



LES STATISTIQUES ... C'EST DE LA MAGIE !



COÏNCIDENCES

Il est facile d'être étonné par des faits qui font penser qu'il vient de se passer quelque chose d'extraordinaire, par des coïncidences que l'on pense totalement improbables.

En cette matière, il faut effectivement se forger une intuition avant de conclure au surnaturel ou à l'extraordinaire.

Voici deux exemples.

Quelqu'un annonce froidement à son assistance qu'il tient pour certain que deux d'entre eux sont nés le même jour, ou que leurs parents se sont mariés le même jour etc.

Cela n'a rien d'extraordinaire : la probabilité pour que ce soit vrai dépend du nombre de personnes, mais à partir de 23, elle est déjà de 50% et si on s'adresse à 50 personnes, il n'y a plus que 3% de chances de se tromper !

En effet la probabilité de se tromper est égale à

$$365 \times 364 \times 363 \times \dots / 365^N$$

où N est le nombre de personnes et où on a mis N termes au numérateur. Ainsi on vérifie que

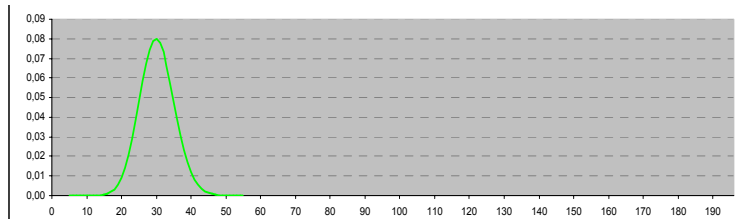
$$365 \times 364 \times 363 \times \dots \times 316 / 365^{23} \approx 0.0296.$$

Dans le même ordre d'idées, un medium annonce pour les trois années à venir 169 dates pour lesquels il y aura des séismes de magnitude supérieure à 6,5. On constate après coup que, sur les 196 séismes qui se sont effectivement produits, 33 avaient été prédits par le medium.

Or la probabilité d'avoir 33 succès de la sorte n'est que de 7,1%. Pourtant il n'y a là rien d'extraordinaire. Pourquoi ?

Parce que la probabilité qu'il y ait entre 20 et 40 succès est environ 98% ! Or le medium n'avait pas annoncé exactement 33 succès ...

En fait c'est s'il avait obtenu plus de 40 ou moins de 20 que cela aurait été surprenant. À bien y réfléchir, si le medium n'avait eu **aucun** succès, alors là, oui, il aurait fallu s'émerveiller !



LE SOPHISME DU PROCUREUR

C'est un paradoxe très largement débattu et dont il faut se méfier dès que l'on manipule des probabilités.

L'argument fallacieux consiste à confondre « la probabilité qu'un événement survienne » avec « la probabilité qu'un événement soit survenu dans un cas étudié ».

En particulier être innocent d'un crime, c'est se trouver dans la situation où
1° une mort est survenue

2° un événement extérieur (par exemple une mort naturelle inexplicquée) s'est produit.

Autrement dit pour évaluer la probabilité d'innocence d'un accusé, il faut chercher le nombre de fois qu'un événement rarissime se produit dans une population très restreinte (celle de ceux qui ont subi un décès) et non le nombre de fois qu'il se produit au sein de la population totale.

Ainsi quand on analyse des événements rares l'important est de comparer le rare au très rare. Et pareillement avec les événements probables : il faut les comparer aux événements très probables !

Prenons l'exemple des tests ADN.

La probabilité pour que deux profils ADN soient identiques est environ de 1/10 000. Si maintenant, suite à une affaire de meurtres, on trouve un suspect dont le profil ADN est exactement celui trouvé sur la victime, doit-on en déduire qu'on n'a qu'une chance sur 10 000 de se tromper en l'accusant ?

NON.

En effet il faut le contexte pour pouvoir dire quelque chose. En fait cette donnée permet de restreindre les suspects, mais pas de pouvoir accuser avec certitude. Pour cela il faut des renseignements complémentaires !

Faisons le calcul ... Supposons que pour parvenir à trouver le suspect, on ait dû comparer le profil ADN trouvé sur la victime avec 20 000 profils (issus d'un fichier de la police scientifique). Avec la probabilité donnée, il y a en fait 86% de chances de trouver au moins un profil identique et il y a 27% de chances de trouver exactement un profil identique.

En effet ces deux probabilités sont respectivement égales à

$$1 - (1 - 1/10\,000)^{20\,000} \approx 0.86$$

et

$$20\,000 \times (1 - 1/10\,000)^{19\,999} \times 1/10\,000 \approx 0.27.$$

Il ne faut pas en déduire que les tests ADN ne servent à rien, ni qu'on ne peut rien dire à partir d'eux. Si par exemple la victime reconnaît son agresseur et qu'*ensuite* on effectue un test ADN qui est positif, alors les calculs précédents ne s'appliquent pas : il faudrait pouvoir remplacer 1/10 000 par la probabilité pour que **simultanément** une victime ait cru reconnaître son agresseur et que le profil des deux, agresseur et faux agresseur, coïncident. Une telle statistique est difficile à établir car le nombre de cas d'étude est très faible.