

Plus étonnant peut-être, un an après, à un mois d'intervalle, le 21 septembre et le 16 octobre 2010, le Loto israélien sortit la série 13, 14, 26, 32, 33, 36, provoquant là encore l'étonnement mondial. Comme l'expliquait David Hand dans son article «Des coïncidences pas si étranges» (*Pour la Science* n°438, avril 2014), si l'on prend en compte le nombre de jeux de Loto dans le monde, et le nombre de tirages que chacun d'eux opère, souvent plusieurs fois par semaine, notre étonnement doit cesser.

Plus précisément, considérons le Loto bulgare où l'on tire 6 numéros entre 1 et 49, ce qui donne une probabilité de gagner de 1/13983816. Demandons-nous combien de tirages sont nécessaires pour que l'événement «deux d'entre eux donnent le même résultat» ait une probabilité supérieure à 50% de se produire. La réponse est 4404.

Le calcul est analogue à celui fait pour le célèbre paradoxe des anniversaires, selon lequel dès que 23 personnes sont réunies, la probabilité que deux d'entre elles aient la même date anniversaire dépasse 50 %. Cela montre que les événements ressentis comme extraordinaires pour les loteries bulgare et israélienne sont en fait nécessaires. Ce n'est pas la survenue des tirages identiques qui est improbable, mais l'inverse: si tous les tirages étaient toujours différents, nous devrions nous en étonner et en rechercher l'explication.

Une autre source d'étonnements provient des séries rapprochées d'événements rares, tels que les accidents d'avion. Les journalistes aiment mentionner une prétendue «loi des séries», pourtant inconnue des mathématiciens, qui expliquerait ces rapprochements jugés à la fois fortement improbables et explicables par cette introuvable loi... Jacques Chirac disait simplement: «Les emmerdes, ça vole »

3

LOI DE POISSON ET ACCIDENTS D'AVION

En août 2005, une série de cinq catastrophes aériennes s'est produite dans un intervalle de 22 jours, faisant plusieurs centaines de morts au total. Cela nous semble extraordinaire, mais l'est-ce vraiment ?

Élise Janvresse et Thierry de la Rue proposent une belle analyse du problème dans leur petit livre *La Loi des séries, hasard ou fatalité ?* (Le Pommier, 2007). En examinant les données disponibles, ils ont évalué que la probabilité totale d'un accident sur un vol d'au moins 30 passagers est de 0,04 par jour. Ils ont ensuite calculé le nombre total d'accidents auquel il faut s'attendre sur une période donnée. On fait appel à la loi de Poisson, une loi des événements rares introduite par Siméon Denis Poisson en 1837. Elle est appropriée ici, puisque la probabilité que 1 des 20 000 vols quotidiens soit victime d'un accident est $0,04/20\,000 = 0,000\,002$.

La loi de Poisson fait intervenir un paramètre d'intensité λ calculé à partir du nombre moyen d'accidents sur

la période considérée — soit, pour une journée, 0,04. D'après la loi de Poisson, la probabilité que le nombre N d'accidents soit égal à l'entier k durant la période choisie est donnée par la formule :

$$\text{Proba}(N = k) = \exp(-\lambda) \lambda^k / k !$$

Si l'on se fixe une fenêtre particulière de 22 jours, alors l'intensité pour cette période vaut $\lambda = 0,04 \times 22 = 0,88$. La probabilité pour qu'il n'y ait eu que 0, 1, 2, 3 ou 4 accidents est donc :

$$\begin{aligned} \text{Proba}(N \leq 4) &= \text{Proba}(N = 0) + \text{Proba}(N = 1) + \text{Proba}(N = 2) + \\ &\text{Proba}(N = 3) + \text{Proba}(N = 4) = \\ &\exp(-0,88)[1 + 0,88^1 + \\ &0,88^2/(2 \times 1) + 0,88^3/(3 \times 2 \times 1) + \\ &0,88^4/(4 \times 3 \times 2 \times 1)] = 0,9978. \end{aligned}$$

La probabilité qu'il y ait eu 5 accidents ou plus dans la période de 22 jours fixée est donc $1 - 0,9978 = 0,0022$, ce qui est faible.

La série malheureuse de catastrophes de 2005 serait donc extraordinaire ! Ce serait oublier qu'il faut envisager non pas une fenêtre de 22 jours, mais toutes les fenêtres de 22 jours présentes durant une année. Le calcul est assez compliqué, car les fenêtres à envisager se chevauchent. Élise Janvresse et Thierry de la Rue ont donc fait usage de ce qu'on nomme les statistiques de balayages, mises au point assez récemment pour traiter ce type de problèmes. Ils concluent que la probabilité que, durant une année donnée, il se produise 5 accidents aériens graves ou plus dans une fenêtre de 22 jours est 11 %. C'est assez faible, mais cela rend la série malheureuse de 2005 peu étonnante, d'autant plus que le calcul mené ne prend pas en compte les variations saisonnières du trafic aérien qui, en concentrant les vols sur certaines périodes, augmentent le 11 % obtenu.

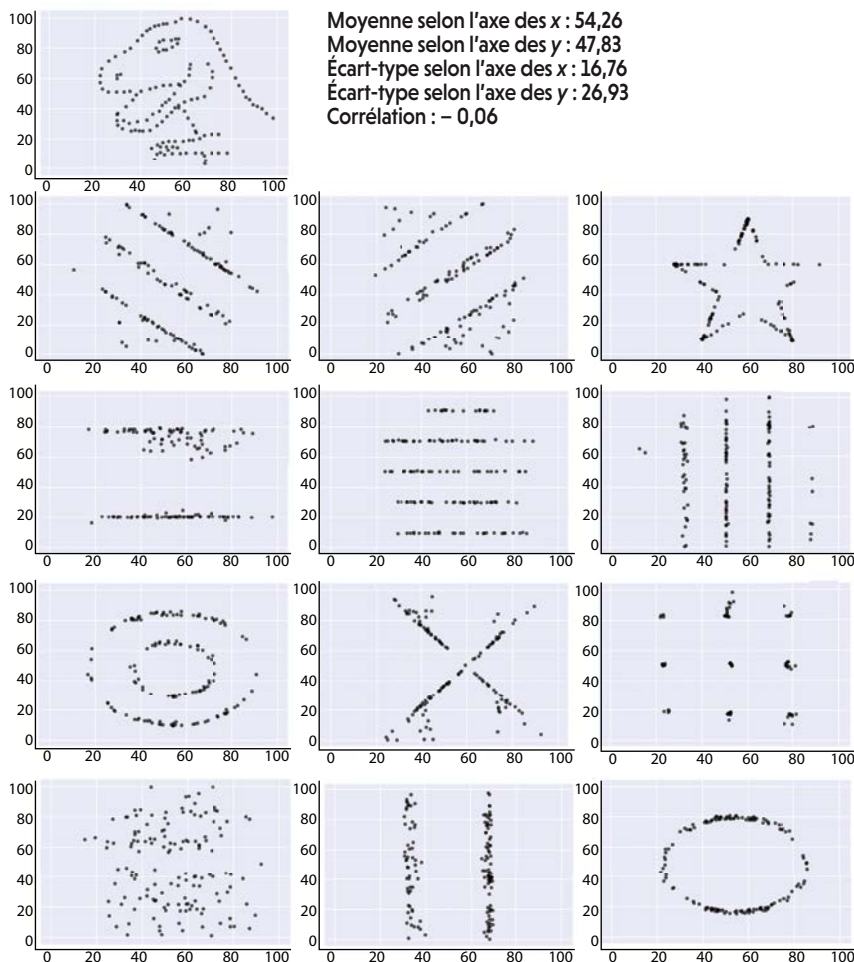


MÊMES PARAMÈTRES... ET POURTANT SI DIFFÉRENTS

4

Nous pensons en général que si des données ont des paramètres statistiques proches (moyennes, écarts-types, etc.), elles se ressemblent nécessairement. Les paramètres statistiques, pensons-nous, sont comme des versions compressées des séries de données. Justin Matejka et George Fitzmaurice prouvent graphiquement que c'est faux. À l'aide d'une méthode qui leur permet de créer facilement des séries ayant exactement les mêmes paramètres statistiques, ils ont notamment produit ces 13 schémas, utiles à tous ceux qui souhaitent prendre conscience de notre fragile compréhension intuitive du hasard et des statistiques.

Il s'agit de graphiques de séries de couples de nombres (x, y) , qui apparaissent très différents. Pourtant, d'une série à l'autre, les 5 paramètres suivants sont les mêmes : moyenne des x , écart-type des x , moyenne des y , écart-type des y , corrélation entre les x et les y . Lorsque la corrélation entre les x et les y est parfaite, ce type de schéma est une droite. (Voir J. Matejka et G. Fitzmaurice, « Same stats, different graphs : Generating datasets with varied appearance and identical statistics », *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1290-1294, ACM, 2017.)



> toujours en escadrille!... Certaines analyses tentent d'en identifier l'origine en parlant d'une « attente excessive d'étalement » (voir mon article « Notre vision du hasard est bien hasardeuse », *Pour la Science* n° 293, mars 2002) : notre intuition nous souffle à tort que, par exemple, les dates des accidents d'avion doivent être régulièrement espacées, alors que les statistiques nous montrent autre chose.

Cette attente excessive d'étalement a été clairement analysée dans les cas précis des accidents d'avions par Élise Janvresse et Thierry de la Rue, de l'université de Rouen (voir l'encadré 3 page 111).

DES SÉRIES TRÈS DIFFÉRENTES, MAIS AYANT LA MÊME STATISTIQUE

Notre bon sens est défaillant pour traiter les probabilités et nous sommes surpris dans des cas où il n'y a pas lieu de l'être. Utiliser le mot « coïncidence » n'explique rien ou alors cela introduit des idées étranges comme la synchronicité de Carl Gustav Jung ou les champs morphiques de Rupert Sheldrake, dont les scientifiques ont vainement cherché à prouver l'existence (voir par exemple www.sceptiques.qc.ca/dictionnaire/).

Cette faiblesse de notre esprit à percevoir correctement les probabilités a encore récemment été mise en évidence. On est tenté de penser que si des données ont des paramètres statistiques proches, elles doivent nécessairement se ressembler ; or Justin Matejka et George Fitzmaurice, de la société Autodesk Research à Toronto, ont créé délibérément des exemples magnifiques montrant que c'est faux.

Pour engendrer leurs séries, ils ont mis au point une technique qui leur permet de jouer facilement avec les formes des graphes créés. Ainsi, la figure de l'encadré 4 ci-contre visualise une collection de 13 séries de couples de nombres (x, y) qui, malgré leurs graphes très différents (chaque couple (x, y) représentant un point du plan), ont les mêmes 5 paramètres suivants : moyenne des x , écart-type des x , moyenne des y , écart-type des y , corrélation entre les x et les y .

SIMPLE ET INATTENDU

Toujours afin de comprendre et de classifier les situations où notre esprit s'étonne alors qu'il ne devrait pas, détaillons à présent un cas moins connu, car lié à une théorie assez récente. La mauvaise compréhension de ce qui est probable, car simple, conduit à percevoir certains événements comme étonnants alors qu'ils ne le sont pas, et donc à croire être en présence de coïncidences miraculeuses sans bonne explication, alors que la situation est banale.

La notion la plus générale de simplicité est celle qui provient de la théorie algorithmique de l'information. Comme nous n'en avons pas