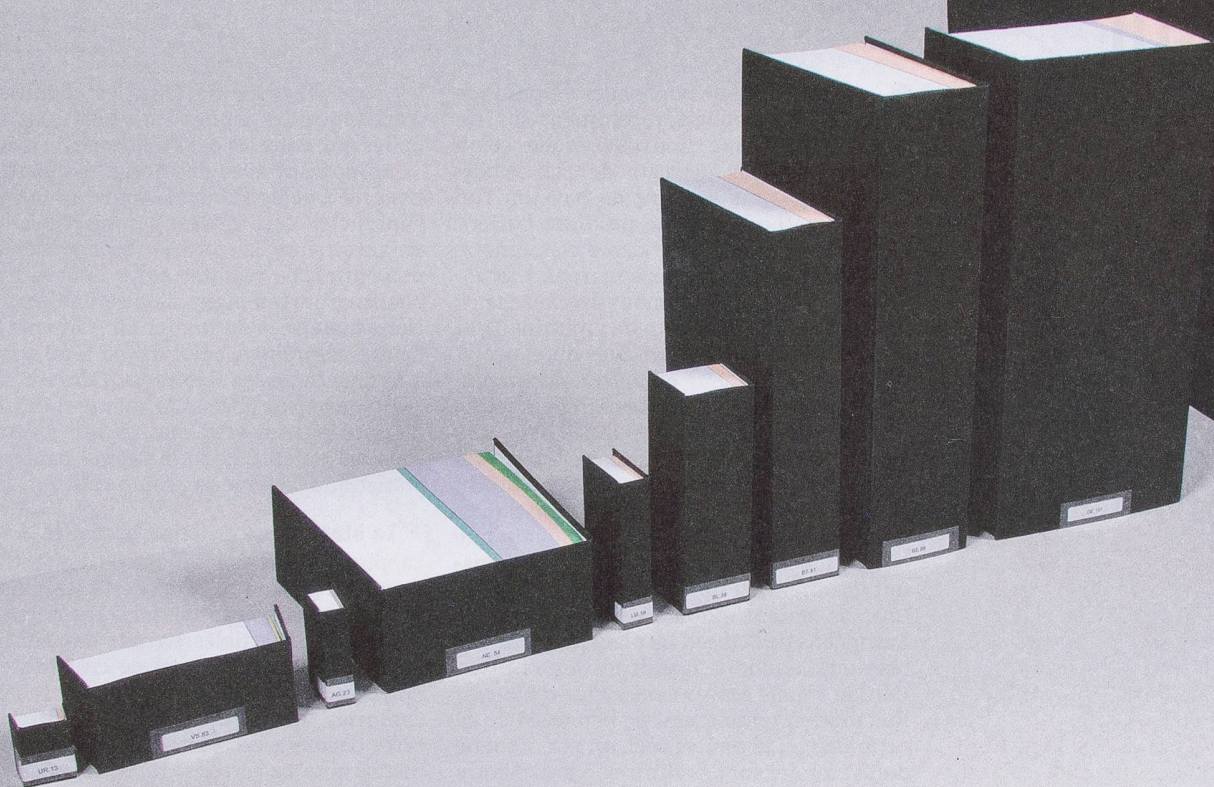
En opérant avec des données quantitatives, l'étude de l'Unicef brosse un tableau superficiel des analyses subjectives des enfants, sous la forme standardisée de moyennes. Le résultat est à double tranchant: ce qui témoigne, d'un côté, des conditions malaisées en Roumanie, cimente, de l'autre, le préjugé, péjoratif au fond, quant à la misère de ses habitants. L'enfant roumain moyen se porte mal, même si ce sujet n'existe pas en tant que tel. En mettant tous les enfants dans le même panier, la statistique façonne sa propre réalité.

### La statistique anticipe la réalité

La plupart des médias entretiennent un rapport étroit au registre de la statistique ou de la statistique triviale. Les communiqués et les articles livrant des chiffres jouissent d'une plus grande crédibilité. A l'inverse, les consommateurs se sentent confortés dans leur vision des choses si cette dernière est partagée par l'instance médiatique. Le cercle se referme avec l'omniprésence des sondages d'opinion: les médias ont tendance à produire des contenus correspondants aux désirs des consommateurs, et ceux-ci désirent ce que les médias produisent. Les sondages d'opinion avant les votations sont particulièrement importants pour les médias. Ils accompagnent la foire d'empoigne politique et font peu à peu monter la tension. Mais en même temps, ces prévisions, recueillies à l'aide de méthodes statistiques, sont assimilées au résultat auquel il faut s'attendre. La statistique anticipe la réalité.

Les statistiques jouent aussi un rôle de premier plan en politique et dans l'administration. En matière de politique migratoire ou de politique de la santé, par exemple, presque aucune décision n'est prise sans référence à des statistiques. Les politiciens se légitiment en se positionnant en fonction d'indicateurs socioculturels, telle l'augmentation ou la diminution des demandes d'asile, ou les corrélations entre style de vie et risque de mortalité. En règle

Comportement des passants à un passage pour piétons entre midi et minuit. L'illustration montre combien de gens, suivant l'heure, traversent la rue au vert ou au rouge. Sabine Affolter, Haute école des arts de Berne.



Le nombre de prêts et de visiteurs ainsi que l'offre médiatique de neuf bibliothèques publiques des cantons d'Uri, du Valais, d'Argovie, de Neuchâtel, Lucerne, Bâle-Campagne, Bâle-Ville, Berne et Genève (de gauche à droite), 2009. Katja Rüfenacht, Haute école des arts de Berne.

générale, la dispute politique s'enflamme lorsqu'il s'agit d'interpréter ces chiffres et les mesures qu'ils appellent, mais elle ne porte ni sur les données ni sur la valeur qu'on leur attribue: un accès privilégié à la réalité.

La pensée statistique, ou plutôt son pendant trivial, s'est hissée au rang de forme de connaissance dominante. Elle prétend être objective, ou au moins plus objective que d'autres formes d'interprétation. Ce statut unique, elle l'a obtenu grâce à une étrange interaction entre deux formes d'autorité, par ailleurs tout à fait différentes, pour reprendre les termes d'Alain Desrosières dans sa *Politique des grands nombres*: celle de la science et celle de l'Etat. A partir du XVII<sup>e</sup> siècle, la philosophie s'est émancipée de l'autorité de la religion et des princes, développant des modes de pensée qui permettaient d'étayer les décisions concernant l'avenir et de déterminer le degré de fiabilité des connaissances scientifiques (à l'aide du calcul de l'erreur).

### Connaissance de l'Etat

En même temps, avec l'absolutisme, les princes gagnaient en influence sur leurs territoires et l'administration de ces derniers. D'après Alain Desrosières, ils avaient besoin d'une statistique, construite comme un «espace d'équivalence cognitif» par rapport à la réalité, à l'aide de laquelle ils pouvaient décrire, administrer et former les sociétés. L'Allemagne, avec ses innombrables principautés, a joué un rôle pionnier dans l'élaboration de cette statistique. Le terme même - connaissance de l'Etat, en allemand - est aussi une création lexicale de Gottfried Achenwall, juriste et historien allemand (ou de l'un de ses collègues). L'Angleterre, en revanche, a connu le développement de l'arithmétique politique et des techniques de calcul, alors que la France a produit de savantes descriptions et enquêtes empiriques.

Ces branches statistiques ont d'abord coexisté, et elles étaient très contestées. La statistique allemande, notamment, a fait l'objet, au début du XIX<sup>e</sup> siècle, d'une critique sévère dans une revue scientifique majeure de l'époque, les *Göttingischen*

*Gelehrten Anzeigen*: «Ces pauvres bouffons diffusent l'idée folle selon laquelle il serait possible de prendre la mesure de la puissance d'un Etat en connaissant sa superficie, sa population, son revenu national et le nombre d'animaux qui broutent dans ses enclos.» Un demi-siècle plus tard, plus personne ne songeait à présenter la statistique (en tant que moyen de connaissance) comme une affaire de fous, du moins personne parmi les employés des nombreux services statistiques étatiques qui venaient d'être créés. En comparaison internationale, le Bureau fédéral de la statistique - l'actuel Office fédéral de la statistique - a vu le jour tard, en 1860 seulement. La Suisse républicaine et fédérative, à ses débuts, faisait preuve de réticence envers cette description chiffrée de la société.

### Un rôle dans la lutte sociopolitique

La percée de la statistique moderne, telle qu'elle est appliquée aujourd'hui en sciences et en politique, s'est produite aux Etats-Unis, au milieu du XX<sup>e</sup> siècle. Le champ statistique a été internationalement unifié dans le paradigme mathématique: sondages d'opinion sur la base d'échantillons représentatifs, comptabilités nationales et recours à l'informatique. La statistique a joué un rôle important dans la lutte sociopolitique autour du développement du capitalisme et de ses institutions sociales, car elle permettait de recenser les nécessaires et les biens nécessaires à leur existence, mais aussi de se disputer à leur propos. A noter que le recul de l'Etat social, en cours depuis quelques années et politiquement tout aussi contesté, se réfère lui aussi à des connaissances issues de la statistique.

Ces connaissances ne sont pas seulement produites par une statistique basée sur des mathématiques sophistiquées dont les opérations ne sont comprises que par certains spécialistes, comme c'est largement le cas au sein des Ecoles polytechniques fédérales. Sous l'égide des «sciences dures», et plus récemment du «big data», les méthodes quantitatives se sont établies dans la plupart des disciplines. Réservées traditionnellement aux sciences natu-

relles, elles ont aussi, depuis plusieurs décennies, fait leur entrée dans les sciences sociales.

Les sciences politiques, la psychologie, l'économie et une partie de la sociologie copient désormais le modèle des sciences naturelles. Leur fierté d'avoir transformé ces disciplines, autrefois «molles» et «imprécises», en sciences «véritables» est manifeste. Mais cela signifie aussi que seuls certains pans de la réalité sociale sont répertoriés: les portions quantifiables et reproductibles par l'expérimentation. Tout le reste - ce que Max Weber appelait «l'aspect qualitatif» - reste invisible: le mental, l'interprétatif, les sentiments mais aussi les actes, autrement dit la réalité sociale, tantôt figée et tantôt fluide.

La statistique est la discipline-phare de la planification et du contrôle socio-technologique de la société d'aujourd'hui. A travers elle, les instances médiatiques, politiques et scientifiques produisent un certain type de réalité: le monde tel un agrégat monochrome de chiffres et de proportions. Et sous la houlette de la statistique, la société cherche à s'assimiler à cet agrégat.

# Les données ont leurs détectives

Adapter des modèles à la réalité, même s'ils restent des approximations: c'est à cela que travaillent les statisticiens.

Par Simon Koechlin

«**L**a statistique est un travail de détective, explique Anthony Davison. On sonde les données à la recherche de points de repère, dans le but de comprendre ce qu'on voit.» Professeur de statistique à l'EPFL, ce chercheur étudie notamment les événements extrêmes: vagues de chaleur et pluies diluviennes, records sportifs ou krachs boursiers. Il développe des méthodes qui permettent de comprendre la fréquence ou le risque d'événements rares, à l'aide de modèles statistiques.

Pour ce faire, il n'existe pas de recette toute faite. «Les possibilités de modéliser ces événements sont nombreuses, explique Anthony Davison. Mais lorsqu'on commence le travail, on se concentre sur les méthodes, en fonction de leurs propriétés mathématiques.» Une fois que l'on a trouvé une structure statistique adaptée, on procède à un test de crédibilité, en vérifiant si les formules sont compatibles avec les données relatives aux vagues de chaleur ou aux précipitations. Le plus souvent, il faut ensuite tout reprendre au début, pour adapter, améliorer et affiner le modèle. «Il est décisif que celui-ci restitue les données de manière conforme», poursuit le statisticien.

## Identifier les paramètres pertinents

Les données sont tout sauf une denrée rare. «Par le passé, on mesurait la pression artérielle ou le pouls, note Sara van de Geer, professeure de statistique à l'EPFL. Alors qu'aujourd'hui, dans une étude médicale, il est courant que l'on détermine 20 000 gènes chez chaque participant, pour ensuite savoir lesquels d'entre eux influencent une maladie donnée.» L'art des statisticiens réside dans leur capacité à identifier les paramètres pertinents pour un questionnement donné. Sara van de Geer consacre sa recherche à une méthode appelée LASSO, qui a gagné en popularité ces dernières années. «Le LASSO permet des prévisions presque optimales», explique-t-elle. Mais seulement presque. Si l'on reproduit l'expérience qui fournit les données pour un modèle LASSO, la prévision se modifie aussi quelque peu. La tâche du statisticien consiste à procéder à une estimation de ce genre de variabilités. Mais dans le cas du LASSO, celle-ci est mathématiquement impossible: «Pour mon travail de détective,

«L'incertitude reste inhérente à la statistique.»

Sara van de Geer

je dois donc me fixer d'autres objectifs, explique la chercheuse. J'examine en détail seulement quelques paramètres au lieu des 20 000, par exemple.»

La manière de procéder est toujours la même: développement de méthodes, programmation et production de preuves mathématiques. Mais les statisticiens doivent aussi collaborer avec des personnes qui, le reste du temps, n'ont pratiquement jamais affaire aux mathématiques, rappelle Anthony Davison. Et les applications venues d'autres domaines influencent aussi leur travail, ajoute-t-il. De nombreux principes fondamentaux en statistique ont leur origine dans une application. Comme le principe de la randomisation. Ce procédé a été inventé par le chercheur britannique Ronald A. Fisher pour conduire des expériences en agriculture. Aujourd'hui, il est appliqué dans la plupart des branches scientifiques, notamment en médecine, où les sujets participant aux expériences sont attribués au hasard à différents groupes de traitement.

Toutefois, l'incertitude reste inhérente à la statistique, rappelle Sara van de Geer. Même le meilleur des modèles demeure une approximation de la réalité. Et Anthony Davison de préciser: «Dans le cas des prévisions climatiques, on part souvent du principe que l'avenir se comportera comme le passé. Mais personne ne sait si c'est vrai. Dans l'hypothèse inverse, il faudrait reconsidérer l'utilisation faite jusqu'ici des données collectées pour les prévisions.»

Mais c'est précisément cette interaction entre l'abstraction et la réalité que le scientifique trouve si stimulante. Il travaille actuellement à un modèle qui décrit la progression et le déroulement de violentes tempêtes de pluie: «Quand on réalise que le monde réel se comporte effectivement comme le modèle, c'est fascinant», note-t-il.