L'INITIATION A LA STATISTIQUE AU NIVEAU SECONDAIRE *

H. BRENY

Université de l'Etat, Liège

1. En Belgique, la question d'existence d'un enseignement de probabilités et statistique au niveau secondaire ne se pose plus : en 1967, les Ministres Grootjans et Toussaint ont chargé une commission de représentants de l'enseignement supérieur d'établir la liste des matières du programme de mathématique; le point 9 de cette liste est « statistique descriptive, probabilités, éléments de statistique inférentielle ». La distribution de cet enseignement par années scolaires n'est pas officiellement connue à l'heure actuelle, on sait pourtant que le programme de 3^{me} scientifique comporte l'étude de la statistique descriptive et on peut présumer que « théorie des probabilités » sera placé en 2^{me}, et « éléments de statistique inférentielle » en 1re. C'est d'ores et déjà la disposition adoptée (à un niveau de moindre difficulté mathématique) pour la section « Sciences humaines », récemment créée.

Le libellé de cette liste des matières tranche aussi une question qu'il reste pourtant intéressant de se poser in abstracto : lorsque l'on dit « théorie des probabilités », s'agit-il, dans ce contexte, de la théorie des phénomènes aléatoires ou de la théorie du comportement cohérent en régime d'incertitude (maximisation de la moyenne par rapport à des probabilités subjectives d'un indicateur cardinal d'utilité personnelle) ? Cette dernière théorie (dont l'intérêt est évident), quelle que soit la façon de l'exposer, repose en quelqu'endroit sur la théorie d'une certaine classe de phénomènes aléatoires, dont l'étude est présupposée; en outre, son formalisme (ensembles convexes, avec un appel inévitable au théorème de von Neumann et Morgenstern), son interprétation (rapport entre comportement efficace et comportement cohérent), ses limites posent des problèmes difficilement accessibles à un adolescent de 17-18 ans. Aussi est-ce à bon droit que la liste des matières a été libellée de façon à ne laisser aucun doute sur ce point : il s'agit bien de la théorie des phénomènes aléatoires.

---

* Conférence faite à la « Journée Statistique » organisée par la Société Belge de Statistique, le 10 février 1971 au Palais des Congrès à Bruxelles.

2. Si la question d'existence ne se pose plus, on peut pourtant s'interroger sur la finalité d'un tel enseignement. On le peut d'autant plus que certains y ont vu l'occasion de développer l'exposé de recettes statistiques utilisables dans divers domaines, un peu à la façon de quelques cours de 1re année d'université où l'on développe, vaille que vaille, les recettes dont le professeur pense avoir besoin par après. Ce serait là une erreur manifeste. Certes, l'utilité d'un enseignement, quel qu'il soit, n'est pas un facteur à négliger (et l'inutilité n'est pas un motif positif de choix) mais, dans l'enseignement secondaire général, ce n'est pas, et de loin, un facteur déterminant; le cours de mathématique, dans son ensemble, est utile à pas mal d'élèves, mais son but essentiel est la formation de l'esprit. Le chapitre « probabilités et statistique » fait partie de ce cours, et sa finalité essentielle est la même. Il ne s'agit pas d'y inculquer, par une sorte de dressage, des techniques (même utiles), il s'agit d'ouvrir l'esprit des élèves à une certaine façon de voir le monde, et de le préparer à recevoir ultérieurement, si le besoin s'en fait sentir, un enseignement davantage professionnel. Les trois parties de ce chapitre ont d'ailleurs chacune sa finalité propre.

3. La modeste statistique descriptive elle-même a une incontestable finalité culturelle. Nos élèves, en effet, doivent retirer de leur cours de mathématique, entre autres bénéfices, une véritable familiarité avec les données numériques (ce que les Anglais nomment « numeracy »); le cours de statistique descriptive sera pour eux l'occasion (unique, peut-être, au niveau secondaire), de rencontrer des tableaux de nombres ne présentant, a priori, aucune régularité mathématique. Il est pourtant vrai que la statistique descriptive telle qu'elle est ordinairement présentée, c'est-à-dire une suite de recettes de calcul, n'offre rien qui accroche l'intérêt des élèves. Il y a à cela deux remèdes possibles. L'un consiste à adopter systématiquement une présentation vectorielle : les tableaux de n nombres réels forment tout naturellement un espace vectoriel, que l'on dote d'une structure euclidienne par l'introduction du « produit scalaire statistique » $(\Sigma_i x_i y_i)/n$; cette façon de faire est très élégante et offre un beau terrain d'exercices vectoriels, mais, à notre avis, limite trop la statistique descriptive à son aspect linéaire (auquel échappent, par exemple, la médiane et les quantités analogues; encore faut-il admettre que l'objection perd sa force en statistique multivariée). L'autre consiste à arriver le plus vite possible au diagramme cumulatif associé à tout tableau de nombres réels, et d'utiliser cette notion centrale pour motiver et définir toutes les autres : le chapitre de statistique descriptive reçoit ainsi une organisation très nette et très claire, et les recettes de calcul apparaissent toutes comme les traductions explicites de définitions intrinsèques préalables.

4. Le cours de calcul des probabilités a comme fin essentielle d'attirer l'attention des élèves sur le caractère aléatoire de maint phénomène du monde sensible. Dans l'état actuel de notre enseignement, c'est — au plus tôt — en 2^me^ année d'Université que l'étudiant rencontre le fortuit; c'est, pour beaucoup, trop tard : leurs études les ont déjà conditionnés au déterminisme, et le calcul des probabilités (y compris la statistique inférentielle) leur paraît irrémédiablement artificiel; de ce point de vue, l'introduction de la théorie des probabilités au programme de la classe de 17 ans sera chose excellente. L'on doit pourtant regretter que les contraintes de temps (et aussi l'âge des élèves) conduisent à renoncer presque entièrement aux manipulations (dés, cartes, urnes, voire comptages de particules ionisées); des expériences actuellement en cours permettent d'espérer que l'école primaire pourra prendre à son compte une initiation au fortuit par les manipulations concrètes (et les instituteurs seront mieux préparés à les diriger lorsque leur propre programme d'études comportera un cours de calcul des probabilités). Le formalisme habituel de la théorie des probabilités (événements, lois de probabilités, variables aléatoires, indépendance) est évidemment l'objet principal de l'ensignement (rappelons qu'il s'agit du cours de mathématique). Ce n'est pourtant pas le seul; il est assez étonnant de voir des professeurs de mathématique (ou des auteurs de manuels) défendre parfois la thèse que « les mots « hasard » ou « phénomène fortuit » n'ont pas de sens mathématique précis, qu'il faut donc les exclure de l'enseignement de la théorie des probabilités, qui est un formalisme mathématique parmi bien d'autres, et rien de plus ». La réponse est fournie par une excellente remarque des « Directions méthodologiques » du Ministère de l'Education Nationale (programme de 2^me^ « Sciences humaines ») : « la partie du programme consacrée aux probabilités doit apprendre à mathématiser une situation aléatoire ». La mathématique d'aujourd'hui, en effet, qui intéresse par elle-même une minorité (les mathématiciens), intéresse par ses possibilités d'application (par sa polyvalence, dit-on volontiers) un nombre considérable de personnes. [C'est bel et bien au professeur de mathématique qu'il incombe de montrer comment, sortant en quelque sorte de lui-même, le formalisme mathématique se met au service de la science du monde sensible; qui d'autre que lui, en effet, le pourrait ?] Or, la théorie des probabilités offre l'occasion de montrer aux élèves, sur des situations à la fois neuves (pour eux) et relativement simples, comment on construit un modèle mathématique d'une réalité donnée pour en imiter théoriquement le comportement et, par là même, le comprendre (en un sens assez particulier) et, s'il le faut, l'utiliser. C'est une occasion qu'il ne faut absolument pas manquer.

Une part non négligeable du cours de probabilités est consacrée aux variables aléatoires (nouvel avatar de l'omniprésente notion de fonction) et à leurs valeurs typiques. Il importe de souligner que si le cours de statistique descriptive a été organisé de façon convenable, les analogies formelles entre un diagramme cumulatif et une fonction de répartition permettent de gagner un temps précieux, sans rien sacrifier du juste esprit de rigueur mathématique. Dans le même ordre d'idées, il serait bon que, même si les seuls modèles effectivement étudiés sont des modèles finis, les définitions et les principaux énoncés soient suffisamment généraux pour pouvoir être repris tels quels dans une étude ultérieure (un bon exemple est fourni par la notion d'indépendance : des élèves habitués, au départ, à parler uniquement de l'indépendance d'événements ont quelque peine à assimiler la nécessaire généralisation à des familles d'événements).

5. En ce qui concerne le cours de statistique inférentielle, la liste des matières en limite très sagement le contenu à l'essentiel : les notions de population et d'échantillon, et un problème d'estimation (« estimation de la moyenne d'une population à partir d'un échantillon nombreux »).

On ne conçoit guère qu'un enseignement de statistique inférentielle puisse éviter de commenter d'abondance les notions de population et d'échantillon. Dans le contexte avant tout culturel du cours envisagé pour la classe terminale du secondaire, il faut même aller plus loin, et mettre très nettement en évidence la notion d'échantillonnage comme phénomène fortuit, ce qui fait d'emblée de ce cours une suite, et une très importante application, du cours de la classe de seconde. Les « théorèmes d'échantillonnage » sont le point de départ de toute procédure statistique, il convient donc d'en énoncer un certain nombre et de les faire bien comprendre. Il est vrai que rares sont les théorèmes de ce type que l'on pourrait démontrer (rares, mais non inexistants : le calcul de la moyenne et de la variance d'échantillonnage de la moyenne d'échantillon est parfaitement possible); au demeurant, même une démonstration en bonne et due forme ne suffirait sans doute pas à en éclairer le sens. Aussi devra-t-on avoir recours à des expériences à faire en classe, à partir d'une population artificielle (une « urne » contenant des jetons marqués de valeurs numériques bien choisies); il n'est pas difficile de faire calculer par les élèves l'ogive de la population (car, ici, on la connaît, puisqu'on la fabrique !), de leur faire extraire ensuite une centaine d'échantillons, calculer la moyenne de chacun d'eux, appliquer à ce tableau de nombres les méthodes de la statistique descriptive, et comparer les résultats obtenus aux caractéristiques de la population.

C'est ce même type d'expériences qui devra être utilisé pour illustrer le théorème de tendance normale, qui est facile à énoncer mais impossible à démontrer (à ce niveau). A partir de ce théorème, la propriété essentielle de l'intervalle de confiance pour la moyenne de population peut en effet se démontrer; mais, ici encore, une illustration expérimentale sera la bienvenue, pour éclairer définitivement cette notion difficile.

6. Ainsi équipé, l'élève sortant du secondaire aura des idées assez nettes sur le noyau central des méthodes statistiques. Qu'après cela l'enseignement supérieur lui impose un certain nombre de « recettes » pratiques, sans démonstration ni éclaircissement, sera peut-être un mal inévitable : il disposera du moins d'une base suffisamment solide pour établir des analogies qui lui permettront de comprendre, et de ne pas appliquer ces recettes de façon purement mécanique. Si, comme il faut l'espérer, il échappe à un pareil endoctrinement, il aura en tout cas enrichi son esprit d'une vue du monde de l'aléatoire, et de notions d'un grand intérêt, même pour quiconque n'aurait pas à les appliquer. Mais qui, de nos jours, échappe entièrement à la statistique ?