

DICTIONNAIRE DE LA STATISTIQUE ET DE LA SCIENCE 2005

AVANT PROPOS

(21 / 03 / 2019)

La **Statistique** est souvent présentée, tantôt sous une **forme « utilitariste »**, à finalités pratiques (« recettes »), tantôt sous une **forme « calculatoire »**, ou encore « procédurale », voire « algorithmique ». Ces deux modes de présentation ont, sans doute, leur intérêt. Mais des études, ou des formations, basées sur ces seules considérations réduisent la **Statistique** à n'être qu'un outils que l'on peut manipuler sans en comprendre les fondements.

Un objectif majeur de ce travail est ainsi un **objectif pédagogique**, dont l'une des conséquences est de montrer que la Statistique, en tant que savoir (« science ») et en tant que savoir-faire (« art »), est, dans son essence, éloignée de ces présentations très restrictives, voire simplistes. Etant à la fois une branche des mathématiques et un instrument général opérant au coeur même de la science, la Statistique apporte donc bien davantage que des recettes ou des calculs, si sophistiqués soient-ils.

En effet, si le savoir est une chose, la compréhension en est une autre. Ce dictionnaire vise à **faciliter l'assimilation** :

- (a) des fondements des **concepts** statistiques ;
- (b) de la **nature des problèmes** à traiter ;
- (c) des **démarches** suivies par le **statisticien** pour mettre au point des **méthodes** appropriées ;
- (d) des **résultats** obtenus à partir de ces fondamentaux (interprétation : signification ou intérêt).

1. Ce dictionnaire a pour but de **rendre service aux utilisateurs de la Statistique**, quels qu'en soient le **niveau** ou l'**étendue** des connaissances, ainsi que les **domaines** d'application.

1.1. *Etudiants et enseignants* des différents degrés universitaires ou des grandes écoles, à titre d'informations de référence. Ce travail peut constituer une « transition » entre l'étude générale de la Statistique, telle qu'elle est enseignée dans les grandes écoles ou les universités, et les travaux statistiques approfondis réalisés pendant la carrière professionnelle ultérieure (revues spécialisées, applications particulières), lesquels nécessitent souvent des connaissances supplémentaires, ainsi que de l'expérience.

1.2. *Scientifiques et praticiens*, appelés ici « **hommes de l'art** », que leurs **domaines de connaissance** respectifs conduisent vers l'utilisation plus ou moins régulière d'« informations statistiques » : données individuelles, séries ou tableaux statistiques, d'une part, articles de revues nécessitant une connaissance des concepts ou des méthodes mises en oeuvre dans les applications, d'autre part.

Ainsi, les *ingénieurs et chercheurs* des entreprises ou des administrations, exercent leur activité dans diverses sphères qui requièrent généralement des bases statistiques, à titre d'outils méthodologique et technique.

2. Pour en asseoir la **valeur scientifique** aussi bien que l'**utilité pratique**, la réalisation de ce dictionnaire a cherché à atteindre des **objectifs importants**.

2.1. D'abord, *combler en France une lacune*, qui a été durable, dans ce domaine.

2.2. *Réunir, dans un même endroit et sous une forme aussi homogène que possible, des définitions et propriétés statistiques* figurant habituellement dans des ouvrages divers, plus ou moins spécialisés, souvent hétérogènes dans leur contenu ou dans leur présentation. Cet objectif doit conduire à un corpus aussi complet et autonome que possible.

2.3. *Fournir des définitions précises et rigoureuses des concepts* statistiques ainsi que des *descriptions concises des méthodes* mises en oeuvre. Cet objectif couvre des domaines très vastes : aussi, quelques règles simples, développées dans l'introduction, ont servi à circonscrire le champ des thèmes abordés. La délimitation de la matière traitée est implicitement résumée dans la **classification générale des thèmes**.

2.4. Mettre à la portée du **statisticien** « **généraliste** » ou de l'**homme de l'art** des informations plus approfondies que celles auxquelles il a usuellement recours et, corrélativement, fournir au « **spécialiste** » des notions relatives à des spécialités différentes de la sienne. Parallèlement, on a cherché à *faciliter l'accès à des notions élaborées* à partir de notions élémentaires.

2.5. *Rassembler*, ce qui est le propre de tout dictionnaire, *les résultats importants* afférents à un même thème (concept, méthode, propriété), notamment afin d'économiser au lecteur l'effort et la dispersion inévitables liés à leur recherche.

2.6. *Dresser un large panorama* des concepts, méthodes et résultats sur lesquels se fonde la Statistique moderne, autrement dit, décrire les *démarches usuelles du statisticien* lorsqu'il traite un problème statistique. Le dictionnaire montre d'ailleurs que, par delà une certaine diversité, nécessaire pour progresser, ces démarches procèdent souvent d'une certaine **unité d'esprit qui peut servir de fil directeur** au lecteur peu expérimenté.

Un **exemple** concret donne un aperçu de ces multiples possibilités.

3. Les **caractéristiques majeures** de ce site sont les suivantes.

3.1. Fruit d'un *travail de compilation*, ce dictionnaire rassemble (objectif d'exhaustivité), trie (objectif de pertinence) et met en forme (objectif de présentation) un domaine qui mérite d'être beaucoup mieux connu (formations, utilisations).

3.2. Le *souci de pédagogie* conduit à doubler le texte statistique (donc mathématique) à l'aide d'un texte littéraire qui autorise, lorsque cela est possible, une interprétation intuitive des notions. Tout en limitant les redondances dans le discours statistique proprement dit, cette méthode permet de mieux expliquer les notions. Un environnement mathématique, constitué des « **méthodes mathématiques de la Statistique** », contribue en outre à préciser ou à rappeler, le cas échéant, un concept ou une notation mathématiques utilisés à titre intermédiaire.

On observe, en effet, que négliger l'aspect pédagogique entraîne souvent, chez l'utilisateur, de mauvaises interprétations ou des retours en arrière sur des notions supposées connues. On peut enseigner la Statistique comme un cours de mathématiques « ordinaire », mais il est nettement avantageux d'associer à cet enseignement la description de divers « contextes » permettant de justifier le sens des démarches choisies par le statisticien.

3.3. Enfin, l'*exigence de rigueur* se traduit par un formalisme concis et par l'adoption d'un système de notations facile à interpréter. Ce système est, autant que possible (cf infra et *notations*), homogène dans la plus grande partie de ce travail.

On a aussi cherché à éviter (a) divers abus de langage (ceux-ci sont, le cas échéant, indiqués aux endroits opportuns) et (b) certains anglicismes dénaturant la langue française, même si l'usage d'un « jargon » technique peut sembler nécessaire.

4. Pour permettre une **consultation pratique et rapide**, le site est organisé comme suit.

4.1. L'**accès direct** vers une notion donnée s'effectue simplement en suivant l'ordre alphabétique des articles (de A à Z), c'est-à-dire par usage des **index analytiques** (liens hypertextes) : en principe (sauf regroupements), il existe un index par lettre de l'alphabet. Un article correspond, généralement, à un terme ou à un concept statistique, ou encore à une expression terminologique ou conceptuelle plus complexe. Il correspond donc à un **concept**, à une **méthode** ou à un **résultat** ; il peut éventuellement en contenir plusieurs (lorsque les notions sont très liées entre elles, ou qu'elles sont très spécialisées). Les **définitions** sont notamment repérables en **caractères gras** dans le texte.

Les index dont il est question reprennent les articles directs, mais aussi les termes ou expressions terminologiques important(e)s localisé(e)s dans le corps des articles. Les divers renvois (liens hypertextes) ou références doivent permettre une consultation aisée, c'est-à-dire rapide et précise.

4.2. L'**accès indirect** vers une notion donnée s'effectue alors grâce aux renvois situés dans le corps des articles auxquels le lecteur a accédé directement. Les index contiennent généralement aussi ces accès indirects. Des redondances, voulues, entre ces divers renvois (liens hypertextes) doivent faciliter les cheminements.

4.3. Hormi l'**environnement mathématique** (méthodes mathématiques de la Statistique), incorporé au dictionnaire sous forme d'articles plus ou moins détaillés, il existe deux catégories d'articles directs :

(a) la première comporte des *articles généraux* qui introduisent une théorie ou une méthode, souvent sous une forme littéraire accessible, et qui renvoient généralement à des articles plus spécialisés ou plus techniques : ceux-ci (concepts, méthodes ou résultats) sont développés ailleurs dans le corps des articles ;

(b) la seconde catégorie regroupe des *articles plus spécialisés ou plus techniques*, auxquels il est fait référence dans la première catégorie, ainsi que d'*autres articles de portée générale* qu'il serait arbitraire de rattacher à une théorie ou à une méthode particulière (formules et inégalités classiques, théorèmes, propriétés ou critères, etc).

4.4. Les *articles* - qu'ils soient banalisés, spécialisés ou techniques - comprennent :

(a) un rattachement à la **classification thématique générale** (codification des thèmes) ;

(b) souvent, une courte **introduction** ;

(c) généralement une **définition**, voire même plusieurs (variantes d'une définition, ou encore suite de définitions de plus en plus générales) ;

(d) des **notations-types** courantes ;

(e) les **propriétés** les plus caractéristiques, lorsqu'elles ne nécessitent pas de développements importants. Dans le cas contraire, il s'agit généralement d'une méthode en elle-même, à laquelle est alors dédié un article ;

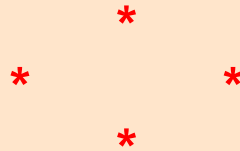
(f) des remarques permettant d'apprécier la portée ou l'utilité de la notion examinée (concept, méthode ou résultat).

4.5. Les *notations mathématiques* ont été uniformisées au maximum. Elles sont ainsi regroupées dans un **index des principales notations**. Les autres notations rencontrées doivent être considérées comme spécifiques de l'article concerné, dans le corps duquel elles sont donc définies. La grande majorité des symboles utilisés est de pratique courante ; à défaut, leur signification est précisée lors de leur utilisation.

4.6. Un **lexique anglais-français** doit faciliter l'accès à la littérature statistique internationale, largement dominée par la langue anglaise. S'agissant d'un texte rédigé en langue française, le lexique inverse (français - anglais) paraissait d'un intérêt limité et n'a donc pas été inséré.

4.7. Enfin, un **index des noms d'auteurs** a été établi. Un travail de cette nature est naturellement redevable des apports passés de la communauté des

statisticiens. Malgré un risque d'arbitraire, l'imputation (à un ou plusieurs auteur(s)) d'un concept, d'une méthode ou d'un résultat a été effectuée : soit en suivant l'usage (règle générale), soit en citant les auteurs qui ont fait progresser la connaissance de façon significative et reconnue. Lorsque plusieurs auteurs sont concernés, ils sont représentés selon l'ordre alphabétique de leur patronyme.



Ce dictionnaire peut donner l'impression d'un « essai » de type BOURBAKI, combinant rigueur mathématique, axiomatisation généralisée, et figeage des développements. Mais, aussi tentante qu'elle soit, une telle prétention aurait certainement été utopique. De plus, elle est démentie par la présence, à tous les niveaux, de développements littéraires et de guides pédagogiques visant à présenter les finalités en fonction du contexte. Ces précautions devraient notamment assurer une pérennité maximale à ce contenu.

Nous espérons avoir ainsi accompli une oeuvre utile et fiable, en offrant des points de départ, préalables ou simultanés, à la lecture d'ouvrages, généraux ou spécialisés, ou de revues périodiques. Chaque article du dictionnaire peut, le cas échéant, être utilement complété en fonction des centres d'intérêt ou des spécialisations.