JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

J. TORRENS-IBERN

Qu'est-ce que la statistique ?

Journal de la société statistique de Paris, tome 97 (1956), p. 289-293
<http://www.numdam.org/item?id=JSFS_1956__97__289_0>

IX

VARIÉTÉ

Qu'est-ce que la statistique ?

PRÉAMBULE

Dans le n° d'avril-mai-juin 1955 du *Journal*, j'ai lu avec beaucoup d'intérêt le texte de la communication développée par M. M. Dumas dans une réunion de notre Société sur « Le statisticien, cet artiste, ou de quelques définitions du mot « Statistique ». La question de la définition de la Statistique a hanté mes pensées presque depuis que j'ai commencé à m'intéresser à cette branche des sciences mathématiques, comme, je pense, celles de tous ceux qui s'y sont passionnés avant ou après moi.

Comme contribution à la réalisation des désirs exprimés par le professeur M. Fréchet je me permets de verser au dossier de la discussion ouverte les définitions qui vont suivre. Ces lignes ne sont rien d'autre que la traduction en langage clair des notes dont je me sers pour expliquer un des cours du programme de « Statistique fondamentale » à l'École d'Ingénieurs Industriels de Barcelone. Elles ont été aussi, en partie, développées dans une communication présentée au Congrès de l'Association française pour l'Avancement des Sciences, en 1948. Étant, donc, en quelque sorte, antérieures à la communication de M. Dumas il n'est pas étonnant que les notes ci-dessous ne s'y rapportent nullement, mais j'ai préféré ne pas retarder leur rédaction plutôt que d'y faire des remaniements et des adjonctions.

1. — Définition mathématique de la statistique.

Il est bien connu l'existence d'un très grand nombre de définitions de la Statistique. Le Dr Willcox a publié dans la *Revue de l'Institut International de Statistique* », en 1935, une collection de 115 définitions (1), laquelle était loin d'être exhaustive. Kendall ose même écrire que « parmi les thèmes à propos desquels les statisticiens ne sont pas d'accord se trouve la définition de leur science » (2).

On connaît la boutade d'un homme d'État anglais qui disait que la statistique est une des formes les plus raffinées que l'homme a inventé pour mentir; bien éloignées d'elle nous trouvons des définitions qui cherchent à nous montrer un certain emploi possible de la statistique. Le « Comité National de l'Organisation française » par exemple, écrit que « la Statistique consiste en une méthode qui, par le relevé en masse et le groupement rationnel des faits, permet de décrire et d'observer les phénomènes collectifs, d'obtenir des rapports numé-

---

(1) W. F. Willcox : « Definitions of statistics » *Revue de l'Institut International de Statistique* (1935), n° 3, p. 388.
(2) M. G. Kendall : *The advanced theory of statistics* (2e édit., 1945), vol. I, p. 1.

riques sensiblement indépendants des anomalies du hasard, de dégager la régularité du changement » (1).

A la rigueur il est possible d'accepter que la Statistique *peut servir* à cela, mais il ne serait pas logique d'affirmer que la Statistique *est* cela. Le faire ainsi équivaudrait à admettre comme mathématiquement valable une définition des logarithmes qui les décrirait comme des collections de tables permettant la résolution simplifiée d'opérations numériques. Il faut, donc, chercher à définir la Statistique par ce qu'elle est et non pas par son utilité.

Pour cela considérons un peu l'ensemble des mathématiques. Nous voyons que dans toutes les branches de cette science il s'agit d'étudier et de déterminer les rapports existants entre certaines quantités qui changent d'un problème à un autre, d'une question à une autre. Nous voyons aussi qu'il existe des transformations appropriées pour obtenir des variables à partir d'autres variables, et qu'on appelle *opération mathématique* le processus de ces transformations. Elles peuvent être entreprises dans les deux sens, ce qui permet de définir des *opérations inverses*, c'est-à-dire des opérations dans lesquelles les résultats sont les données de l'autre opération, et réciproquement. Cette notion d'opération inverse est celle qui va nous permettre de rattacher la Statistique à toutes les autres parties de la science mathématique; c'est seulement en partant de cette notion qu'on arrive à en comprendre clairement l'essence et les possibilités.

Nous pouvons, en effet, définir ensemble comme des opérations inverses le Calcul des Probabilités et la Statistique (2). Le Calcul des Probabilités étudie les théorèmes qui permettent de déterminer les effets lorsque les causes sont connues et il n'existe aucune loi apparemment simple pour les relier univoquement. Si nous avons des urnes qui contiennent des boules blanches et des boules noires mélangées en quantités égales ou différentes mais connues, le Calcul des Probabilités nous enseigne les chances qu'il y a de prélever dans ces urnes des boules blanches et noires en suivant un ordre bien déterminé lorsque nous les prenons une à une ou plusieurs ensemble. Inversons le problème: si nous extrayons de ces urnes, dont nous ne connaissons pas la composition, plusieurs séries de boules, blanches ou noires, la Statistique peut nous permettre d'estimer, avec une certaine approximation, la proportion des boules de chaque couleur qui se trouvent à l'intérieur des urnes. C'est-à-dire que la Statistique permet la détermination de certaines causes lorsque les effets ont été mesurés et nous sont numériquement connus.

Du point de vue d'une philosophie mathématique très élémentaire il est intéressant de s'arrêter un peu à étudier si la dénomination d'opération est appropriée aux analyses probabilistiques et statistiques. Dans toute opération, même dans toute expression mathématique où existent des variables, on peut les classifier en variables indépendantes et variables fonction. Ceci est encore vrai si nous considérons leur caractère physique puisque certaines variables, par exemple le temps avec son irréversibilité, sont essentiellement indépen-

---

(1) P. Donzallaz : *Méthode d'analyse des entreprises*, 3e édit., p. 11.

(2) Nous considérons comme un pléonasme de qualifier de *mathématique* la Statistique. On ne concevrait pas qu'on parle de l'algèbre ou du calcul infinitésimal *mathématiques*. Si la Statistique n'était pas mathématique, que serait-elle?

dantes et ne peuvent pas, du point de vue strictement physique, être regardées comme une conséquence de la conjonction de certaines valeurs des autres variables, c'est-à-dire, en être fonction. L'étude mathématique de ces phénomènes ne s'arrête pas, cependant à ces considérations byzantines et l'on transforme une variable indépendante en une variable fonction par des simples passages des termes d'un membre à l'autre de l'égalité.

Dans notre domaine d'étude nous pouvons définir, au sens physique des termes, les causes d'un phénomène comme des variables indépendantes et les effets produits par le phénomène comme des variables fonction. On peut, alors, en toute rigueur affirmer que le Calcul des Probabilités et la Statistique sont des opérations mathématiques inverses permettant de déterminer respectivement ces variables fonction (effets) et ces variables indépendantes (causes) en suivant des processus analytiques appropriés.

Il ne faut pas, cependant, concevoir la Statistique comme une prolongation, un élargissement ou une généralisation du théorème classique de Bayes sur la probabilité des causes. R.-A. Fisher est déjà sorti au pas de cette interprétation erronée de la Statistique (1). D'un point de vue un peu différent nous pouvons examiner aussi sur un problème concret la dissemblance fondamentale qui existe entre les méthodes statistiques et le théorème de Bayes. Considérons, par exemple, le problème que Poincaré proposa :

« On joue à l'écarté avec un inconnu ; il retourne un roi dès la première partie. Quelle est la probabilité pour que cet inconnu soit un tricheur? »

Laissons de côté les critiques que l'on pourrait faire, suivant les vues de Fisher, à l'énoncé de la question et limitons-nous à considérer l'esprit du problème. Celui-ci, tel qu'il est proposé, appartient d'une façon bien caractérisée au Calcul des Probabilités. Les seuls théorèmes qui interviennent dans sa résolution sont ceux qui constituent les bases du Calcul des Probabilités : le principe des probabilités totales et celui des probabilités composées.

L'application de l'analyse statistique à la résolution d'une question semblable se ferait tout autrement. Pour commencer il ne serait pas concevable qu'un seul événement puisse être considéré suffisant pour établir un jugement sur lui. Le problème que le statisticien se proposerait de résoudre serait plutôt celui-ci :

« On joue à l'écarté avec un inconnu; pendant N parties il retourne n fois le roi. Est-il vraisemblable que l'inconnu soit un tricheur? Ou, mieux, quel est le degré de confiance que nous pouvons avoir dans l'hypothèse que l'inconnu est un tricheur? »

La comparaison de la Statistique avec le Calcul des Probabilités a quelque ressemblance avec la comparaison que des mathématiciens ont fait entre les opérations des Calculs Intégral et Différentiel. En effet, nous lisons dans le livre de Leconte et Delteil (2) : « Après Cauchy, il a été démontré que l'opération de l'intégrale définie s'applique à une classe infiniment plus étendue de fonctions que celle des fonctions continues (Riemann, Darboux, M. H. Lebesgue),

---

(1) R. A. Fisher : *Les méthodes statistiques adaptées à la recherche scientifique* (trad. Y. Bertrand) (1re éd., 1947), p. 7-8.

(2) Th. Leconte et R. Delteil : *Éléments de calcul différentiel et de calcul intégral* (3e éd., 1945), t. I., p. 126.

en sorte que l'intégration apparaît aujourd'hui comme une opération primordiale de l'analyse mathématique. Ce fait est mis en évidence dans les lignes suivantes : « Une fonction continue a nécessairement une fonction primitive. Au contraire, il existe des fonctions continues qui n'ont pas de dérivées. La notion d'intégrale a donc à ce point de vue des propriétés plus simples que celle de la dérivée... » (1). Ainsi qu'avec l'intégration définie on obtient une extension très grande du domaine des fonctions continues, la Statistique élargit beaucoup l'horizon du Calcul des Probabilités; les distributions théoriques des fréquences basées dans celui-ci existent certainement en nombre limité, tandis que celles que la Statistique est appelée à étudier ne le sont pas.

En résumé, donc, *nous proposons de définir la Statistique comme l'opération inverse du Calcul des Probabilités,* sans la confondre cependant avec le théorème de Bayes ou des probabilités inverses. *La Statistique,* en effet, *cherche à estimer non pas la probabilité pour qu'un phénomène soit dû à une certaine cause, mais le degré de vraisemblance de l'hypothèse qui affirme que le phénomène est une conséquence de cette cause-là.*

Nous nous trouvons alors très loin de l'image conventionnelle des longues colonnes de chiffres, non pas pour leur inutilité, mais parce qu'elles ne sont pas la Statistique comme les facteurs ne sont pas la multiplication (2). La Statistique commence lorsque ces nombres nous informent de quelque chose qu'ils contiennent implicitement, de leurs rapports, de leur dépendance, des lois que régissent leur distribution.

2. — La Statistique du point de vue pratique

Dans tous les traités et toutes les études sur la Statistique ont leur place les phrases si connues de Lord Kelvin : « Lorsque vous pouvez mesurer ce dont vous parlez et l'exprimer au moyen de chiffres, vous savez quelque chose de cela, mais lorsque vous ne pouvez pas le mesurer ni l'exprimer numériquement, votre connaissance en est faible et peu satisfaisante ». Ne soyons pas si absolus et acceptons que certaines sciences peuvent se développer brillamment sans qu'il soit nécessaire un accompagnement mathématique à l'exposition de leurs lois. Cependant, il est encore plus évident que l'avancement des sciences et des techniques en général a été facilité par la représentation mathématique de phénomènes étudiés et la prévision numérique de leurs résultats. La géométrie nous apprend à calculer la longueur de la circonférence et l'aire du cercle à partir de leur diamètre. La loi d'Ohm nous permet de déterminer l'intensité du courant électrique qui circule dans un conducteur lorsque nous connaissons la différence des potentiels de deux points de ce conducteur et sa résistance ohmique. La résistance des matériaux nous donne les formules nécessaires pour choisir les dimensions d'une poutre qui se trouve soumise à certains efforts et appuyée en des points et dans des conditions déterminées.

---

(1) Hadamard : *Cours d'Analyse de l'École Polytechnique.*

(2) Dans le livre de vulgarisation statistique de M. J. Moroney, *Facts from figures,* Pelikan Books n° A 236 (1951) on peut lire (p. 82) en épigraphe d'un chapitre : « Lorsque j'étais plus jeune, la Statistique était la science des grands nombres. Maintenant il me semble qu'elle devient rapidement la science d'aucun nombre. » Dr. Oswald George. »

Si nous considérons d'un peu près ces différents exemples nous nous apercevrons, cependant, qu'entre eux il n'existe pas d'identité absolue. Alors que dans le problème géométrique l'exactitude du résultat dépend en premier lieu de celle avec laquelle nous aurons mesuré le diamètre et de la rigueur de l'évaluation de π — coefficient que l'on peut obtenir avec autant de précision que l'on veut par des méthodes de raisonnement pur (en 1874 Shanks en a calculé, par exemple, 707 chiffres exacts) — dans les autres cas l'inexactitude provient essentiellement de la connaissance toujours incomplète qu'on peut avoir des qualités physiques de la matière considérée. Les essais des matériaux nous informent bien des coefficients à employer dans nos calculs, mais ces coefficients n'ont pas la consistance immuable de π; ils varient, en général, plus ou moins d'une éprouvette à une autre, même sans changer de matériau, de façon que l'on peut très bien obtenir autant de valeurs différentes d'un certain coefficient que mesures expérimentales on aura réalisées.

La variabilité de ces coefficients ne provient que des seules erreurs de mesure (qui intervenaient déjà dans l'évaluation des dimensions géométriques) mais aussi de l'impossibilité de maintenir à des niveaux exactement constants les facteurs multiples qui influent d'une façon ou d'une autre sur le caractère étudié : composition chimique, traitement thermique préalable, défauts physiques même de grandeur infime, température ambiante, et beaucoup d'autres causes qui, si prises séparément sont d'un effet inappréciable, par leur somme ou par leur interaction, modifient parfois un chiffre plus ou moins proche de la virgule du coefficient en question.

Dans l'étude des sciences expérimentales la rigueur des liaisons fonctionnelles des mathématiques classiques souvent disparaît. A leur place apparaît une gamme infinie de liaisons stochastiques, plus ou moins lâches, qui peuvent arriver jusqu'à l'indépendance complète. Il faudra donc prendre en main l'outil statistique chaque fois qu'il s'agit d'étudier les rapports existants entre des variables dont les valeurs ne sont pas en correspondance univoque, ce qui se produit lorsqu'il y a beaucoup de facteurs, beaucoup de causes, qui influent sur les résultats, ou lorsque trop de facteurs nous sont inconnus, trop de causes indéterminées. Le domaine des méthodes statistiques est donc très vaste; les exemples de leur application se trouvent partout où les méthodes des mathématiques classiques font défaut par manque de données complètes du phénomène étudié. La Statistique permet alors d'exploiter au maximum l'information recueillie en lui faisant subir son analyse critique.

En nous plaçant au point de vue pratique nous pouvons donc dire que *la Statistique est une méthode mathématipue d'analyse qui permet d'étudier avec le maximum de précision les phénomènes incomplètement connus* (1).

J. Torrens-Ibern.

***

(1) Cette définition a été inspirée par celle citée ici même par M. A. Sauvy : « L'art de préciser les choses que l'on ignore » (*Journal de la Société de Statistique de Paris*, nos 4-5-6, avril-mai-juin 1952, p. 89).