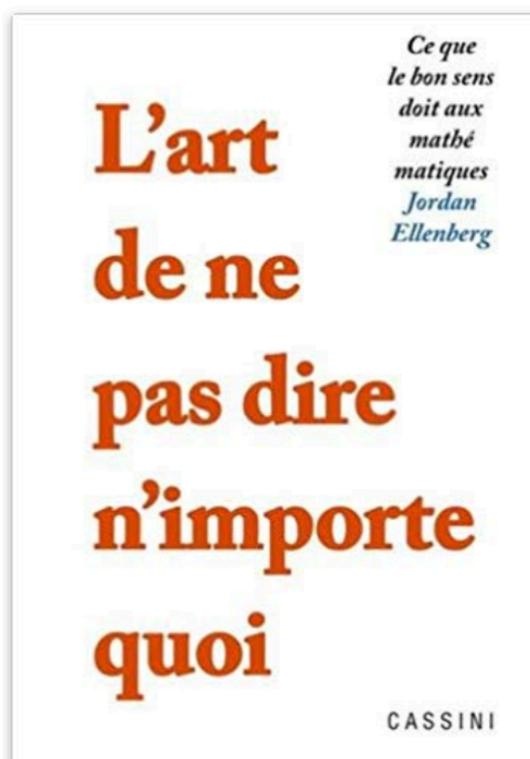


Note de lecture rédigée par Jean-Louis PIEDNOIR¹

L'ART DE NE PAS DIRE N'IMPORTE QUOI

Jordan ELLENBERG

Livre (496 pages)
Édition : CASSINI – 2017



Jordan Ellenberg n'est pas statisticien, mais algébriste. Il a effectué des travaux en théorie des nombres, enseigne la géométrie algébrique, mais a écrit de nombreux articles de vulgarisation. L'objectif poursuivi dans l'ouvrage analysé est de montrer comment des erreurs très communes peuvent être analysées et démontées dès lors que l'on fait appel à une formalisation mathématique. Les mathématiques sont donc « la continuation du bon sens par d'autres moyens ». L'ouvrage est drôle, facile à lire, sans calcul complexe et bourré d'exemples, provenant essentiellement des Etats-Unis.

L'auteur doit être doté d'une solide culture dans de nombreux domaines des mathématiques. Les exemples traités, issus de réalités concrètes, permettent de faire appel à différents champs : théorie des fonctions, de l'information, des jeux, logique. Mais près de la moitié de ces exemples issus de situations réelles font référence à des traitements de type statistique, descriptive et inférentielle. Certes les pièges évoqués sont connus des statisticiens mais, pour l'enseignement de la discipline, il est bon de diversifier les exemples, de choisir un

¹ Inspecteur général honoraire de l'Education Nationale, amjl.piednoir@orange.fr

Note de lecture : « L'art de ne pas dire n'importe quoi » (J. Ellenberg, 2017)

point de vue différent. Aussi on ne peut que recommander cet ouvrage aux enseignants de statistique quel que soit le niveau de leur intervention.

La description statistique et son interprétation

Abraham Wald a fait s'interroger les militaires sur les données fournies par les avions revenant de mission, oubliant ceux qui avaient été abattus. Des politiques manipulent (inconsciemment ?) des proportions de créations d'emplois, oubliant ceux qui avaient été détruits. On polémique sur les classes sociales bénéficiaires de la croissance économique, mais on oublie celles qui ont vu leur niveau de vie baisser. Il est dangereux d'extrapoler des liaisons mises en évidence par une régression linéaire comme quand on regarde les frais de scolarité universitaires et les performances des étudiants.

Les proportions sur des effectifs trop petits ou trop différents n'ont pas de sens. On a établi un classement sans signification des états selon le taux de cancers du cerveau observé oubliant les différences de populations, un palmarès des performances des écoles élémentaires sans tenir compte du nombre d'élèves. A ce propos on introduit les notions de hasard, de loi des grands nombres, de théorème central limite sur lesquels on revient dans d'autres chapitres.

L'inférence statistique

Quand il y a un grand nombre de situations analogues, la loi des grands nombres nous fait comprendre que « les choix improbables arrivent tout le temps », comme les performances du courtier en bourse. La théorie des tests de Fischer, ses démêlés avec Neyman et Pearson font l'objet de développements illustrés par des exemples comme l'existence ou non de la main chaude des joueurs de basket. Un test statistique peut être significatif sans que l'utilisation du terme significatif soit pertinente dans la vie courante comme dans l'utilisation d'une étude sur les pilules contraceptives ; même multiplié par 2 ou 3, un risque très petit reste très petit.

La théorie des tests avec son hypothèse nulle que l'on souhaite éliminer est analogue au raisonnement par l'absurde, il s'agit d'une réduction à l'improbable mais les deux concepts sont différents. Quand on associe en génétique gènes et maladie, beaucoup de gènes ne sont pas pris en compte. Attention au biais de publication qui a rendu obsolète de nombreuses publications médicales sans compter le charcutage inévitable de p (il s'agit de la p -valeur du test). En conclusion le test statistique est vu comme un détective mais pas comme un juge.

Un chapitre est consacré à la manipulation des probabilités conditionnelles avec un passage sur la limite de l'utilisation des données massives avec en illustration la détection des terroristes. Cela conduit à des interrogations sur l'inférence bayésienne et l'interprétation subjective de la probabilité, peut-être le chapitre le moins convainquant.

L'introduction de la notion d'espérance mathématique est l'occasion de raconter une histoire extraordinaire : une bande d'étudiants ont exploité une faille de la réglementation d'une loterie et gagné beaucoup de dollars ! Plus sérieusement, on expose une démonstration élémentaire du problème de l'aiguille de Buffon. L'évocation du paradoxe de Saint-Petersburg est l'occasion d'introduire la notion d'utilité, d'organiser un débat sur le pari de Pascal, de présenter sa critique par Ellesberg qui distingue hasard et incertitude.

J.-L. Piednoir

La régression, la corrélation, la question de la causalité sont abondamment illustrées au long de trois chapitres. Des erreurs célèbres sont évoquées comme le triomphe de la médiocrité déduit de la régression vers la moyenne avec Secrist, économiste, Hotelling, qui en fait la critique et l'inévitable Galton, le criminologue Bertillon, Fisher, et l'influence du tabac sur la santé. Comme il est fait allusion aux ellipses de Galton, cela permet d'évoquer la force des représentations graphiques ou géométriques avec l'exemple de la retraite de Russie, les décès de l'armée anglaise pendant la campagne de Crimée par Florence Nightingale. L'interprétation en géométrie euclidienne des indicateurs statistiques permet de montrer pourquoi la corrélation n'est pas transitive. En étude de Santé, méfiez-vous des études faites uniquement en milieu hospitalier : vous oubliez les personnes qui ne sont pas hospitalisées !

D'autres sujets

Si les sujets relevant de la statistique font la majorité des sujets traités, il est d'autres domaines où les mathématiques évitent de dire n'importe quoi. Ainsi sont détaillés, après introduction de la notion de variance, des arbitrages à faire entre temps de traitement et détection dans le cas des codes correcteurs d'erreurs. Un autre sujet reçoit aussi un traitement important : celui de l'agrégation des préférences individuelles en préférence collective avec l'effet Condorcet et le théorème d'Arrow. On illustre avec plusieurs exemples comment la présence d'un troisième candidat, sans aucune chance, influence le résultat du vote, y compris quand il s'agit d'une population de champignons !

Des sujets plus éloignés des applications sont évoqués : les géométries non euclidiennes, la logique avec le théorème de Gödel, les 23 problèmes de Hilbert, etc.

Avis donné

On aura compris que l'auteur de ces lignes a beaucoup apprécié l'ouvrage, il a élargi son stock d'exemples, apprécié le style et l'humour, même s'il a trouvé une présentation un peu désordonnée. Il le recommande à tous les enseignants de statistique, surtout ceux qui enseignent aux non spécialistes.