MATHÉMATIQUES ET SCIENCES HUMAINES

G. MORLAT

Statistique et théorie de la décision

Mathématiques et sciences humaines, tome 8 (1964), p. 1-7

<http://www.numdam.org/item?id=MSH_1964__8__1_0>

G. MORLAT

STATISTIQUE ET THEORIE DE LA DECISION*

Même si les statisticiens, qui sont gens raisonnables, s'en plaignent rarement, la statistique est une discipline bien mal nommée, et de ce fait, parfois mal renommée. Combien de nos contemporains pensent encore, avec un personnage de Courteline, que le métier du Statisticien consiste à compter le nombre de sexagénaires qui traversent le Pont Neuf un dimanche après-midi.

Il est bien vrai que le rôle de la statistique, il y a un siècle ou deux, c'était de compter les choses.

"Science, dit Littré, qui a pour but de faire connaître l'étendue, la population, les ressources agricoles et industrielles d'un Etat.... Les économistes ont créé un mot pour désigner la science de cette partie de l'économie politique - les dénombrements - et l'appellent statistique".

Cette vocation de compter les choses dont on pouvait se gausser au siècle dernier, a revêtu quelque noblesse, depuis qu'on a reconnu combien elle est essentielle à la bonne marche des systèmes économiques, et donc des sociétés humaines. D'ailleurs, si les Incas prenaient très au sérieux, avec raison ceux de leurs fonctionnaires qui étaient chargés de faire des noeuds sur des ficelles de couleurs diverses, tenant ainsi à jour les statistiques vitales de leur Etat - comment nos contemporains n'auraient-ils pas de la considération pour des hommes qui disposent maintenant de machines à calculer électroniques, puissantes et coûteuses, qui savent avec une bonne précision de quoi sont faites la population et la production de leur pays et qui peuvent fournir, sur demande, au gouvernement, des chiffres pour éclairer ses décisions?

Nous devons donc constater que la statistique, au sens de Littré, a connu au cours des dernières décennies une promotion considérable, et cela suffirait à en faire une discipline pour l'honnête homme. Mais en même temps le terme de statistique a progressivement désigné un domaine d'activité intellectuelle infiniment plus vaste, et c'est de cette évolution là que je veux parler.

La première phase d'un traité élémentaire de statistique, publié il y a quelques années aux Etats-Unis, est celle-ci: "La statistique est un ensemble de méthodes pour prendre des décisions raisonnables en présence d'incertitudes". Comme toutes les décisions humaines sont prises en face d'incertitudes, cette modeste phrase indique donc que la statistique, c'est la méthode pour prendre des décisions. Nous voici assez loin de la science des dénombrements. Par quel chemin la statistique - sans changer son étiquette, ce en quoi elle a peut être eu tort, puisque c'est source de malentendus, mais il est trop tard de toute façon pour y revenir - par quel chemin a-t-elle pu passer des dénombrements à la théorie des Décisions? C'est

---

* Cet article est le texte d'une conférence prononcée en Novembre 1963 au cours d'une réunion commune à la Société de Statistique de Paris et à la Société des Ingénieurs Civils de France.

ce que nous proposons ici de mettre en lumière, en retraçant sommairement les étapes par lesquelles la méthode statistique a progressé; nous n'aurons guère le souci de respecter l'ordre historique de cette évolution, mais seulement de montrer comment s'enchaînent, d'une manière toute naturelle, les diverses conceptions de la statistique qui vont de la "comptabilité des choses" à la logique des décisions.

*

* *

Les premières étapes sont les mieux connues. Pour rendre commodes les dénombrements, et pour présenter leurs résultats sous forme succincte et maniable, on a l'habitude de calculer des moyennes et des écarts moyens, d'établir des polygones de fréquence, ou des histogrammes, des fibrogrammes, des courbes d'observations classées, et quelques autres représentations commodes. La manière d'établir des tableaux à double entrée - ou davantage - les graphiques représentant les variations d'un phénomène selon les valeurs d'un autre, le concept de régression, et le calcul des courbes de régression par la méthode des moindres carrés, la définition et le mode de calcul des divers indices de liaison entre des séries d'observations, tels le coefficient de corrélation, le carré moyen de contingence, et beaucoup d'autres - tout cela fait partie de la statistique descriptive; c'est l'art et la méthode pour manier, décrire et résumer des observations nombreuses. Mais lorsqu'il s'agit d'interpréter - lorsqu'on se demande, par exemple, si deux séries d'observations peuvent être regardées comme homogènes - ou si tel phénomène dépend de tel autre - les recettes de la statistique descriptive ne nous permettent guère de fonder des raisonnements solides. Il faut, pour de telles interprétations, faire appel à des modèles, et les modèles convenables sont fournis par le calcul des probabilités.

Est-ce suffisant? Le calcul des probabilités, au sens strict nous apprend par exemple comment on peut tirer, d'un modèle probabiliste donné, la loi de probabilité de la moyenne d'une série d'observations. Cela éclaire grandement la situation, mais le statisticien voudrait, connaissant des observations, et ayant quelque idée sur les modèles convenables, préciser ces modèles ou les mettre à l'épreuve: il veut induire et non pas déduire. Voilà pourquoi le calcul des probabilités n'est pas suffisant, et pourquoi il a fallu élaborer une discipline nouvelle, étroitement fondée sur le calcul des probabilités, mais parfaitement originale cependant, par les problèmes qu'elle aborde, et par son objet précis, qui est de formaliser l'induction, et de fournir le modèle de ce que l'observation nous apprend des modèles. Cette nouvelle discipline a continué à s'appeler statistique - ou pour être plus précis, on la dénomme statistique mathématique.

On convient généralement de considérer que la première pierre de la statistique mathématique fut posée par Karl Pearson, avec son célèbre mémoire de 1900 dans le Philosophical Magazine: "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". En réalité, le problème de l'induction statistique avait été bien vu depuis fort longtemps, et le Révérend Thomas Bayes avait écrit un mémoire, publié après sa mort, en 1764 dans les Philosophical Transactions - "An essay towards solving a problem in the doctrine of chances" - où il proposait une solution générale à ce problème. Mais la "formule de Bayes" - dont Condorcet, Laplace et beaucoup d'autres usèrent et abusèrent parfois - pouvait paraître bien arbitraire. Au début de ce siècle il semblait souhaitable d'établir la statistique mathématique comme une discipline indépendante d'éléments subjectifs ou arbitraires, comme les probabilités a priori de

Bayes - et il est assez plaisant de constater que c'est en s'assignant un tel but, que nous savons aujourd'hui littéralement insensé, que de grands statisticiens tels que Karl Pearson, Sir Ronald A. Fisher, Jerzy Neyman (pour ne citer que les noms les plus justement célèbres) ont pu en effet échafauder les principaux chapitres de la statistique mathématique, et donner à cette science une impulsion telle qu'elle a connu, au cours des dernières décennies, d'immenses succès théoriques et pratiques, et qu'elle est devenue indispensable dans la plupart des domaines d'activités scientifiques.

*

* *

Mais les différents problèmes de la statistique mathématique - estimations des paramètres, tests d'hypothèses, analyse de la variance, plans d'expériences, statistiques d'ordre, etc... - pouvaient apparaître comme une collection de problèmes ayant certes en commun ce trait fondamental d'impliquer une induction, des observations au modèle, mais pour le reste assez différents entre eux, et presque hétéroclites, en ce qui concerne leurs formalisations respectives. Il a fallu attendre l'oeuvre d'Abraham Wald pour voir se réaliser l'unité de la statistique mathématique, sous la forme de la Théorie des Fonctions de Décision Statistique, dont les divers chapitres que nous avons énumérés à l'instant apparaissent comme des cas particuliers. C'est une étape un peu moins généralement connue, et nous nous y arrêterons un instant, pour décrire d'une façon sommaire la Théorie proposée par Wald. L'intérêt est double à nos yeux: montrer l'unité des problèmes de la statistique mathématique et mettre en lumière la nécessité de fondements logiques que l'on ne peut trouver que dans une autre théorie, dont nous dirons quelques mots in fine.

Un problème statistique comporte les éléments formels suivants:

L'ensemble des observations, **X**, est l'ensemble de tous les "points" x qui constituent tous les résultats d'observations possibles a priori, pour le phénomène étudié. Naturellement, x peut être un nombre, un vecteur, une fonction du temps, ou tout autre élément que l'on voudra.

Les diverses lois de probabilité prises en considération, F, forment un ensemble Ω. Ce sont des distributions de probabilité sur **X**, et elles peuvent être en nombre fini, ou bien former une famille paramétrique, ou être définies de quelque autre façon.

Le statisticien doit, au vu des observations, prendre une décision. Les décisions possibles d forment un ensemble **D**. Dans certains cas, on peut être amené à prendre en considération le tirage au sort entre plusieurs décisions, avec des probabilités choisies au mieux: cela signifie qu'on prend en considération l'ensemble **D*** des distributions de probabilité sur **D**. Du point de vue qui nous occupe ici, cela constitue un détail technique, dont nous ne nous encombrerons guère par la suite.

Le problème posé est de choisir une décision d chaque fois qu'on disposera d'observations x, c'est-à-dire de choisir une manière de faire correspondre un d à toutx; en d'autres termes, nous devons envisager les applications de **X** dans **D**. Une telle application, δ , constitue ce que Wald appelle une fonction de décision, et nous noterons Δ l'ensemble des fonctions de décision δ.

Maintenant, en vertu de quelles considérations peut-on être amené à choisir, une fonction de décision plutôt qu'une autre? Si l'on prend une décision d, alors

que la loi qui représente le mieux le phénomène étudié - ce que nous appellerons conventionnellement la "vraie" loi - est F, alors on encourt certains avantages et certains désagréments. Admettons, avec Wald, que ces conséquences peuvent être représentées valablement par un nombre, que nous appellerons la "perte": ce nombre est donc une fonction de F et de d, que nous représenterons par

$$W\ (F,\ d) \qquad \text{(fonction de perte)}$$

Dans certains problèmes de contrôle des fabrications, ce concept peut être rattaché à une perte au sens ordinaire, chiffrée en argent (ce sera quelque chose comme l'espérance des divers coûts actualisés résultant d'une décision de rejet ou d'acceptation d'un lot de fabrication). Mais dans la plupart des problèmes statistiques, il faut admettre que la fonction de perte représente, forfaitairement, des inconvénients de toute nature: on verra plus loin qu'il suffit de lui attribuer des formes extrêmement simples pour retrouver et dans une certaine mesure justifier, les techniques classiques de la statistique.

On doit choisir, avons-nous dit, non pas une décision, mais une fonction de décision: la fonction de perte n'est qu'un intermédiaire, et l'on attachera à toute fonction de décision δ l'espérance de la perte correspondante, ce que nous écrivons:

$$r\ (F,\ \delta) \quad = \quad \int W\ (F,\ d) \qquad dF \qquad (x)$$

et que nous appellerons, en reprenant toujours la terminologie de Wald, fonction de risque. Le sens de ce nouveau concept peut être rendu plus explicite à une distribution de probabilité F sur X, la fonction de décision δ, qui est une application de X dans D, associe une distribution de probabilité sur D, et en conséquence également une distribution de probabilité pour W: d'où la possibilité d'en calculer l'espérance, qu'exprime bien la formule ci-dessus.

La considération de la fonction de risque, permet-elle de choisir une règle de décision? Ce sera vrai dans des cas exceptionnels où il existera une fonction de décision δ_0 rendant minimum le risque quel que soit F. Mais en général, on peut s'attendre à ce que la règle de décision, qui minimise le risque dépende de F. Avant de signaler comment on peut surmonter cette difficulté, illustrons les notions que nous venons d'introduire, en examinant quelques problèmes statistiques classiques.

*

* *

On connaît le problème de l'estimation d'un paramètre: la famille Ω des lois de probabilité est par exemple une famille à un paramètre, F (x, θ) et on veut, d'après des observations, assigner une valeur approximative à θ. C'est bien un cas particulier des problèmes que nous venons de décrire. Une décision est constituée par la valeur t qu'on attribuera au paramètre, une règle de décision n'est pas autre chose qu'une fonction t(x) - c'est ce qu'on appelle en statistique mathématique un estimateur.

Supposons que la fonction de perte admise soit proportionnelle au carré de l'écart entre la valeur vraie du paramètre et sa valeur estimée:

$$W = k\ (\theta - t)^2$$

Alors la fonction de risque sera, dans le cas d'un estimateur sans biais:

$$r\left[\theta, t(x)\right] = \int k\,(\theta - t)^2 \quad dF\,(x) = k \text{ var } t.$$

Minimiser la fonction de risque, c'est donc ici rechercher les estimateurs de variance minimum: on voit comment on pourra justifier la méthode du maximum de vraisemblance et quelques autres méthodes classiques d'estimation.

Considérons maintenant un problème de test - et pour simplifier l'exposé, nous admettrons qu'on s'intéresse exclusivement à des modèles entièrement déterminés: on veut tester une hypothèse simple F, contre une alternative simple, G. L'ensemble X étant au reste quelconque, Ω est formé des deux lois F et G. L'ensemble des décisions possibles comporte aussi deux éléments, à savoir:

d_1 : conserver l'hypothèse F

d_2 : rejeter l'hypothèse F

Dès lors, une fonction de décision est simplement une partition de X en deux sous-ensembles - ou encore, c'est la donnée d'une région critique, ou région de rejet.

Prenons pour fonction de perte la fonction représentée par le tableau ci-dessous:

	F	G
d_1	0	1
d_2	1	0

Autrement dit, la perte est nulle quand on prend une décision correcte, égale à l'unité quand on commet une erreur. On calcule sans peine la fonction de risque, qui vaut α pour F et β pour G, α désignant le seuil de signification du test, et β la probabilité d'erreur de seconde espèce. Là encore, nous retrouvons tout naturellement des notions classiques.

Dans l'exemple de l'estimation - et du moins pour la plupart des formes de lois de probabilité d'usage courant - l'estimateur de variance minimum, ou l'estimateur de Fisher, ne dépend pas de la valeur du paramètre que l'on cherche à estimer: nous sommes dans ce cas privilégié où il existe une fonction de décision uniformément meilleure que toutes les autres - cette proposition étant vraie sans restriction lorsqu'il existe un estimateur exhaustif, et n'étant vraie qu'asymptotiquement dans d'autres cas.

Nous constatons qu'il en va tout différemment dans l'exemple des tests; même dans les cas les plus simples, il faut choisir un compromis entre les deux risques d'erreurs α et β: on sait bien qu'il n'existe pas de test qui rende à la fois α et β aussi petits que possible.

C'est la difficulté déjà signalée plus haut: la fonction de risque dépend des deux arguments F et δ. Il faut décider d'un principe de choix.

Wald a suggéré qu'on pourrait trancher cette ultime difficulté en adoptant le principe du minimax: la fonction de décision retenue serait celle qui réalise la condition:

$$\min_{\delta} \max_{F} r(F, \delta)$$

C'est une règle de prudence extrême, et l'analogie formelle entre les problèmes de la statistique et ceux de la Théorie des Jeux à deux joueurs a pu rendre cette règle fort séduisante. Mais il convient de noter que la justification du minimax dans la Théorie des Jeux réside essentiellement en ceci, que le minimax conduit à un point d'équilibre: si l'un des joueurs s'écarte de la stratégie optimale, il sera pénalisé. Dans les problèmes statistiques, l'un des joueurs est le statisticien, l'autre est la "nature", qui est censée choisir la loi de probabilité F inconnue. On ne peut guère considérer sérieusement que les intérêts de la nature sont opposés à ceux du statisticien - et le minimax perd de ce fait toute espèce de justification.

Wald a étudié par ailleurs une catégorie de fonctions de décision dont il a montré tout l'intérêt: ce sont les fonctions de décision de Bayes, c'est-à-dire celles qui réalisent une condition du type

$$\min_{\delta} \mathop{E}_{\xi} r(F, \delta)$$

ξ étant une distribution de probabilité sur l'ensemble Ω des lois de probabilité F (ξ est appelée habituellement "distribution de probabilité a priori"). Wald a montré que, dans des cas très généraux, les fonctions de décision de Bayes sont les seules fonctions de décision admissibles - c'est-à-dire telles qu'il n'existe aucune autre fonction de décision dont le risque serait plus faible pour tout F. Aux yeux de Wald, il s'agit là d'une propriété purement formelle et la distribution de probabilité a priori n'est qu'un élément formel intermédiaire, permettant de sélectionner une classe de fonctions de décision intéressante: c'est que la statistique était encore fortement tournée, il y a une dizaine d'années, vers la recherche de principes objectifs).

Aujourd'hui, un nombre croissant de statisticiens admettant que la recherche d'un principe de choix objectif est vaine. Les problèmes de décisions statistiques dont nous venons d'esquisser sommairement les contours, entrent dans la catégorie des problèmes de décision en face d'incertitudes, et la Théorie Générale des Décisions nous apprend que des choix cohérents, ou rationnels - la définition précise de ces objectifs ne peut être traduite dans des postulats aussi convaincants qu'on peut le souhaiter - de tels choix sont nécessairement ceux que l'on commettrait en maximisant l'espérance d'une fonction d'utilité des conséquences - ou, ce qui revient au même, en minimisant l'espérance d'une fonction de perte convenablement choisie. L'espérance doit être prise par rapport à une distribution de probabilité traduisant tous les éléments d'incertitude sur lesquels n'influe pas la décision du statisticien - et en particulier des probabilités doivent être affectées aux éléments de l'ensemble Ω des modèles.

C'est dire que la seule façon de donner un fondement logique solide à la statistique, semble bien être de considérer que les distributions a priori de Wald représentent un certain degré de connaissance, ou de confiance, préalable, à l'égard des diverses lois F.

Cela, c'est en quelque sorte le droit le plus général. Dans la pratique faut-il recommander de traduire toujours les connaissances a priori par une distribution de probabilité, et considérer que tous les efforts de la statistique classique - celle qu'on appelait moderne il y a dix ans - doivent être passés par pertes et profits? Certains n'hésitent pas à l'affirmer.

Personnellement, nous admettons sans réserve que la Théorie Générale des Décisions, ou encore Théorie de la Probabilité-Utilité, constitue le seul modèle logique valable. Mais le choix effectif des probabilités a priori éveille dans beaucoup de

situations pratiques des scrupules trop vifs pour qu'on puisse penser qu'ils sont dénués de fondement. Il convient de juger en connaissance de cause, et dans chaque cas particulier, quels inconvénients pèsent le plus lourd, de ceux d'un choix hasardeux des probabilités a priori, et de ceux d'une règle de choix conventionnelle, permettant de s'en passer. Heureusement, dans les problèmes les plus courants de la statistique, on peut reconnaître que ce dilemne est facilement tranché. Nous avons eu l'occasion de constater que le problème de l'estimation des paramètres donne lieu, sous certaines conditions, à une règle de décision optimale évidente. On peut montrer, d'autre part, que dans beaucoup de problèmes, de tests, des hypothèses très généralement raisonnables concernant à la fois des probabilités a priori des diverses hypothèses et les coûts des diverses erreurs, permettent de justifier les méthodes classiques - de montrer que la règle optimale, au sens de la Théorie des Décisions, s'écarte assez peu de la convention courante, qui consiste à choisir des tests ayant un niveau de signification faible. Bien entendu, cette propriété n'est nullement générale, c'est une des raisons pour lesquelles, il nous semble nécessaire que les Ingénieurs, de même que les autres utilisateurs possibles de la statistique, soient informés des rapports entre les techniques statistiques et la Théorie de la Décision; nous avons essayé, d'une façon beaucoup trop succincte sans doute, d'en donner seulement les grandes lignes. Peut être eût-il été bon d'esquisser avec plus de précision le contenu de la Théorie Générale des Décisions; mais ceci est une autre histoire....

- - - - - - - - - - - - -