

Réunion du Comité d'Orientation, Institut Diderot (institutdiderot.fr), le 17 novembre 2021.

Préparation à l'intervention de la rentrée 2022 à l'Institut Diderot :

« Grandeurs et Faiblesses de la Statistique »
Deprins D.

Résumé

La Statistique est un univers. La Statistique classique cherche à réduire l'incertitude du monde en réduisant sa variabilité ; d'un désordre microscopique, la statistique fait émerger un ordre macroscopique, mettant en évidence des invariants, des régularités, des tendances¹. Son principe inférentiel consiste à dégager un savoir prenant son sens à un niveau global, à partir de « particules élémentaires » d'information brute qu'on appelle des données brutes. Cette méthodologie empêche la statistique de prétendre au risque zéro ; même si la statistique appartient au domaine des mathématiques, certes, appliquées, aucun résultat statistique ne peut être assorti d'une certitude à laquelle on pourrait accorder une incontestabilité mathématique. Pour que, dans de nombreux domaines, la statistique puisse être une aide à la décision sans toutefois se substituer à la décision elle-même, il faut consentir à accepter un certain niveau de risque de se tromper. Les modèles probabilistes de la statistique classique sont des fictions ; les variables explicatives ne sont jamais que des causes seulement probables de la variable qu'elles sont censées expliquer. Leur partie aléatoire n'est qu'une érosion du déterminisme et la partie restante, dite déterministe, offre une détermination partielle du phénomène aléatoire étudié. L'interprétation des résultats obtenus introduit une tension signifiante propre aux modèles probabilistes ; elle demande une rigueur extrême si l'on veut en respecter les balises théoriques. La multiplicité des modèles possibles pour ajuster les données disponibles, les nombreuses hypothèses simplificatrices qui sont forcément associées à chaque modèle, l'interprétation des résultats, la démarche inférentielle qui introduit du biais et de l'incertitude, l'incontournable risque de se tromper qui accompagne toujours les décisions statistiques, autant de raisons qui interdisent de prendre les résultats statistiques pour des certitudes ou pour la « vérité vraie ». L'incertitude qui est une dimension essentielle du vivant est l'envers de la vérité.

¹ Currien N., (inédit) *L'impact du Tsunami numérique*, Panel : « La Production des Données de l'INSEE à Google », Séminaire : Des Faits aux Données, (org.) F. Ewald et alii, École Nationale d'Assurance, 22 juin 2011, Paris.

En revanche, la Statistique contemporaine du Big Data peut exploiter le geyser de données foisonnant que nous générons en continu pour notre vie économique, sociale et affective, d'une manière qui s'écarte des méthodes historiques de la statistique ; non plus, du particulier au général, de l'échantillon à la population, mais du particulier au particulier. Un nouveau paradigme prédictif permet des prédictions au niveau de granularité de l'individu, ou plutôt à celui de la singularité de chacun, ce qui le différencie des autres. Il sera question d'interroger la rupture épistémologique dont ces données massives et leur usage sont désormais l'expression et quelles en sont les conséquences dans le monde des data. Par le biais du changement de logiques entre le monde du probable auquel appartient la statistique classique et celui des données massives qui s'en affranchit, on peut se demander quel type de pouvoir peut s'exercer à partir de la statistique qui génère, en flux continu, ce sur-savoir. L'hybridation qui fleurit dans de nombreux domaines (enseignement, guerre, économie, genre, voitures, technologie, etc.) serait-elle la réalité positive de ce changement de logique, mode selon lequel se peuplent la pensée, le mental, les affects et les politiques des hommes du monde numérique ?