Yves Chevallard et Floriane Wozniak

# ENSEIGNER LA STATISTIQUE AU SECONDAIRE

*Entre genre prochain et différence spécifique*

**Abstract :** The recent reform of the teaching of statistics in French secondary schools comes up against two difficulties which this paper tries to analyse. The first one lies in the fact that teachers of mathematics have gradually unlearnt to teach the non-purely mathematical parts of the mathematics curriculum since the so-called "new mathematics" reform, whereas statistics belongs to "mixed" mathematics. The second yet more specific difficulty originates from the fact that statistics mostly deals with variability and uncertainty, two notions usually ignored and even repressed in the worldview shared by most people and institutions.

## 1. ENSEIGNER LA STATISTIQUE ?

### *1 Généricité et spécificité en didactique*

Précisons d'abord, à titre de principes organisateurs de la réception de notre propos, les hypothèses cruciales qui sous-tendent notre travail. Tout d'abord, nous nous situons dans une problématique de l'activité scientifique qui refuse de reprendre à son compte le grand partage, vieux de quelque quatre siècles, entre sciences de la nature d'un côté et sciences de l'homme et de la société de l'autre. Nous considérons au contraire, dans une visée unitaire, que *toute* science, qu'elle soit « de la nature » ou « de l'homme et de la société », se donne pour objet d'étude un certain ensemble *de conditions et de contraintes de la vie des sociétés*. Le lien social, par exemple, est bâti sur des contraintes physiques, chimiques, physiologiques, mathématiques que les sciences correspondantes élucident et, dans le meilleur des cas, permettent de déplacer ou d'annuler. Et on ne conçoit guère, dès lors, de sciences « sociales » qui n'intègrent ces contraintes de la vie des sociétés, dont la prise en charge adéquate exigera sans doute une réforme profonde de la formation et de la culture des acteurs et producteurs des sciences de l'homme.

Dans le concert tantôt harmonieux, tantôt discordant des diverses sciences que deux millénaires et demi de l'histoire de nos sociétés ont porté au jour, un son nouveau, si discret que beaucoup encore ne l'ont pas entendu, s'est élevé tout au long des dernières décennies du XX$^{e}$ siècle : le son d'une *scienza nova* qui s'est constituée en se donnant pour objet d'étude un ensemble longtemps ignoré mais essentiel de conditions et de contraintes de la vie des sociétés, celles de la diffusion (et de la rétention) des connaissances et des savoirs, ou, pour user d'un néologisme englobant, à la fois plus large et plus précis, des *praxéologies*. On appelle ici *didactique* cette science nouvelle : son territoire, c'est le continent des didactiques disciplinaires – didactique des mathématiques, didactique de l'EPS, etc. – qui ont émergé jusqu'ici ou qui naîtront demain. Une question de vocabulaire doit en ce point être tranchée : *la* didactique, de ce point de vue, n'existe pas moins que *la* physique, *la* biologie, etc. Et un didacticien des mathématiques peut dire à bon droit, en certains contextes, qu'il est un *didacticien*, tout court, de même que d'autres disent « Je suis physicien » sans préciser nécessairement s'ils sont physicien du solide, ou des particules, etc. Bien entendu, la même personne se présentera aussi comme didacticien *des mathématiques* – comme d'autres diraient qu'ils sont physiciens des particules – pour préciser que son camp de base et son point de visée, à l'intérieur du continent didactique, ce sont les conditions et les contraintes de la diffusion sociale des praxéologies *mathématiques*.

Le développement du continent didactique appelle des *études comparées* entre didactiques disciplinaires comme entre domaines d'une *même* didactique disciplinaire. De tels travaux ont d'abord pour objet de nourrir le développement *des didactiques disciplinaires* et de leurs différents domaines, et de fonder ainsi le développement de *la* didactique, consubstantiel au développement *des* didactiques. Mais pourquoi les études comparées en didactique sont-elles nécessaires au développement des didactiques disciplinaires ? Pour répondre, nous reprenons ici un schéma qui oriente, à nos yeux, toute étude en didactique et singulièrement toute étude « comparée » : l'échelle des *niveaux de détermination didactique* (figure 1).

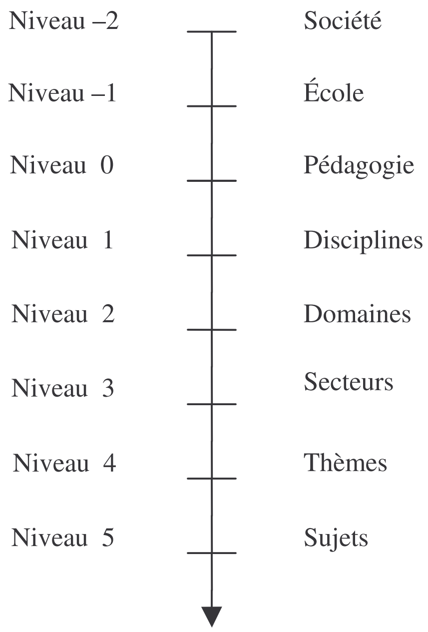


*Figure 1*

Le principe de la réponse à la question posée tient en fort peu de mots : si le didacticien opère bien aux niveaux de plus grande spécificité de l'échelle des niveaux de détermination didactique, ce qu'il étudie est conditionné, de manière directe ou indirecte, par des contraintes relevant de *l'ensemble* de ces niveaux. Bien entendu, les contraintes en question opèrent en général d'une manière *spécifique* sur les niveaux de plus grande spécificité. Mais encore faut-il identifier ces contraintes, notamment celles qui ont pour siège la société, l'École ou la « pédagogie ». Encore faut-il dégager les notions génériques de contrainte disciplinaire, de contrainte de domaine, de contrainte sectorielle, etc., dans une dialectique toujours ouverte entre généricité et spécificité. Pour cela, l'effort comparatiste, à peine engagé aujourd'hui, apparaît comme une condition clé du développement du continent didactique.

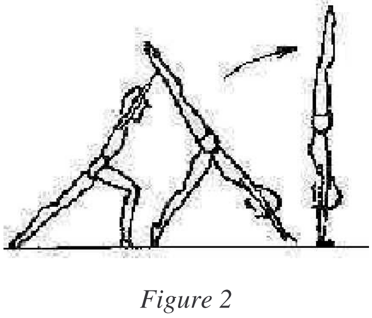

*Figure 2*

Notons encore cette conséquence : dès lors qu'on envisage le territoire du didacticien comme le donne à voir l'échelle des niveaux de détermination didactique, la question des connaissances à mobiliser dans une recherche en didactique se pose autrement qu'on ne le suggère parfois. Imaginons la situation suivante : dans cette discipline scolaire qu'est l'*éducation physique et sportive* (EPS), dans ce *domaine* d'études qu'est la *gymnastique*, considérons cette praxéologie dont les programmes scolaires font un *sujet* d'étude classique : l'*appui tendu renversé* (figure 2). Que faut-il savoir de cette *praxéologie gymnique* pour mener à bien telle ou telle étude didactique qui prenne pour objet sa diffusion scolaire ? Sur l'ATR lui-même, il faut sans doute, dans la plupart des cas, en savoir plus que ce que le schéma précédent en fait connaître ! Mais est-il nécessaire, par exemple, de savoir là-dessus « tout » ce que sait qui a pratiqué la gymnastique à haut niveau ? Rien ne permet de l'affirmer. Solidairement, mais en sens inverse, rien ne permet non plus d'affirmer que les connaissances ainsi évoquées, de manière forcément floue, seraient dans tous les cas *suffisantes*. Car le rapport idoine à tel ou tel objet qui structure sa recherche est, pour le didacticien, un rapport *sui generis* qu'il lui faut *définir et redéfinir* parce qu'il ne saurait qu'exceptionnellement s'identifier à tel ou tel rapport institutionnel déjà existant dans la société. C'est même là le principal ressort de l'émergence du continent didactique depuis trois ou quatre décennies ! Car c'est en particulier parce qu'on ne fait pas de la didactique de la statistique (par exemple) ou de la didactique de la gymnastique avec les connaissances actuelles d'un professeur de mathématiques ou avec celles d'un professeur d'EPS, et que l'on n'y réussit pas davantage avec les connaissances qui permettent à la recherche « non didactique » en statistique ou en gymnastique de progresser, que la figure du didacticien s'est mise à exister dans une noosphère déjà fort peuplée.

Dans la recherche « non didactique » en une discipline donnée, il y a bien entendu du didactique, parce qu'il y a toujours une *intention didactique*. Mais celle-ci vise en principe un public de pairs ou de quasi-pairs, configuration qui pose au chercheur « non didacticien » des problèmes didactiques regardés en général – peut-être trop vite – comme fort éloignés de ceux que s'efforce de résoudre, ordinairement, le didacticien qui se réfère à l'institution scolaire. C'est dans cette perspective, soulignons-le en passant, qu'il faudrait situer l'affirmation – refusée aujourd'hui encore en nombre de disciplines, qui n'arrivent pas même à la *penser*, faute déjà de s'être dépris de l'*habitus* sécessionniste engendré par le grand partage historique des sciences – que la didactique d'une discipline donnée doit être regardée comme une composante de cette discipline même. Relève en effet, ou *devrait* relever, selon nous, d'une discipline donnée toute sous-discipline qui accroît de manière spécifique notre connaissance de certains des objets de cette discipline regardée comme *réalité sociale large* – et non comme réalité sociale limitée à un « petit monde » choisi, monde savant ou monde du haut niveau par exemple.

## 2 *Un problème de transposition didactique*

On examine dans ce qui suit le problème de l'étude scolaire d'un domaine déterminé de l'actuel curriculum mathématique : *la statistique*. On se limitera, à cet égard, à l'exploration de quelques-unes des conditions et contraintes, de niveau de spécificité variable, qui déterminent actuellement le champ des possibles en matière d'enseignement de la statistique, afin de réunir des matériaux susceptibles d'entrer dans l'élaboration d'une réponse à la question suivante, qu'il faut entendre comme une question *de la profession* (et non de tel ou tel professeur) : *comment développer un enseignement scolaire de la statistique qui apparaisse à la fois fidèle à la science statistique telle qu'elle existe hors de l'École et pertinente pour la formation scolaire des jeunes générations ?* Sans faire ici de longs développements sur la transposition didactique, rappelons toutefois que la « fidélité » mentionnée ne saurait être entendue comme un pur et simple mimétisme : car les transpositions extrascolaires d'un savoir donné développent fréquemment des idiosyncrasies institutionnelles qui, reprises sans esprit critique, par simple souci d'authenticité formelle, pourront en certaines façons se révéler pathogènes dans l'institution de formation où, sous couleur de transposition, on a cru bon de les reconstituer scrupuleusement.

La question posée vaudrait pour d'autres domaines, secteurs ou thèmes du curriculum mathématique secondaire. Pourquoi alors interroger l'enseignement de la statistique ? Prenons acte ici d'une raison qui tient à l'objet de la recherche plutôt qu'à la dynamique interne à la recherche en didactique des mathématiques : un nouveau programme de mathématiques est entré en vigueur dans les classes de seconde françaises à la rentrée 2001, qui marque sa volonté de renouveler l'enseignement de la statistique [1]. Que peut-on espérer de cette volonté d'*aggiornamento* ? Et que peut-on faire, dans cette perspective, pour atteindre les deux objectifs conjoints explicités dans la question formulée ci-dessus ?

### *3 La statistique ou les statistiques ?*

Notons d'abord un paradoxe. Alors que le programme en vigueur se veut l'opérateur d'un renouvellement de l'enseignement de la statistique, nulle définition de la statistique n'y est proposée ! Il est vrai que, traditionnellement, l'identification du savoir à enseigner comme *totalité culturelle* est supposée aller de soi. Le paradoxe n'en est donc pas un : comme il est d'usage, le programme se contente de détailler les *parties* de « la statistique » à étudier dans la classe particulière à laquelle ce programme se rapporte, en ne définissant pas plus « la statistique » qu'il ne définit « la géométrie ». Traiter ainsi le savoir à enseigner comme une *évidence de la culture* participe, en vérité, d'une technologie curriculaire toute classique, peu informée des connaissances accumulées en matière de transposition didactique au cours des deux dernières décennies. La non-problématicité supposée du savoir à enseigner constitue, on le sait, une fiction cardinale des transpositions didactiques « à l'ancienne » : les élèves seuls feraient problème ! Cette illusion de transparence constitue, ici comme ailleurs, une condition déterminante de la transposition didactique que les concepteurs du programme ont enclenchée. En particulier, cette condition laisse libre cours, au plan du lexique, à un détournement quasi instantané : de même que nombre de professeurs rebaptisent spontanément « analyse » le domaine d'études que le nouveau programme nomme humblement « calcul et fonctions », de même la corporation des professeurs de mathématiques retouche le projet d'enseignement de la statistique porté par le nouveau programme en en modifiant la dénomination : le programme leur impose d'enseigner *la* statistique, la plupart des professeurs parlent d'enseigner *les* statistiques – « les stats ».

Confortant en cela une réduction de la compétence professorale qui, si traditionnelle soit-elle, ne saurait être regardée comme un indépassable destin, le programme dialogue avec les professeurs en s'en tenant, dans l'échelle de détermination didactique, aux seuls niveaux des *thèmes* et des *sujets*, en s'élevant le moins possible au niveau des *domaines* qui compose le programme – « statistique », « calcul et fonctions », « géométrie ». Or cet évitement est, pour le professeur, synonyme d'enfermement et fonctionne comme un *interdit de penser* touchant les secteurs et domaines d'études du programme. Une telle censure, en particulier, ne permet guère de faire entendre une précision du type de celle qu'apportent par exemple les auteurs d'une *Histoire de la statistique* au début de leur ouvrage : « … nous n'avons pas voulu, notent-ils, aborder l'*histoire des statistiques*, publiques, officielles ou privées [2]. » Ici, point de déclaration précisant qu'on demande au professeur de mathématiques de se faire enseignant, non de statistiques, au pluriel, mais de *statistique*, au singulier : le professionnel attentif devra se contenter de noter par lui-même que, si « statistiques » (au pluriel), en lieu et place de « statistique », se rencontre une fois dans le programme, par un *lapsus calami* sans doute significatif de la pression du jargon professoral [3], et une fois encore dans le document d'accompagnement du programme [4], « statistique » (au singulier) est le mot consacré par

---

[1] BOEN hors série n° 2 du 30 août 2001 : http://www.education.gouv.fr/bo/2001/hs2/default.htm.

[2] Droesbeke et Tassi 1990, p. 4.

[3] On le trouve dans l'affirmation suivante : « À titre indicatif, le temps à consacrer aux différents chapitres pourrait être de 1/8 pour les statistiques, le reste se répartissant équitablement entre les deux autres chapitres. »

[4] « Les calculatrices sont par ailleurs un premier outil de simulation simple pour la partie "statistiques" du programme. »

lequel les rédacteurs nomment aussi bien ce qui s'est enseigné au collège [5] que ce qui s'enseignera dans les classes de première et de terminale [6]. Chose plus remarquable, le singulier résiste même dans l'expression « cahier de statistique », alors que ce cahier, nouvellement suggéré par le programme, est par excellence le registre où viendront s'inscrire *des* statistiques, sans lesquelles il ne serait guère possible d'étudier *la* statistique [7].

Les puissances tutélaires semblent donc, sur ce point – *la* statistique ou *les* statistiques ? –, dénuées d'ambiguïté. Lors de la présentation à l'Académie des sciences, le 3 juillet 2000, du rapport sur la statistique élaboré par un groupe de travail animé par Paul Malliavin, Jean-Pierre Kahane, qui à l'époque préside la Commission de réflexion sur l'enseignement des mathématiques (CREM), note pour s'en réjouir : « Le rapport distingue bien *les* statistiques, issues d'une grande variété d'activités, et *la* statistique, qui conceptualise et développe les méthodes de recueil et de traitement *des* statistiques [8]. » L'introduction du rapport de l'Académie s'intitule en effet *Des statistiques à la statistique*. « Les statistiques, y lit-on, sont des dénombrements de sujets, d'objets, d'événements dans une population ou des sous-populations. La statistique est une démarche permettant de recueillir, de traiter et d'interpréter les données qu'on recueille dans divers domaines où celles-ci présentent une caractéristique essentielle : la variabilité. » Le rédacteur note : « Les statistiques existent depuis des siècles, la statistique par contre repose sur un mode de pensée original qui ne s'est réellement développé qu'à partir du XIX^e^ siècle. »

Par contraste avec ce point de vue – l'intention didactique officielle est bien qu'il y ait en seconde un enseignement de *la* statistique –, notons d'emblée un effet pervers de la substitution subreptice, non questionnée, de « statistiques » à « statistique » dans la langue des professeurs : si l'on prétend enseigner *les* statistiques, il devient plus difficile de poser nettement le problème – crucial – de la place, de la nature et du rôle *des* statistiques dans l'enseignement de *la* statistique, tout particulièrement dans le cadre du *cours de mathématiques*. Le jeu de langage – statistiques, statistique –, que d'aucuns jugeront d'abord anodin, ressemble fort, en ce cas, à un tour de passe-passe : étiqueter « enseignement des statistiques » un enseignement dont les principaux objets seront en fin de compte la moyenne, l'écart type, etc., conduit assez facilement à oublier la question vitale *des* statistiques.

## 2. LE GENRE PROCHAIN : DES MATHEMATIQUES METISSES

### *1 La statistique, dites-vous ?*

Qu'est-ce donc que cette statistique qu'il s'agirait d'enseigner ? L'étude de cette question commande en grande partie ce que l'on pourra répondre à la question mise au principe de cette étude. De cette question, les programmes postulent tacitement que les professeurs n'ont pas à disputer parce qu'ils connaîtraient déjà la réponse, réduite classiquement à une énumération de secteurs d'études, de thèmes et de sujets en quoi le domaine statistique s'analyserait apparemment sans reste. Question cruciale, qui renvoie à un *type* de tâches – qu'est-ce que la géométrie ? qu'est-ce que l'algèbre ? qu'est-ce l'analyse ? – dont les textes officiels font plus généralement l'économie : contrainte qu'il faut situer au niveau de l'École

---

[5] La classe de troisième, écrivent-ils ainsi, est le lieu d'une « initiation à l'utilisation des tableurs-grapheurs en statistique ».

[6] « En classe de première et de terminale », peut-on lire, « l'enseignement de la statistique sera présent dans toutes les filières mais sous des formes diverses. »

[7] « L'élève pourra se faire un "cahier de statistique" où il consignera une grande partie des traitements de données et des expériences de simulation qu'il fait… »

[8] Malliavin 2000, p. 180.

(le niveau –1 dans l'échelle reproduite plus haut), et dont nous essaierons d'abord de desserrer un peu l'étreinte.

Dans son intervention devant l'Académie des sciences, Jean-Pierre Kahane rappelle l'importance, en France, de la « statistique officielle » productrice de statistiques (au pluriel), et qui, à côté des probabilités, constitue le principal atout du pays en matière de statistique puisqu'elle lui permet d'occuper « le second rang international, derrière les États-Unis naturellement, sous l'angle de la représentation à l'ISI (*International Statistical Institute*) [9] ». Ce type d'analyse conduit à lier fortement *la* statistique (y compris dans ses aspects mathématiques) et *les* statistiques, en prenant ainsi le « continent statistique » *comme un bloc*. Ce continent, il est vrai, est loin d'être unifié : s'y côtoient un premier sous-bloc, lui-même divisé, comprenant la statistique mathématique et le calcul des probabilités pratiqués par des mathématiciens (qui travaillent pour la plupart au sein de l'Université) ainsi que des disciplines ayant pignon sur rue (médecine, psychologie, biologie, sociologie, etc.) qui exploitent leurs propres techniques statistiques, et un second sous-bloc formé d'une part d'entreprises commerciales productrices de sondages, d'autres part de services publics ou privés produisant, au moyen d'enquêtes ou à partir de documents administratifs, ces informations économiques et sociales qui sont l'un des points forts de la France dans la compétition internationale [10]. Sans remonter même à l'*arithmétique politique* de William Petty (1623-1687) et John Graunt (1620-1674), notons que l'arrimage du continent statistique aux mathématiques est ancien : dès les années 1830-1840, nous rappelle Alain Desrosières [11], Adolphe Quetelet (1796-1874) noue, par le moyen de la distribution gaussienne, discours probabiliste et observations statistiques, pensant ensemble « l'aspect aléatoire et imprévisible des comportements individuels » et « la régularité et donc la prévisibilité de la sommation statistique de ces actes individuels, à travers la notion d'*homme moyen* ». Mais le premier sous-bloc se mettra véritablement en place au début du XX$^{e}$ siècle, « quand sont routinisées et diffusées les techniques de la régression et de la corrélation, à partir du centre de biométrie de Karl Pearson, puis celles de la statistique inférentielle (estimation, tests, analyse de variance) développées au laboratoire expérimental d'agriculture de Ronald Fisher [12] ».

## 2 *Le continent statistique à l'épreuve de l'enseignement*

Fruit d'une histoire plurielle, deux régimes de mathématisation de la statistique coexistent alors : l'un, élémentaire, correspond en gros à ce que nous nommons la statistique *descriptive*, l'autre, qui mobilise des mathématiques supérieures, articule les conquêtes de la jeune statistique *inférentielle*. Lorsque se crée en 1922 l'Institut de statistique de l'Université de Paris (ISUP), le premier niveau y fait l'objet d'un cours intitulé *La méthode statistique*, qui comporte 25 leçons. Douze leçons sont consacrées à l'enseignement de second niveau : intitulé *Éléments de statistique mathématique*, il est longtemps assuré soit par Georges Darmois (1888-1960), soit, plus rarement, par Émile Borel (1871-1956). Pour l'année 1925-1926, par exemple, le contenu de ces leçons est le suivant [13] :

> 1) Statistique des caractères. Association. 2) Statistique des variables. Courbes de fréquence.

---

[9] Malliavin 2000, pp. 179-180.

[10] Voir là-dessus Volle 1984, p. 11.

[11] Desrosières 1993, p. 18. Voir aussi Desrosières 2002.

[12] Desrosières 1993, pp. 21-22. Dans l'introduction de son livre *Statistical methods for research workers* (1925), Ronald Fisher (1890-1962) écrit en effet : *« The science of statistics is essentially a branch of Applied Mathematics and may be regarded as mathematics applied to observational data. As in other mathematical studies the same formula is equally relevant to widely different groups of subject matter. Consequently the unity of the different applications has usually been overlooked, the more naturally because the development of the underlying mathematical theory has been much neglected. »*

[13] D'après Pressat 1987, p. 25.

3) Moyennes. Écarts. Corrélation. 4) Applications. 5) Corrélation multiple. 6) Applications. 7) Stabilité des fréquences. Probabilité. Principes fondamentaux. 8) Épreuves répétées. Théorème de Bernoulli. Loi de Laplace. 9) Applications. 10) Polygones dissymétriques. Loi des petits nombres. 11) Ajustement des statistiques. Dispersion. Schéma des urnes. 12) Écart des observations.

Mais l'enseignement de « statistique » dispensé à l'ISUP ne se réduit pas à cela, comme le montre le tableau des cours de l'année 1939-1940 : autour de la méthode statistique et des éléments de statistique mathématique (qui occupent désormais 20 leçons), se déploient des enseignements portant sur la démographie et la statistique sanitaire (20 leçons), les assurances sur la vie (20 leçons), les opérations financières (25 leçons), l'économie politique mathématique (20 leçons), la science des affaires (16 leçons), la législation, l'hygiène et l'assistance sociales (12 leçons)[14]. Le poids de ces enseignements « périphériques » s'explique sans doute en grande partie par la volonté des institutions parties prenantes dans l'organisation de cet enseignement « interuniversitaire » d'être représentées solidement dans la formation donnée ; il n'en reste pas moins que, à la tension mathématique déjà notée entre un niveau élémentaire et un niveau supérieur se surimpose une tension entre le cœur de la discipline et ses périphéries.

Le mouvement de réduction du continent statistique à son « corps central tel qu'il s'est constitué de nos jours » – ce que les auteurs de l'époque nomment souvent *la* méthode statistique, au singulier – est alors amorcé ; mais il serait erroné de le croire accompli. Dans un ouvrage intitulé *Statistique et applications*, dont la première édition est de 1934 et la cinquième de 1957, Darmois précise que la méthode statistique, qui « développe ses applications dans un champ très étendu », comporte essentiellement trois centres d'activité[15] : la présentation des observations ; leur réduction ; la description, l'interprétation et l'explication des régularités statistiques. Le premier pôle d'activité est à l'origine du nom même de statistique, qui dérive « de *status*, pris soit au sens d'État, soit à celui de situation » ; les deux autres centres névralgiques de l'activité statistique, réduction et interprétation, appartiennent davantage à ce qui deviendra le « corps central » de la statistique. Mais l'auteur ne se limite pas à une statistique réduite à une technologie elle-même réduite à ses composants mathématiques. Ainsi, pour illustrer la notion de régularité statistique, Darmois se réfère-t-il successivement aux jeux de hasard, au taux de masculinité, aux lois mendéliennes de l'hybridation, à la radioactivité, aux taux des mariages, de natalité, de mortalité. Dans le corps de l'ouvrage, trois chapitres seront successivement dévolus à l'analyse démographique, aux « indices de l'activité économique », aux « permanences de l'hybridation » (lois de Mendel). Inversement, ces chapitres d'application sont encadrés par des chapitres de technologie statistique. On retrouve au reste à l'échelle d'un chapitre ce que la table des matières montre à l'échelle du livre : l'encadrement de la technologie statistique par des emplois extramathématiques de cette technologie, et réciproquement. Le chapitre II, intitulé significativement « L'outillage et les idées », comporte ainsi les subdivisions suivantes :

Dénombrements et mesures – Diagramme intégral – Courbe de fréquence – Moyenne arithmétique – Écart moyen quadratique ou écart type – Valeur médiane – Quartiles ou quartiers – Écart moyen – Introduction à la théorie des probabilités – Notion de variable aléatoire – Espérance mathématique – Signification de l'espérance mathématique – Le résultat d'A. de Moivre – Nature des interprétations et explications fournies par la théorie des probabilités – Taux de masculinité – Le cas le plus simple des lois de Mendel – Radioactivité.

Cette structure se retrouve dans le chapitre VI, intitulé « Répartitions statistiques à une variable », qui présente le découpage suivant :

Dimensions d'organismes – Temps de réaction – Fréquences de désintégration des atomes radioactifs – Distribution de revenus – Répartition des villes d'après le nombre d'habitants –

[14] *Op. cit.*, p. 23.
[15] Darmois 1957, p. 3.

Première utilité de ces représentations – Peut-on espérer d'autres résultats – Spécification préalable de la loi de fréquence – Exemples des tailles – Estimation des paramètres – Qualité d'une représentation – Stabilité d'une courbe de fréquence – Autres formes de distributions – Répartitions discontinues – Le problème général du jugement sur échantillon – Médiane et déciles – Emploi d'autres représentations.

Le caractère métissé des contenus est frappant. Si on le regarde à son tour comme un sous-continent des mathématiques, le continent statistique relève sans doute aucun des mathématiques *mixtes* : la statistique, *toute* statistique mêle nécessairement des objets mathématiques et des objets non mathématiques. La mise en œuvre de la « méthode statistique », c'est-à-dire de la technologie statistique, enclenche ainsi, dans le meilleur des cas, de véritables *synergies codisciplinaires*, en articulant les énergies de deux disciplines au moins, l'une mathématique, l'autre non. Sur ce patron, on pourra faire de la statistique *en* médecine, ou *en* démographie, ou *en* lexicologie, ou *en* psychologie, ou *en* didactique, etc. D'une manière générale, la statistique apparaît comme un complexe plus ou moins intégré d'organisations mathématiques mixtes, parmi lesquelles on peut sans doute distinguer – jusqu'à un certain point – une statistique « médicale », une statistique « démographique », une statistique « lexicologique », une statistique « psychologique », une statistique « docimologique », et ainsi de suite [16].

### 3 *Une statistique doublement amputée*

Voué *a priori* à intégrer toutes les statistiques possibles, le continent statistique est pourtant fréquemment soumis à deux opérations de *réduction* qui, l'une et l'autre, peuvent dans certaines conditions adultérer la science statistique, même quand on se plaît à en évoquer le caractère pluriel. D'un côté, la « discipline d'accueil » – qui gouverne spécifiquement les études dans le champ des phénomènes où l'on doit mettre en œuvre la technologie statistique – peut tendre à masquer, à refouler la discipline mathématique à laquelle elle devrait normalement se soumettre. Cette neutralisation est fréquemment recherchée pour des motifs didactiques, parce qu'elle permet d'abaisser le coût de la percolation des praxéologies statistiques dans les institutions consommatrices. C'est ainsi que, dans les premières lignes d'un petit livre sous-titré *La statistique et le vivant*, un auteur déjà mentionné [17] indique sans détour que, « bien que s'appliquant à des données chiffrées », le « mode de pensée statistique » – sinon la statistique – « peut être exposé sans faire appel aux mathématiques », et même sans aucune formule… Plus explicites encore, les auteurs d'une *Statistique en psychologie* précisent que leur ouvrage n'a pas pour objectif « d'étudier les fondements mathématiques de l'analyse statistique, mais de *comprendre les principes qui permettent de réaliser correctement ces analyses*, quelle que soit la formation initiale, scientifique ou littéraire, du lecteur [18] ». La connaissance de ces fondements mathématiques, ajoutent-ils, « serait évidemment préférable pour acquérir une bonne compréhension des fondements de la statistique et de ses champs d'application. Mais c'est un travail considérable, voire trop difficile pour des non-mathématiciens, et qui constitue une spécialité à part entière ». Les mathématiques n'apparaissent plus, dès lors, que comme des *ingrédients techniques*, et non comme des composants *technologiques* permettant la *production*, la *justification* et l'*intelligence* de techniques qui tendent alors à devenir de simples recettes. Il est vrai que cette

[16] C'est en ce sens qu'on pourrait à bon droit parler au pluriel *des* statistiques, en désignant par là des organisations de savoir en partie distinctes mais aucunement disjointes. Des auteurs réputés (Schwartz et Lazar 1978) ne donnent-ils pas à l'un de leurs ouvrages le titre d'*Éléments de statistique médicale et biologique* ? À propos d'une variante terminologique usitée plus haut (« statistique en démographie », etc.), notons encore qu'une collection publiée chez Flammarion s'intitule « Statistique en biologie et en médecine ».
[17] Schwartz 1994, p. XI.
[18] Rude et Retel 2000, p. 15.

démathématisation peut être en partie symbolique, et destinée surtout à marquer que la discipline « cliente » *reste maîtresse dans son domaine* [19].

La mise en retrait des technologies statistiques *à forte teneur mathématique* n'est cependant pas l'apanage des consommateurs de statistique pour qui celle-ci est un instrument dont ils cherchent simplement à faire une utilisation appropriée. Semblable évanescence technologique existe aussi, à bien des égards, dans le « cours de mathématiques » même ! C'est ainsi que le programme de la classe de seconde, traitant de la question (facultative) des sondages, préconise d'inciter les élèves « à connaître l'approximation usuelle de la fourchette au niveau de confiance 0,95, issue d'un sondage sur $n$ individus ($n > 30$) dans le cas où la proportion observée $\hat{p}$ est comprise entre 0,3 et 0,7, à savoir : $[\hat{p} - 1/\sqrt{n}\ ;\ \hat{p} + 1/\sqrt{n}]$ ». On est ici, typiquement, dans une situation analogue à celle de l'étudiant en psychologie pour la sérénité mathématique duquel les praxéologies statistiques à étudier ont été purgées des mathématiques vivantes ayant servi à la production des techniques qui y sont enchâssées, techniques qui ne laissent plus guère apparaître, dès lors, que des fragments mathématiques désormais figés, en quelque sorte cristallisés dans les praxéologies proposées [20]. Même si l'on ne peut généraliser, on tient là un exemple parmi bien d'autres possibles du fait que la démathématisation de la statistique n'est pas l'apanage des « non-mathématiciens ».

D'un autre côté, l'enseignement « généraliste » de la statistique en mathématiques tend à faire de l'extramathématique un théâtre d'ombres, qui met en scène une réalité d'opérette avec laquelle on entretient un commerce incertain, opportuniste, le cours de mathématiques marquant souvent à son endroit, de façon cavalière, une attitude distante et versatile. On retrouve ici, bien entendu, tout le problème de la *motivation* des organisations mathématiques étudiées et de leur signification sociale – c'est-à-dire de leur reconnaissance comme expression de certaines conditions et contraintes de la vie des sociétés. Mais le cas examiné, celui de la statistique, est quelque peu autre et pourrait être autrement révélateur que les cas, de longue date naturalisés, de la géométrie ou de l'arithmétique. Pour ces domaines d'intervention de la raison mathématique, en effet, il existe chaque fois une réalité extramathématique relativement hypostasiée dans la culture courante, la « spatialité » pour la géométrie, la « numérosité » pour l'arithmétique, qui constitue en même temps un foyer de questions génératrices des organisations mathématiques à diffuser et le terrain empirique où mettre à l'épreuve, par exemple par l'expérimentation, ces organisations mathématiques *in statu nascendi*. Dans le cas de la statistique, l'équivalent de la spatialité ou de la numérosité n'est rien d'autre que la *variabilité*, réalité du monde social et naturel dont l'arraisonnement scientifique est historiquement récent et dont nous verrons plus loin que la reconnaissance et l'assomption par la culture commune (scolaire et extrascolaire) sont aujourd'hui encore problématiques. En statistique plus encore qu'en géométrie ou en arithmétique, il convient donc de prendre au sérieux les singularités du monde extramathématique dont la statistique permet d'étudier la variabilité. Or c'est en ce point que l'on rencontre un formidable obstacle : celui de *l'interdit de connaissance* qui tend à s'imposer à la plupart d'entre nous à propos de la plupart des faits sociaux ou naturels, et dont nous dépeindrons d'abord les effets.

---

[19] C'est ainsi que l'ouvrage déjà cité de Ronald Fisher – *Statistical methods for research workers* – est présenté sur tel site Internet (http://psy.ed.asu.edu/~classics/index.htm) comme un classique de… la psychologie !

[20] La fourchette suggérée par le programme de seconde est une *simplification* de la fourchette que proposent les auteurs de la « Statistique en psychologie » déjà citée, à savoir $[p - 1{,}96\sqrt{pq/n}\ ;\ p + 1{,}96\sqrt{pq/n}]$, où $q = 1 - p$ (Rude et Retel 2000, p. 109). La fonction $p \mapsto p(1 - p) = pq$ ayant un maximum égal à 1/4 pour $p = 1/2$, on obtient en effet le premier encadrement à partir du second, puisqu'on a : $1{,}96\sqrt{pq/n} \leq 1{,}96/(2\sqrt{n}) < 1/\sqrt{n}$. On peut penser au reste qu'un tel travail de « stylisation » mathématique aurait toute sa place en classe de seconde.

*4 Les renoncements de la statistique calculatoire*

Il n'est pas rare que les auteurs d'ouvrages de statistique imaginent un dialogue entre deux personnages, dont l'un est le « statisticien » et l'autre son « client », supposé légitime dans une sphère d'activité qui, dans la réduction mathématique de la statistique, est généralement laissée dans l'ombre [21]. La dichotomie entre statisticien et client est sans doute inscrite en nombre d'institutions du continent statistique mais elle ne cesse d'être dénoncée par les statisticiens eux-mêmes comme une source de difficultés souvent irrémédiables [22]. Ici s'applique l'observation faite plus haut selon laquelle, dans le travail transpositif, il n'est pas *a priori* illégitime de prendre ses distances vis-à-vis de certaines configurations existant dans l'univers extrascolaire que l'on prétend transposer. En matière de statistique, ainsi, il n'est pas nécessaire que l'élève s'identifie – fût-ce à son insu ou à l'insu du professeur – à la figure au demeurant incertaine du « statisticien ».

Pour poser le problème plus complètement, nous userons d'un schéma formel simple. L'élève $x$ de l'enseignement scolaire étudie les mathématiques, la physique, la biologie, etc., non en s'identifiant à un mathématicien, à un physicien, à un biologiste, etc., mais en apprenant, dans la position *clé* d'élève de l'École de la République et de *citoyen* en devenir, à apporter à des questions $Q_i$ des réponses $R_i$ (qui sont des praxéologies ou des fragments de praxéologies), et cela en *usant* de ces disciplines de production praxéologique que sont les mathématiques, la physique, la biologie, etc. Quand $x$ étudie en classe la question $Q$, dont on suppose qu'elle ne tombe pas sous une juridiction disciplinaire déterminée (du moins aux yeux de $x$), il doit apprendre à mobiliser telle ou telle discipline de production praxéologique – mathématiques, physique, biologie, etc. –, et tout d'abord à en reconnaître la pertinence dans l'abord de $Q$, en même temps qu'il participera, au sein de ce collectif qu'est la classe, à la mise en place et à la mise en œuvre solidaires de praxéologies disciplinaires spécifiques, utiles pour élaborer certains des matériaux qui permettront éventuellement de construire une réponse $R$ à $Q$. En cela, l'élève $x$ s'identifie, non au spécialiste de la discipline mobilisée, mais au citoyen qui la mobilise en vue de répondre, avec d'autres, à une question $Q$, laquelle ne relève pas nécessairement de cette discipline même, mais appartient d'abord à l'ensemble des problèmes que, comme le disait jadis l'historien Lucien Febvre (1878-1956), « l'homme non spécialisé porte en lui ».

Si l'on revient alors au schéma dichotomique du statisticien et de son « client », l'identification de l'élève doit à coup sûr *se faire avec le « client »*. Prenons ici l'exemple d'un ouvrage intitulé dans sa version originale *Principles of statistics* (1971), traduit et publié en français sous le titre *Principes de statistiques* [23], et qui s'adresse, non à des élèves du secondaire, mais à des étudiants en sciences sociales. En ce cas, donc, $x$ n'est pas simplement le futur citoyen : il est un citoyen futur psychologue, ou ethnologue, etc., en sorte que le rapprochement que nous opérons n'est qu'indicatif. L'ouvrage, précise son auteur, « est destiné aux étudiants de sciences sociales qui doivent lire les publications d'un œil critique et analyser leurs propres données expérimentales ». Interroger le fruit de l'activité d'autrui, interroger la matière de sa propre activité : si l'on remplace « étudiants de sciences sociales » par « futurs citoyens », on obtient quelque chose d'assez proche de la position de l'élève évoquée plus haut. Mais qu'en est-il alors des mathématiques dans ce projet de formation ?

---

[21] Cet interlocuteur du statisticien est appelé *l'utilisateur* dans Reeb et Fuchs 1967, *l'expérimentateur* dans McGee 1975, etc.

[22] Voir ainsi Volle 1984, *passim*. On prête à Fisher la remarque, ironique et désabusée, selon laquelle *« to call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of »*…

[23] McGee 1975. L'*s* final de *statistics* (comme celui de *mathematics*) n'est pas, en anglais, une marque du pluriel. C'est ainsi que l'ouvrage *Seeing through Statistics* de Jessica M. Utts (1999) s'ouvre par ces mots : « Statistics deal*s* with complex situations involving uncertainty. »

L'auteur répond ainsi : « Si vous vous mettez à la place de l'expérimentateur (et c'est ce que l'auteur suppose), vous n'aurez à faire que des additions, des soustractions, multiplications et divisions ainsi qu'à élever certains nombres à des puissances données, extraire des racines carrées ou bien attribuer des valeurs aux variables de certaines équations pour calculer différents nombres statistiques (nombres résumés tels que la moyenne, la variance, le coefficient de corrélation) [24]. » Telle est la formule de ce que nous nommerons la *réduction arithmétique* de la statistique, qui fait de celle-ci un calcul arithmétique simplement plus sophistiqué que le calcul de l'école primaire.

L'ouvrage cité n'en reste cependant pas à cette *statistique calculatoire*. Supposé tenir fermement sa position d'expérimentateur, son lecteur est invité à venir occuper régulièrement la position du statisticien, car l'ouvrage vise à le « préparer à discuter intelligemment avec un statisticien », ce qui, l'avertit-on [25], ne manquera pas de lui imposer un plus riche commerce avec les mathématiques de la statistique, l'auteur mentionnant même, à cet égard, « deux livres qui font autorité » – les *Mathematical Methods of Statistics* de Harald Crámer (1946) et le volume 1 de *The Advanced Theory of Statistics* de M.G. Kendall et A. Stuart (1958). Mais ce n'est pas seulement cette ouverture, discrète et insistante à la fois [26], vers les mathématiques supérieures qui épargne à l'ouvrage examiné de succomber à la réduction arithmétique de la statistique et lui permet, positivement, de s'inscrire à l'intérieur du continent statistique, loin du liséré côtier où la statistique calculatoire s'étiole.

Le processus d'étude qu'inspire et outille la statistique est illustrée ici à propos de la question suivante[27] : « Quelle est la taille moyenne des étudiants hommes admis dans les universités américaines en 1968-69 ? » La question conduit d'abord à s'interroger sur la *population* concernée : appartenir à cette population suppose que l'on soit de sexe masculin, que l'on soit inscrit en première année d'université à la rentrée 1968, et cela… dans une université américaine. Même si le sujet n'est que brièvement traité, l'auteur évoque les difficultés inattendues qui peuvent s'élever : que faire d'un étudiant cul-de-jatte par exemple ? Puis il brosse rapidement la technique statistique de base [28] : ne pas essayer de considérer toute la population, mais choisir un *échantillon* de cette population ; mesurer la taille des individus de cet échantillon ; répondre à la question pour ce qui est de l'échantillon examiné ; s'interroger sur ce que, à partir de là, il est possible d'inférer à propos de la population tout entière. Même si l'auteur ne mentionne pas le long débat historique entre partisans des études exhaustives et partisans des sondages [29], il souligne que l'*échantillonnage* pose à l'expérimentateur des problèmes non moins redoutables que la définition de la population, ce qui conduit souvent à utiliser tel échantillon pour la simple raison qu'il est… *disponible*, l'auteur soulignant à cet égard le problème posé par ces universités américaines qui sélectionnent leurs étudiants avec l'ambition première d'avoir de bonnes équipes de basket-ball, ce qui risque de faire d'un échantillon disponible un échantillon *biaisé* [30].

Cette entrée en statistique a le mérite de faire rencontrer, fût-ce sur le mode du récit, ce qui fait la chair du travail statistique : formulation d'un problème, définition d'une population, définition d'un plan d'échantillonnage, constitution d'échantillons, calcul de paramètres statistiques, détermination de l'intervalle d'un paramètre ou test d'hypothèse. Or d'un tel ensemble, on le sait, la statistique calculatoire élémentaire ne retient guère que le calcul de

---

[24] *Ibid.*

[25] *Ibid.*, pp. 18-19.

[26] Dans une note infrapaginale du chapitre III, consacré aux « modèles statistiques usuels », « le lecteur est instamment prié de lire le chapitre 13 (pages 137 à 151) » des *Mathematical Methods of Statistics* de Crámer.

[27] *Op. cit.*, p. 27.

[28] *Ibid.*, p. 28.

[29] Voir par exemple Droesbeke et Tassi 1990, chapitre IV.

[30] *Ibid.*, pp. 29-30.

quelques paramètres de tendance centrale et de dispersion (mais non de paramètres de forme, du moins en France).

En vérité, même, le statisticien « officiel » se bat d'abord sur la définition des *nomenclatures*, c'est-à-dire sur les découpages qui « fabriquent » le réel économique et social. Les exemples seraient ici innombrables : Michel Volle note ainsi que, au XIX^e^ siècle, les professions libérales incluaient « de nombreux salariés, en particulier au service de l'État » (la restriction des professions libérales au statut non salarié remonte à la deuxième guerre mondiale), et que, tandis que la catégorie des fonctionnaires renvoyait alors plutôt à notre catégorie actuelle des hauts fonctionnaires (elle incluait les professeurs de faculté, mais non les professeurs des lycées et collèges, assimilés à des employés), on ne connaissait encore ni techniciens, ni agents de maîtrise, ni cadres [31]. Ce façonnage du réel socio-économique comporte sa part d'arbitraire et doit à ce titre être questionnée : interrogé sur les raisons qu'il avait eues de classer les contremaîtres avec les ouvriers (et non avec les cadres), le concepteur de la première version (1951) du code des CSP (catégories socioprofessionnelles) répondit simplement : « Les contremaîtres, ce sont des gens qui ont des gros bras et qui sifflent. »

Ces problèmes de spécialistes sont pourtant au cœur d'une éducation citoyenne en matière statistique. Lorsque sont connus les résultats du recensement de 1975, l'hebdomadaire *L'Express*, dont le lectorat revendiqué est surtout composé de cadres, titre : « Les cadres : 4 250 000 problèmes [32] ». La surévaluation, sans doute intéressée, est ou devrait être flagrante. À cette époque, en effet, les catégories « ingénieurs » et « cadres administratifs supérieurs », qui bornent l'imaginaire de l'hebdomadaire, comptent respectivement 256 000 et 654 000 personnes : le total – 910 000 – nous laisse loin du compte ! Pour parvenir à l'effectif annoncé, il faut ajouter d'abord les « professions libérales » (172 000) et les « professeurs, professions littéraires et scientifiques » (337 000), mais aussi les « cadres moyens », groupe composé de 737 000 « instituteurs », 299 000 « services médicaux et sociaux » (infirmières, kinésithérapeutes, psychologues, assistantes sociales, etc.), 759 000 « techniciens » et 970 000 « cadres administratifs moyens », cette dernière catégorie comprenant « près d'un tiers de fonctionnaires du cadre B (niveau contrôleur) auxquels s'ajoutent des représentants de commerce, des comptables et des secrétaires de direction qui peuvent, eux, avoir le statut de cadre » !

De tels problèmes ne semblent trouver aucun écho dans la statistique calculatoire scolaire : ils sont pourtant légion, et il y a donc là une occasion manquée de contribuer, dans le cadre même des programmes actuels, à une solide éducation au bon usage de l'information chiffrée. Un quotidien national titre : « 50 000 clandestins chinois en Île-de-France [33] ». L'article correspondant indique que « sur les 217 000 immigrés clandestins que compte l'Île-de-France, 10,82 % sont asiatiques ». Outre qu'il y a quelque bizarrerie à préciser sans commentaire l'effectif d'une population de clandestins, on observera ici que, puisque 10,82 % × 217 000 = 23479,4 (notons la précision du pourcentage), il y aurait ainsi, en Île-de-France, moitié moins de clandestins asiatiques que de clandestins chinois, ce qui illustre de manière cocasse la difficulté à définir une population. Un second exemple, emprunté à l'un des animateurs de l'association *Pénombre* [34], permettra de faire mieux apparaître encore l'impéritie de la statistique calculatoire scolaire. Dans un article sur la délinquance paru dans une édition locale d'un quotidien national le 9 février 1999, on apprend que la délinquance des mineurs est en forte augmentation : l'information, nous dit-on, a été officiellement confirmée par le préfet, qui indique que 43,8 % des infractions de voie publique (en 1998) sont le fait de mineurs ; le quotidien ajoute que « les délinquants sont de plus en plus jeunes ».

---

[31] Volle 1984, pp. 154-155.

[32] Cet exemple est emprunté à Desrosières et Thévenot 1996, pp. 5-6.

[33] *France-Soir*, 31 mai 1990 ; cité *in* Klatzmann 1996, pp. 50-51.

[34] Voir http://www.unil.ch/penombre/inedits/04.htm.

Il précise à l'appui de son affirmation que, sur les 7625 mineurs mis en cause en 1998 (contre 6547 l'année précédente), 49 % ont de 16 à 18 ans, 41,2 % ont de 13 à 16 ans et 9,8 % ont moins de 13 ans. Or toutes ces données ne peuvent en aucun cas venir à l'appui de ce qui est affirmé : *l'augmentation* de la délinquance des mineurs. En 1998, nous dit-on, 43,8 % des infractions seraient de leur fait. Mais cela ne nous dit rien du *nombre* d'infractions qui peuvent leur être imputées, ni surtout de *l'évolution* de ce paramètre entre 1997 et 1998. Si, au reste, nous savions que, en 1997, le pourcentage analogue avait été de, disons, 47,4 %, nous ne serions pas plus avancés : nous pourrions conclure que la *part* des infractions de voie publique due à des *mineurs* a *diminué* d'une année sur l'autre, sans que cela permette de répondre à la question abordée dans l'article ! La même observation peut être faite à propos de l'*âge* des mineurs délinquants : rien ne permet d'affirmer ici que « les délinquants sont de plus en plus jeunes ». Comme le note l'auteur auquel cet exemple est emprunté, « le lecteur attentif, pas fâché avec les opérations arithmétiques et muni d'une calculette, peut voir que la délinquance des mineurs a augmenté de 16,5 %. Mais comme on ne lui donne pas l'augmentation globale de ce phénomène, ni celle de la population, il ne peut rien en conclure ».

Dans le premier exemple, une information chiffrée était donnée *qui ne répondait à aucune question* (même formulée affirmativement). On peut surtout craindre, en un tel cas, que le lecteur ne tire *motu proprio* cette conclusion qu'il y a décidément beaucoup de « clandestins » en Île-de-France, alors même que rien ne permet de conclure dans ce sens : comment savoir si 50 000 « clandestins », c'est « beaucoup » ou c'est « peu » ? Dans le second exemple, au contraire, une question est, si l'on peut dire, abordée – est-il vrai que la délinquance des mineurs ait augmenté ? – mais les chiffres exhibés sont sans pertinence pour répondre à la question. D'une façon très générale, devant toute information chiffrée, une formation statistique minimale devrait conduire à se poser par exemple les questions suivantes : quelle est ou quelles sont les *questions* auxquelles les chiffres proposés sont censés apporter un élément de réponse ? Ces chiffres sont-ils *cohérents* entre eux ? Sont-ils au contraire *redondants* ou sont-ils utilisés pour provoquer des *effets de redondance* ? Les éléments de réponse explicites ou implicites sont-ils *réellement étayés* par les chiffres produits ? Sur quels caractères, sur quelles variables, relatifs à quels échantillons de quelles populations, fournissent-ils une information statistique ? Les données statistiques ont-elles été définies et recueillies de manière claire et univoque ? Sinon, quelles ambiguïtés, quelles incertitudes entachent leur définition ou leur recueil ? En particulier, peut-on contester, dans le processus de fabrication des chiffres présentés, certains des choix faits (définition de la population, du ou des échantillons, des valeurs du caractère observé, etc.) ou encore la mise en œuvre statistique concrète de ces choix ?

L'ingénuité distante de la statistique calculatoire à l'endroit de ce qui n'est pas la pure arithmétique des résumés statistiques fait un dur contraste avec les incertitudes d'un véritable arraisonnement statistique du réel, où chaque étape compte. Dans la perspective de la statistique calculatoire, en effet, ces concepts statistiques conquis historiquement de haute lutte ne sont guère plus que « du vocabulaire ». Et loin que ce vocabulaire arme la pensée pour modéliser le réel, nombre de professeurs semblent n'y voir que l'une de ces pesantes obligations qu'impose le programme ! Le sentiment prévaut, ici, que les *raisons d'être* de ce « vocabulaire », et avec lui toute la *problématique de la statistique*, restent inconnues de qui a pourtant mission de l'enseigner. Une distance quasi rédhibitoire se creuse entre savoir à enseigner et savoir enseigné – une distance qui apparaît comme une donnée lourde de l'enseignement actuel.

## 5 *Légitimité et illégitimité de la connaissance statistique*

La *production* d'une information statistique ne saurait être formelle, immotivée. Ainsi qu'on l'a dit plus haut, elle part d'une *question*, à laquelle on s'efforce d'apporter réponse. Arrêtons-nous un instant sur une question soulevée par un professeur stagiaire de mathématiques ayant en responsabilité une classe de seconde : « L'écart type des moyennes trimestrielles de ma classe de seconde est de 4. Est-ce beaucoup ? » Pour répondre à cette question, la technique à mettre en œuvre est, dans ses grandes lignes, la suivante : 1) on traduit la question posée en termes statistiques : sur la *population* $\Omega$ des classes de seconde, quelle est la *fréquence* de l'*événement* $\{ X \geq 4 \}$, où $X$ est le caractère « écart type de la série des moyennes trimestrielles de mathématiques des élèves de la classe » ; 2) on recueille les écart types $X(\omega)$ d'un ou de plusieurs *échantillons* $(\omega_1, \omega_2, \ldots, \omega_n)$ de classes de seconde : $(x_1, x_2, \ldots, x_n) = (X(\omega_1), X(\omega_2), \ldots, X(\omega_n))$ ; 3) on décrit les séries $(x_1, x_2, \ldots, x_n)$ des écart types ainsi recueillis en vue d'en *inférer* la fréquence de $\{ X \geq 4 \}$ dans la population des écarts types $\{ X(\omega) / \omega \in \Omega \}$. C'est cette démarche statistique que met en œuvre, sans en avoir conscience, le professeur débutant qui, soucieux de contrôler ses « résultats », s'enquiert discrètement auprès de, disons, deux de ses collègues du lycée où il effectue son stage en responsabilité de l'écart type de leur classe de seconde, « pour comparer ». Imaginons qu'il obtienne pour réponses 2,16 et 1,68 : sur ce très petit échantillon – qui a surtout le mérite d'être disponible –, l'événement $\{ X \geq 4 \}$ a une fréquence nulle. Le jeune professeur se risque à en inférer, *sous réserve de confirmation*, qu'il semble bien qu'un écart type de 4, « c'est vraiment beaucoup ». Puis il se décide à poursuivre son *enquête* et obtient ainsi l'échantillon suivant : 3,55 ; 3,03 ; 2,22 ; 2,35 ; 2,70 ; 3,04 ; 3,96 ; 3,66 ; 3,33. La conjecture formulée paraît se confirmer ! Notre élève professeur demande alors à un camarade de promotion à l'IUFM d'enquêter dans le lycée où il est en stage. Son jeune collègue lui rapporte la série suivante d'écarts types : 3,66 ; 3,5 ; 2,9 ; 3,41 ; 2,87 ; 2,6 ; 2,55. Le professeur stagiaire décide d'arrêter là son enquête, car la conclusion lui paraît ne plus faire de doute : 4, c'est vraiment beaucoup. En d'autres termes, il semble bien que, dans le cas examiné, la fréquence de l'événement $\{ X \geq 4 \}$ dans la population $\Omega$ des classes de seconde soit, sinon nulle, du moins fort proche de zéro.

L'exemple est sans doute très simple ; mais il illustre un comportement qui ne va nullement de soi, même aujourd'hui, même dans l'institution scolaire : enquêter, pour connaître et faire connaître. Nous vivons en effet dans des sociétés où, de manière souvent explicite, des cloisonnements rigides délimitent des domaines de légitimité exclusifs, l'interdit les concernant étant intériorisé en un véritable *habitus* : il y a ce dont on s'occupe (sa classe, pour un professeur), et il y a ce qui n'est pas de sa compétence, *ce qu'on n'a pas à connaître*, et qu'il vous est interdit d'approcher, de commenter, etc. Pour faire sentir mieux la chose, nous mentionnerons ici un épisode que rapporte, dans ses *Mémoires*, André Morellet (1727-1819), cet abbé que ses amis philosophes – au nombre desquels figuraient Voltaire et d'Alembert – appelaient *Mords-les !* pour sa promptitude à polémiquer :

> En 1764, M. de Laverdy, alors contrôleur général, ayant fait rendre un arrêt du conseil qui défendait d'imprimer sur les matières d'administration, sous peine d'être poursuivi extraordinairement, ceux qu'on appelait alors philosophes furent indignés ; et j'étais de ce nombre. Je combattis pour la liberté de la presse, et j'intitulai mon ouvrage *De la liberté d'écrire et d'imprimer sur les matières de l'administration*. C'était le développement d'une partie du *Traité de la liberté de la presse* que j'avais commencé à la Bastille [...]. Je gardais ici une extrême modération, afin de ne pas rencontrer d'obstacles ; mais cette réserve ne me servit de rien, et je ne pus obtenir pour moi-même la liberté que je demandais pour tous. Cependant mon travail n'avait pas déplu à M. Trudaine ; son fils l'avait communiqué à M. Chauvelin, intendant des finances, et celui-ci au contrôleur général : mais le ministre y fit une réponse à mi-marge, tout entière de maximes despotiques, ou de la théorie des premiers commis : « pour parler d'administration, il

> faut tenir la queue de la poêle, être dans la bouteille à l'encre ; et ce n'est pas à un écrivain obscur, qui, souvent, n'a pas cent écus vaillants, à endoctriner les gens en place ». J'ai longtemps gardé ce précieux document. On comprend bien que mon ouvrage ne fut pas alors imprimé ; mais en 1774, M. Turgot étant arrivé au ministère, je le publiai avec l'épigraphe de Tacite : *Rara temporum felicitate, ubi sentire quae velis, et quae sentias dicere licet* [« … le rare bonheur d'une époque où l'on peut penser ce que l'on veut et dire ce que l'on pense »]. (Morellet 1822, p. 158)

Cet interdit de connaître, à propos duquel l'un de nous [35] a émis l'hypothèse, lors de l'école d'été 2001, qu'il était une des causes de la fragmentation savante de la production des connaissances, se saisit *a fortiori* de l'École. Rappelons seulement les deux données fondamentales qui ordonnent l'existence, le développement et le dépérissement des curriculums scolaires : d'une part, la partition relativement stricte des domaines de connaissance étudiés, chaque discipline scolaire se voyant assigné le *monopole à peu près exclusif* de la connaissance d'un champ restreint de conditions et de contraintes de la vie des hommes, ce qui aboutit au *confinement épistémologique* des disciplines scolaires ; d'autre part, l'évidement, par *purification épistémologique*, des domaines de connaissance disciplinaires, phénomène particulièrement net dans le cas des mathématiques aujourd'hui. Ce second attribut des disciplines scolaires est, apparemment, ce qui bloque presque immédiatement la mise en place d'un enseignement de la statistique articulé à l'exploration d'au moins quelques catégories de conditions et de contraintes de la vie en société. Mais le premier phénomène est peut-être d'abord le plus révélateur d'un fait de civilisation massif, dont les lois gouvernant la formation et l'évolution des curriculums scolaires ne sont au mieux qu'un révélateur. D'une manière générale, on l'a dit, l'accès à la connaissance est soumis à autorisation : un interdit de connaissance fondamental organise notre rapport au monde naturel et social, interdit qui n'est jamais levé que partiellement, dans des circonstances particulières, comme on accorde un privilège. Le découpage en disciplines est à cet égard une conquête réelle contre une tyrannie maintenue, mais une conquête que cette tyrannie même dénature : car si, à l'intérieur d'une discipline, une certaine permission de connaître est accordée, elle est en même temps *étroitement contrôlée*, avec, en particulier, ce paradoxe que le fait d'avoir été *élève* (ou *étudiant*) ne vous autorise pas à faire état légitimement des connaissances acquises, hormis bien sûr si vous continuez d'occuper la position d'élève en une institution autorisée à diffuser la connaissance en cause. Le circuit est clos. La connaissance d'un domaine fait, scolairement, l'objet d'un monopole local, même si plusieurs disciplines se disputent quelquefois une partie du domaine.

La fragile compatibilité de l'École avec l'interdit de connaissance se construit par le tracé, régulièrement repris, de frontières disciplinaires officielles qui définissent le partage curriculaire de la curiosité autorisée et de l'enquête déplacée. Le confinement disciplinaire respecte d'autant mieux l'interdit de connaissance qu'il est renforcé par un phénomène déjà mentionné : l'éviction de la discipline des contenus « ambigus » qui pourraient donner lieu à des conflits de frontière. Ce processus de purification épistémologique de la discipline est généralement corrélatif de la montée d'un « chauvinisme » disciplinaire qui aboutit fréquemment à ce qu'on peut appeler un *narcissisme* disciplinaire – quand la discipline ne s'occupe plus que d'elle-même, comme il en va trop souvent aujourd'hui en mathématiques. La fermeture au monde extérieur prend quelquefois aussi une forme trompeuse, dans laquelle le monde extérieur est certes accueilli, mais pour constituer au sein du monde disciplinaire « vrai » un *univers d'opérette*, un monde sans réalité, pur décorum à la manière d'un village Potemkine. L'incitation à *prendre au sérieux* le monde extramathématique en est diminuée d'autant. Le programme de seconde actuel, ainsi, paraît mû davantage par une exquise urbanité à l'égard des élèves et de leurs professeurs que par le souci de la pertinence

[35] Voir Chevallard 2002.

formative des sujets d'étude retenus. Certes, l'enseignant, y lit-on ainsi, devra traiter « des données en nombre suffisant pour que cela justifie une étude statistique » ; mais on ajoute aussitôt : « il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'actualité et de ses goûts ». Manière de dire que les questions extramathématiques prétendument étudiées ne sont guère abordées que dans une optique *occasionnaliste*, pour « faire de la statistique », afin d'ailleurs surtout de « faire des maths », toutes choses qui, ainsi qu'on l'a suggéré, ont au reste fort peu de chances de se produire.

En matière de statistique, le problème général de l'interdit de connaissance se pose dans des conditions qu'il convient de spécifier. Si l'État a, selon Max Weber (1864-1920), le monopole de la violence légitime, il possède non moins le *monopole de la connaissance légitime*. Or la production statistique est, en cette matière plus peut-être qu'en aucune autre, *une affaire d'État*. En France, l'information statistique produite par l'INSEE et les différents ministères recouvre aujourd'hui l'essentiel du domaine économique et social. Cette situation de monopole partagé est le fruit d'une évolution qui, à partir des années 1960 surtout, conduit plusieurs des statisticiens de l'INSEE à travailler dans les ministères (lesquels doivent, selon la doctrine de l'époque, développer la statistique de leur domaine propre, Industrie, Agriculture, Équipement, etc.), pratique que la décentralisation ne fera ensuite qu'approfondir. Ainsi une part de « l'institution statistique » est-elle placée au plus près du pouvoir d'État, ce qui ne manque pas de créer des tensions « entre les statisticiens restés à l'INSEE et ceux qui travaillent dans les ministères [36] ». Car on ne saurait trop exagérer la fascination que la puissance de l'État peut exercer sur les meilleurs esprits : « il devint " chic " à l'INSEE, note Volle, d'être comptable national – ou, mieux encore, de participer aux travaux du Plan – les tâches proprement statistiques étant par contrecoup relativement dévalorisées et considérées comme du " charbon " [37] ». La toute-puissance de l'État a une conséquence sur laquelle il est habituel de se montrer discret, mais qu'il convient pourtant d'avoir à l'esprit : la forte dépendance des études scientifiques de type sociologique ou économique par rapport aux services de l'État dont ces études utilisent la production statistique. Le chercheur, fût-il statisticien « officiel », est alors confronté à des limites qu'il n'est guère en son pouvoir de déplacer, même quand il en perçoit les effets invalidants au plan scientifique. Volle note sobrement, à cet égard, que si les statisticiens « définissent des champs d'observation comportant des lacunes qui nous paraissent choquantes, c'est peut-être parce qu'ils se heurtent, lorsqu'ils veulent combler ces lacunes, à des obstacles et à des interdits qu'ils ne parviennent pas à surmonter [38] ». On peut comprendre que de telles difficultés découragent [39]. Le même auteur suggère au reste un possible facteur d'indifférence face au problème de l'accès à des données chiffrées conceptuellement pertinentes et techniquement idoines. Selon un schéma d'inspiration weberienne, il brosse une opposition entre culture « protestante » et culture « catholique » : pour des raisons qu'on laissera le lecteur découvrir [40], la première conduirait à valoriser, voire à sacraliser, la comptabilité et, par continuité, la connaissance statistique ; la seconde déterminerait une attitude faite de détachement et même de scepticisme à l'endroit de l'information quantitative. Quels que soient les effets réels de tels facteurs culturels de longue durée, on retiendra cette idée d'un axe opposant deux attitudes

---

[36] Volle 1984, p. 122. Pour une présentation sur une plus longue durée – depuis la création, dans la première partie du XIX[e] siècle, de la Statistique générale de la France (SGF) – de l'appareil statistique d'État en France, voir Desrosières 1993, pp. 185-203.

[37] *Op. cit.*, p. 122.

[38] *Ibid.*, p. 129.

[39] Certains domaines restent ainsi mal connus : activités des grands groupes de sociétés, aide de l'État aux entreprises, hauts revenus, etc.

[40] *Op. cit.*, p. 84.

générales face à l'information chiffrée – l'une de recherche exigeante, voire passionnée, l'autre d'ignorance insoucieuse et même dédaigneuse.

S'autoriser à enquêter sur le monde n'est jamais anodin : une telle décision est en essence *politique*, y compris dans sa dimension cognitive, et se heurte aux « autorités » établies, à l'État en tout premier lieu. Aussi enquêtent d'abord ceux qui, soit s'opposent à l'État et à l'ordre établi, soit, à l'inverse, tiennent « la queue de la poêle » et, d'une manière ou d'une autre, participent de l'État. À la première catégorie appartient par exemple Karl Marx (1818-1883), dont Volle nous rappelle qu'il utilisait abondamment et d'une main sûre les statistiques disponibles, corrigeant à l'occasion ses prédécesseurs, tel Ricardo pour une erreur commise dans l'emploi de moyennes [41]. De la seconde catégorie relève celui en qui on peut voir le créateur de la statistique et de la démographie, William Petty, dont Marx saluait la « géniale audace » et l'esprit hardi – tout en reconnaissant qu'il était « tout aussi doué pour piller en Irlande sous l'égide de Cromwell que pour se faire donner à force de courbettes, par Charles II, le titre de baronnet en couverture de ses rapines [42] » ! Par ce que le démographe Hervé Le Bras a dénoncé comme une falsification intéressée, l'histoire officielle de la statistique a substitué à Petty l'un de ses contemporains, John Graunt, drapier féru de chiffres qui, souligne Le Bras, « incarne l'image désintéressée et professionnelle du savant dans laquelle les statisticiens et les démographes aiment encore à se reconnaître de nos jours [43] ». Pourtant, si l'on admet que Graunt n'est qu'un prête-nom, que pour l'essentiel les *Observations naturelles et politiques sur les bulletins de mortalité de la ville de Londres* (1662) sont dues au génie de William Petty, homme de science et de pouvoir tout à la fois [44], une autre réalité apparaît : car avec Petty, « la statistique et plus encore la démographie révèlent une étroite proximité avec le pouvoir d'État, absolu et discrétionnaire [45] ». La naissance de ces sciences va de pair avec une certaine configuration *politique*.

La genèse de la notion statistique de *population* fournit de ce constat une illustration pertinente. Petty avait été l'assistant, l'*amanuensis*, de Thomas Hobbes (1588-1679) – qui lui-même avait été celui de Francis Bacon (1561-1626). Dans le chapitre XIII de son *Léviathan* (1660), Hobbes fonde politiquement la notion de population humaine en définissant la notion d'égalité entre les hommes, soit le fait qu'on puisse les regarder, ainsi que nous dirions aujourd'hui, comme des « unités statistiques » dont chacune vaut autant que toutes les autres. Le critère de cette égalité, Hobbes le trouve en dernière instance dans un fait terrible, la capacité de chacun à… tuer son prochain ; car, observe-t-il, « pour ce qui est de la force corporelle, l'homme le plus faible en a assez pour tuer l'homme le plus fort, soit par une machination secrète, soit en s'alliant à d'autres qui courent le même danger que lui ». Une telle relation d'équivalence entre les hommes n'existe ni chez Platon, ni chez Aristote, note Le Bras, qui propose alors cette image éclairante : si j'ai dans mon panier à provisions un chou-fleur, 3 concombres, 18 pommes de terre et 2318 petits pois, je ne dirai pas que mon panier contient 2340 légumes ! On n'imagine pas d'égaler un petit pois et un chou-fleur, pas davantage qu'on ne pouvait, dans l'Antiquité, égaler un citoyen, un esclave, un métèque. Qu'il soit loisible de parler d'un nombre de citoyens, d'un nombre d'esclaves, d'un nombre de métèques n'autorise pas pour autant à parler d'un nombre d'*hommes* – « population » purement virtuelle, faute d'être *politiquement instituée*. On voit ainsi que le fait même de modéliser un collectif comme une « population » suppose un engagement, une prise de

---

[41] *Op. cit.*, p. 97.
[42] Cité *in* Volle 1984, p. 91, n. 13.
[43] Le Bras 2000, p. 7.
[44] L'auteur que nous suivons ici le présente ainsi : « médecin, mathématicien, courtisan proche à la fois des Cromwell et des Stuart, disciple de Hobbes et fondateur de l'"Arithmétique politique" et de la première académie scientifique moderne, la Royal Society. » (Le Bras, *op. cit.*, 4e de couverture)
[45] *Ibid.*, p. 10.

position face à un certain état du monde. L'interdit *disciplinaire* pesant sur un tel engagement masque en général ce qui est fondamentalement un interdit *politique* – l'interdit disciplinaire fonctionnant comme un substitut plus présentable de l'interdit politique.

### *6 Statistique et émancipation citoyenne*

À titre d'illustration, esquissons d'abord un exemple simple d'étude statistique que l'on peut mener en seconde [46]. Supposons que, à la rentrée 2003, deux élèves d'une seconde marseillaise, s'appuyant sur leur expérience du système scolaire, et avec l'assentiment du professeur, proposent à leur classe d'étudier la question suivante : *D'après un reportage vu à la télé, dans l'académie de Paris l'effectif des élèves des lycées représenterait presque les deux tiers de l'effectif des élèves des collèges. Ce pourcentage (66 %) est-il élevé ? Est-il supérieur à la moyenne des académies ? Le pourcentage correspondant pour l'académie d'Aix-Marseille est-il du même ordre ? Ou bien est-il inférieur, de l'ordre par exemple de 50 % ? Ce pourcentage, en tout cas, est-il inférieur à la moyenne des académies ?*

Imaginons aussi que la classe accepte la proposition. Une recherche sur le site Internet du ministère de l'Éducation nationale conduit à une note d'information [47] publiée en juin 2003, intitulée *Les élèves du second degré dans les établissements publics et privés à la rentrée 2002*. Cette note contient un tableau où l'on découvre d'abord que, sous le nom de « lycées », les élèves auteurs de la question se réfèrent implicitement à ce que le tableau nomme « 2$^{\text{nd}}$ cycle général et technologique », et non aux lycées professionnels (« 2$^{\text{nd}}$ cycle professionnel »). Pour l'académie de Paris, en suivant ce tableau, on a : $\frac{\text{Effectif lycée}}{\text{Effectif collège}} = \frac{56\ 880}{86\ 609}$ $\approx$ 65,7 %. Pour l'académie d'Aix-Marseille on a semblablement : $\frac{\text{Effectif lycée}}{\text{Effectif collège}} = \frac{67\ 682}{148\ 345} \approx$ 45,6 %. L'information rapportée par les deux élèves à propos de l'académie de Paris est donc exacte. La valeur envisagée pour l'académie d'Aix-Marseille est en revanche un peu surestimée, mais n'est pas si loin de la vérité ! La série complète des rapports lycée/collège est : Aix-Marseille, 45,6 % ; Amiens, 44,5 % ; Besançon, 45,1 % ; Bordeaux, 45,2 % ; Caen, 44,3 % ; Clermont-Ferrand, 46,5 % ; Corse, 44,9 % ; Dijon, 44,5 % ; Grenoble, 46,2 % ; Lille, 48 % ; Limoges, 48,7 % ; Lyon, 46,4 % ; Montpellier, 43,2 % ; Nancy-Metz, 47,2 % ; Nantes, 45,7 % ; Nice, 43,7 % ; Orléans-Tours, 44,7 % ; Poitiers, 45 % ; Reims, 44,8 % ; Rennes, 49,8 % ; Rouen, 46,6 % ; Strasbourg, 42,3 % ; Toulouse, 46,1 % ; Paris, 65,7 % ; Créteil, 46,3 % ; Versailles, 47,6 %. On voit que l'académie d'Aix-Marseille semble être « dans la moyenne », alors que l'académie de Paris est en fait l'unique « point aberrant » – le seul *outlier*, comme on dit en anglais – de cette série. La série ordonnée obtenue en multipliant par 100, en retranchant 40, puis en multipliant par 10, est : 23 ; 32 ; 37 ; 43 ; 45 ; 45 ; 47 ; 48 ; 49 ; 50 ; 51 ; 52 ; 56 ; 57 ; 61 ; 62 ; 63 ; 64 ; 65 ; 66 ; 72 ; 76 ; 80 ; 87 ; 98 ; 257. L'intervalle médian est ici [56, 57]. Pour la série d'origine, les pourcentages médians appartiennent à l'intervalle [45,6 %, 45,7 %] : l'académie d'Aix-Marseille est donc bien une académie « moyenne ». La moyenne arithmétique des pourcentages est 46,5 % quand on conserve le pourcentage de l'académie de Paris dans la série et 45,7 % quand on l'en exclut : on retrouve que l'académie d'Aix-Marseille est « dans la moyenne ».

Toute étude statistique qui ne se réduit pas à une arithmétique sans contenus suppose que *l'on intervienne sur le monde* – intellectuellement, symboliquement, politiquement. Est-il vrai, par exemple, que les anciens footballeurs professionnels meurent plus de cancer que le reste de la population ? Est-il vrai, encore, que les femmes cadres soient davantage victimes

---

[46] On ne détaille pas ici l'exploitation qu'il serait judicieux de faire, au cours d'une telle étude, des différents outils dont la maîtrise est inscrite au programme de cette classe (calcul de moyennes, de médiane, etc.).
[47] Voir http://www.education.gouv.fr/stateval/ni.htm.

de violences conjugales que les ouvrières ? Etc. On ne peut guère prétendre enseigner la statistique, fût-ce en un cours de mathématiques, si on la réduit à la mise en œuvre de quelques algorithmes sur des données controuvées qui ne renvoient à aucun enjeu de connaissance. Le parti contraire, on l'a suggéré, bute pourtant sur un obstacle majeur – cet interdit de connaissance qui est le négatif de l'existence des divers monopoles, disciplinaires ou étatiques, scolaires et extrascolaires, de la connaissance légitime. Hors du secondaire général, mais déjà dans l'enseignement secondaire de type professionnel, une manière classique de se prémunir sur ce point est de s'enfermer dans le domaine extramathématique dont on est censé s'instruire : on étudiera ainsi la statistique en psychologie, ou la statistique en gestion, ou la statistique en médecine, ou la statistique en écologie, ou en contrôle de qualité, etc., non sans qu'un tribut soit payé à l'interdit de connaissance sous la forme, notamment, d'un renoncement mathématique.

De manière sans doute plus restrictive en réalité, mais en apparence moins « disciplinée », plus libre, on pourra s'intéresser à des phénomènes variés à la condition qu'ils soient « proches » dans la culture partagée avec les élèves : ce vers quoi pointait l'ancien programme de seconde en précisant que les « documents nécessaires » pouvaient, de manière privilégiée, être « empruntés à l'environnement de l'élève ». On retrouve alors un type de solution au problème de l'interdit de connaissance que mettait classiquement en œuvre (sous une forme « réaliste », matérielle) la *leçon de choses* chère à Ferdinand Buisson (1841-1932) : en travaillant à partir d'objets ou d'informations normalement accessibles à faible coût institutionnel dans l'environnement ordinaire de l'élève, on limite automatiquement les risques de conflit liés à une curiosité illégitime pour les choses du monde. Sans doute ce principe peut-il encore servir à contourner prudemment des sujets d'étude trop exposés. Dans cette perspective, plusieurs cas peuvent au reste être distingués, par degré croissant d'illégitimité. Il n'est pas équivalent, ainsi, que l'on étudie statistiquement, dans la classe de mathématiques, un phénomène qui ne tombe sous la juridiction d'aucune discipline connue dans la culture de la communauté éducative, ou, plus simplement, qui ne relève d'aucune discipline enseignée au secondaire, ou au contraire qui est étudié dans l'une des disciplines enseignées au secondaire, mais pas dans la classe concernée, ou qui, entrant dans le champ de compétence d'une discipline enseignée dans la classe en question, relève du programme d'une année antérieure ou postérieure, ou à l'inverse qui figure au programme de l'année en cours, etc. Il n'en reste pas moins que la création d'un enseignement de statistique dans le cours de mathématiques, et en vérité la reconstruction aujourd'hui nécessaire de tout l'enseignement des mathématiques, suppose que l'on s'autorise des curiosités dont, étant donné l'histoire de nos sociétés, le seul fondement légitime, commun à tous, est notre statut de *citoyen* ou de futur citoyen. L'élève ne se forme pas seulement à devenir citoyen ; *c'est parce qu'il se pose en futur citoyen qu'il peut se former*. Car en tant que tel, nul ne peut légitimement contester sa légitimité à mettre en œuvre le mot d'ordre de Condorcet : « Il faut oser tout examiner, tout discuter, tout enseigner même. »

## 3. LA DIFFERENCE SPECIFIQUE : LA VARIABILITE

### 1 *Le déni de la variabilité*

Pour construire une « théorie » mathématisée de la spatialité ou de la numérosité, il est nécessaire de pouvoir développer, contre les censures éventuelles, des rapports adéquats avec des formes appropriées de spatialité ou de numérosité. Pour construire une théorie mathématisée de la *variabilité*, il convient de même de créer et de maintenir un commerce convenable avec des formes empiriques adéquates de variabilité. On sait que, même dans les deux premiers cas – géométrie et arithmétique –, la reconquête du métissage objectal des milieux didactiques scolaires est fort loin d'être accomplie – il n'est pas certain que la « chute dans l'insensible » des mathématiques enseignées ne se poursuive pas. S'agissant de variabilité, le problème est plus difficile encore, dans la mesure où, par delà l'interdit général de l'enquête sur le monde, une donnée massive doit être prise en compte : la résistance de la culture courante au *fait* de la variabilité.

La forme première de ce déni de la variabilité se trouve dans un célèbre petit apologue que reprend Daniel Schwartz dans un livre déjà mentionné [48] : débarquant à Calais, un Anglais aperçoit une femme rousse et en conclut que toutes les Françaises sont rousses… Les anecdotes sont à cet égard innombrables, telle celle-ci que rapporte le même auteur : « Un ami mathématicien, connu pour sa logique impeccable, me rencontre à Paris un samedi : “ Tiens, s'exclame-t-il. Tu ne vas plus le samedi à la campagne ? ” » Revenant sur ce sujet à la fin de son ouvrage, Schwartz souligne que « nous vivons dans le probable, et nos réflexes, profondément enracinés, sont ceux du certain [49]. » Et de conclure : « Ainsi dans tous les pays la pensée statistique a-t-elle du mal à trouver la place qu'elle mérite. »

Tout se passe en effet comme si la culture ordinaire supposait que chaque objet dont elle reconnaît l'existence était le support de grandeurs *bien déterminées*. C'est ce *postulat d'univocité* que nous mettons en œuvre quand nous demandons quel est le poids d'un éléphant, le prix d'un réfrigérateur, le salaire d'un footballeur professionnel, la durée de vie d'un équipement sportif, etc. Ou quand nous demandons si un joueur de tennis, ça gagne plus qu'un footballeur, si un éléphant, c'est plus gros qu'un hippopotame, si une moto c'est plus cher qu'un scooter, etc. Formellement, étant donné un domaine d'objets $\mathcal{O} = \{ o_1, \ldots, o_n \}$ et des espèces de grandeur $X$, $Y$, $Z$ (prix, poids, salaire), nous supposons que les grandeurs $X(o_i)$, $Y(o_i)$, $Z(o_i)$, où $o_i$ est un « agrégat » (footballeur, réfrigérateur, éléphant, etc.) que nous fabriquons avec des mots et des pratiques sociales, sont bien définies. En conséquence, nous ne doutons pas qu'on puisse décider si l'on a par exemple $X(o_i) \prec X(o_j)$ ou l'inverse, etc. Sans doute pourrait-on montrer que nous agissons souvent pour que l'univers créé par l'homme ait une variabilité *réduite* – en fabriquant des objets à durée de vie contrôlée, en imposant des réfections d'équipements selon un calendrier convenu, en calibrant les fruits et légumes, etc. L'action anthropique vise ainsi à imposer au monde, de vive force, un axiome d'univocité, afin de le rendre compatible avec l'appréhension spontanée que nous en avons.

Montrons sur un exemple – celui de la *mesure des grandeurs* – combien l'assomption de la variabilité reste un problème ouvert pour la culture commune. Dans la version la plus simplifiée possible, un objet plan, plateau de table ou champ, a des dimensions *exactes* : un champ de 14 m de large et 25 m de long a une aire de $14 \text{ m} \times 25 \text{ m} = 350 \text{ m}^2$. Tel est l'univers métrologique dans lequel baignent les jeunes générations pendant l'essentiel de leur scolarité. De fait, les textes gouvernant l'enseignement actuel des mathématiques au collège ne distinguent l'exact de l'approché qu'en matière de *calcul*. Ils ne mentionnent les *mesures* approchées que marginalement, dans le document d'accompagnement du programme de troisième, en une section intitulée « Place des grandeurs dans l'enseignement des mathématiques au collège », et de façon quelque peu fuyante – pour dire que « la distinction entre “ mesure exacte ” (…) et “ mesure approchée ” est une question très difficile » et que « le travail mené en mathématiques au long du collège sur calcul exact et calcul approché peut en favoriser l'approche. » Maigre viatique ! Sous le titre *Mesures et incertitudes*, le document d'accompagnement du programme de *sciences physiques* de troisième comporte toutefois des développements plus substantiels. On y souligne qu'il n'y a pas « tout à fait coïncidence entre la valeur exacte d'une grandeur mesurée et la lecture de la mesure correspondante », et cela pour de multiples raisons. Ainsi, « de nombreuses grandeurs mesurées dépendent légèrement

[48] Schwartz 1994, p. 9.
[49] *Op. cit.*, p. 91.

de facteurs physiques (température, pression...) susceptibles d'évoluer en fonction du temps de manière incontrôlée ». En outre, certaines grandeurs, « comme les dimensions d'une table rectangulaire en bois », ne sont définies « qu'avec une certaine précision ».

Comment formaliser ces observations ? Le document cité en explicite la modélisation scolairement correcte à propos de l'exemple d'un voltmètre dont la fiche fournie par le constructeur porterait les indications suivantes : « précision : 0,8 % VL + 2 UR ». Ici, VL désigne la valeur lue, UR l'unité de représentation, valeur de l'unité du dernier chiffre affiché : si, indique le document, on lit U = 12,54 V, on a VL = 12,54 V et 1 UR = 0,01 V ; l'incertitude sur U est alors (d'après le constructeur) : $\Delta U = \left(\frac{0{,}8}{100} \times 12{,}54 + 0{,}02\right)$ V, soit $\Delta U$ = 0,12 V. Le document cité ajoute alors : « La valeur exacte de U, d'après le constructeur, est telle que : 12,54 V – 0,12 V < U < 12,54 V + 0,12 V », soit 12,42 V < U < 12,66 V ; et de conclure : « On peut écrire : U = 12,54 V ± 0,12 V. » On retrouve là l'interprétation traditionnelle au secondaire de la notion de précision de mesure, celle-là même qu'a longtemps popularisée l'enseignement des mathématiques.

Une *Arithmétique* de 1954 pour la classe de mathématiques [50], c'est-à-dire pour la classe terminale scientifique des lycées, consacre ainsi, comme il était alors d'usage, ses ultimes pages au thème des *calculs approchés*. La métrologie exacte de l'école primaire y est tout à coup dépassée : les auteurs partent en effet du constat si longtemps différé que « les nombres dont on dispose dans les calculs sont […] rarement des nombres exacts ». L'incertitude qui les affecte, poursuivent-ils, provient soit de mesures approchées, soit de calculs approchées. D'où un problème, qu'ils explicitent sur l'exemple suivant [51] : « Calculer l'aire d'un champ rectangulaire dont les dimensions $a$ et $b$ mesurées en mètres sont telles que l'on ait : $25{,}4 < a < 25{,}5$ ; $13{,}7 < b < 13{,}8$. » La réponse est connue : « On a : $25{,}4 \times 13{,}7 < ab < 25{,}5 \times 13{,}8$ ou : $347{,}98 < ab < 351{,}90$. L'aire du champ est donc comprise entre 347,98 m$^2$ et 351,9 m$^2$. »

Il est en tout cela un silence éloquent : comment peut-on être sûr, dès qu'un des nombres engagés dans un calcul provient d'une *mesure* (éventuellement par l'intermédiaire d'un calcul lui-même approché), que la « vraie valeur » $x$ cherchée est *encadrée* par deux valeurs $a$ et $a'$ ? En vérité, l'écriture U = 12,54 V ± 0,12 V ne signifie *nullement* que 12,54 V – 0,12 V < U < 12,54 V + 0,12 V, soit 12,42 V < U < 12,66 V. Pour restituer la « vraie » théorie des erreurs de mesure – que l'enseignement secondaire et, en grande partie, l'enseignement supérieur ignorent, du fait sans doute d'une diffusion insuffisante de la culture statistico-probabiliste en ces institutions –, nous suivrons ici brièvement l'auteur d'un ouvrage intitulé *Probabilités et statistiques dans les sciences expérimentales* [52]. Une grandeur $x$ à mesurer se modélise ordinairement par une variable aléatoire continue de densité $\rho$ ; si l'on connaissait cette densité, la valeur moyenne $\overline{x}$ de $x$ serait donnée par $\overline{x} = \mu = \int_{-\infty}^{+\infty} x\rho(x)\mathrm{d}x$ et $x$ aurait pour écart quadratique moyen $\overline{\Delta x^2} = \sigma^2 = \int_{-\infty}^{+\infty} \rho(x)(x-\mu)^2\mathrm{d}x$. En écrivant que $x = \mu \pm \Delta x$, où $\Delta x$ s'identifie à l'écart type $\sigma$, on ne dit alors nullement que $x$ se trouve dans l'intervalle $[\mu - \Delta x ; \mu + \Delta x]$. Du moins lorsque la distribution de $x$ est *gaussienne*, cette écriture signifie seulement que la probabilité de trouver $x$ dans l'intervalle $[\mu - \sigma, \mu + \sigma]$ est égale à… 68,3 % ; semblablement, la probabilité que $x$ appartienne à l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$ est égale à 95,4 %, etc. « Soulignons, insiste l'auteur, que l'interprétation de $\Delta x$ comme une valeur maximale de l'incertitude sur $x$ est fausse. Il faut toujours interpréter le résultat $\mu \pm \Delta x$ au sens des probabilités. » S'il en est ainsi, quand dire alors « qu'il y a désaccord entre deux valeurs

[50] Maillard et Millet 1954. L'ouvrage se réfère au programme du 24 juin 1948.
[51] *Op. cit.*, pp. 203-204.
[52] Belorizky 1998, p. 49.

lorsqu'on compare deux expériences différentes ou une théorie avec l'expérience » ? Une règle usuelle, dite « des 3$\sigma$ », permet de conclure au désaccord dès que la différence entre les deux valeurs comparées excèdent trois fois l'écart type, ce qui est bien autre chose que la simple disjonction des « intervalles d'incertitude ».

On dispose en général, non pas d'une mais de plusieurs mesures $x_i$ ($i = 1, 2, \ldots, N$). On appelle alors *moyenne expérimentale* $x_e$ la moyenne des mesures $x_i$ et *dispersion expérimentale* $s_e = \frac{1}{N-1} \sum_{1 \leq i \leq N} (x_i - x_e)^2$ : on sait que, sous des conditions très générales, quand $N$ tend vers l'infini, $x_e$ et $s_e$ tendent respectivement vers $\mu$ et $\sigma$. Lorsque $N$ est grand, des résultats classiques conduisent à écrire $x = x_e \pm \Delta x$, où on prend pour valeur de $\Delta x$ l'*écart quadratique expérimental* $\frac{s_e}{\sqrt{N}}$ ; lorsqu'il n'en est pas ainsi, on doit introduire le *coefficient de Student* $t_{PN}$ en prenant $\Delta x = t_{PN} \frac{s_e}{\sqrt{N}}$. Ce coefficient correctif dépend de $N$ et de la probabilité de trouver la vraie valeur de $x$ dans l'intervalle $[x_e - \Delta x, x_e - \Delta x]$. L'exemple suivant proposé par l'auteur suivi ici éclaire la chose : on a obtenu six mesures ($N = 6$) d'une masse $m$ (en grammes), à savoir 72,361, 72,357, 72,352, 72,346, 72,344, 72,340 ; la moyenne expérimentale est $m_e = 72{,}350$ g, la dispersion expérimentale vaut $s_e = 8{,}07 \times 10^{-3}$ g et l'écart quadratique expérimental de $m_e$ est $\frac{s_e}{\sqrt{6}} = 3{,}30 \times 10^{-3}$ g. Pour avoir une probabilité de 99 % de trouver la vraie valeur de la masse $m$ dans l'intervalle $[m_e - \Delta m, m_e + \Delta m]$ on doit prendre le coefficient de Student $t_{P=0{,}99\,;\,N=6} = 4$, ce qui donne $\Delta m = 0{,}013$ g : la mesure de la masse $m$ est donc, avec une probabilité de 99 %, comprise entre 72,337 et 72,363.

## 2 *Résistances à la variabilité et droit de connaître*

Dans un cas particulier qui a eu historiquement une valeur paradigmatique en statistique [53], ce qui précède montre deux formes de résistance à la variabilité. La première forme, rarement pertinente mais constamment rencontrée, consiste à faire comme si nous vivions dans un monde *dépourvu de variabilité*. La seconde forme est celle qui conduit à regarder l'écriture $x = x^* \pm \Delta x$ comme exprimant sous une forme compacte l'encadrement $x^* - \Delta x < x < x^* + \Delta x$. En termes probabilistes, cette modélisation revient à dire qu'on suppose une *distribution uniforme* [54] des valeurs de $x$ sur l'intervalle $[x^* - \Delta x, x^* + \Delta x]$. On pourrait, au reste, regarder l'hypothèse classique d'une distribution *normale* des erreurs de mesure comme une troisième forme de résistance à la variabilité, du moins dès lors qu'on tente d'en faire la loi par excellence de tout phénomène variable.

Par contraste avec ces formes de refus, de résistance ou d'acceptation sous réserve de la variabilité, la « vision statistique » conduit à regarder les objets du monde naturel ou social, non comme les supports de grandeurs *fixes*, mais de grandeurs *variables* ou, comme on le dit encore, de grandeurs *aléatoires*, c'est-à-dire, ainsi que le précisait Darmois dans son petit livre déjà cité, « d'une grandeur telle qu'à chacun de ses états corresponde une probabilité déterminée, l'ensemble représentant la loi de probabilité de la grandeur [55] ». Derrière chaque « objet », et pour chaque espèce de grandeur, le regard statistique sur le monde voit une *distribution de probabilités* déterminée. Insistons sur ce point. Dans le livre de Darmois, et de manière au demeurant classique, les premiers exemples proposés concernent le lancer d'un

[53] À nouveau, il faudrait ici évoquer Quetelet : voir à ce propos Desrosières 1993, chapitre 3, et Desrosières 2002.
[54] Une telle distribution des erreurs de mesure avait été envisagée en 1756 par Robert Simpson : voir Droesbeke et Tassi 1990, p. 27.
[55] Darmois 1957, p. 20.

dé : le nombre de points amené, par exemple, est une grandeur variable, dont la loi de probabilité est par définition, si le dé est « bon », la répartition *uniforme* sur l'ensemble { 1, 2, 3, 4, 5, 6 } des valeurs possibles. Or on peut voir là une difficulté : la distribution uniforme apparaît souvent comme la formalisation *d'un refus paradoxal de la variabilité* qui consiste à dire que l'on ne peut rien dire, que, par exemple, la « vraie valeur » $x$ peut être n'importe où dans l'intervalle $[x^* - \Delta x, x^* + \Delta x]$, etc. On pourrait dire de même qu'on meurt à tout âge… Qui n'a pas entendu dire, par exemple, que les violeurs se recrutent dans toutes les classes sociales ? Qu'il en va de même pour les maris violents ? Sur ce dernier point, aiguillonné par une affaire aussi retentissante que tragique, et commentant à ce propos un ouvrage intitulé *Les violences envers les femmes en France* [56], le quotidien *Le Monde* écrivait récemment : « L'image […] de la femme battue dans un foyer pauvre par un mari alcoolique a vécu : la violence touche tous les milieux sociaux (8,9 % des femmes concernées sont des cadres, 3,3 % des ouvrières) [57]. » Or une saine culture statistique aurait voulu ici que l'on fît connaître au lecteur, non deux valeurs isolées, supposées emblématiques, mais *l'ensemble de la distribution des fréquences* à travers les « classes sociales », telle du moins que l'enquête de référence la fait connaître.

La résistance à la variabilité prend ainsi fréquemment la forme d'une adhésion spontanée, parfois intéressée, à ce qu'on peut appeler le *schéma de la variabilité sans loi*, où l'on suppose que « tout peut se passer », « que n'importe quoi peut arriver », « qu'on a tous les cas de figures possibles ». Paradoxalement, ce postulat semble inscrit profondément dans l'inconscient de certaines disciplines établies, et non des moindres : on peut songer ici à l'*histoire* et à la *géographie*, qui se présentent au profane comme des terres d'élection de la variation sans loi – *contre* l'hypothèse implicite de continuité à laquelle, souvent, le profane adhère spontanément en ces matières. De manière très classique sans doute, un géographe, auteur d'un ouvrage intitulé *Géographie et statistique*, écrit ainsi : « … l'objectif de la géographie […] consiste à démontrer que quelle que soit l'échelle d'observation l'espace est hétérogène, anisotrope et discontinu alors même qu'il est le plus souvent perçu comme homogène, isotrope et continu [58]. » *A priori*, il s'agit là de reconnaître la variabilité liée à *l'espace*, essentielle pour le géographe [59]. Mais, abordant plus loin la question des *« lois en géographie »*, et notant alors le faible usage dans sa discipline de la confrontation d'une distribution observée à des distributions théoriques, le même auteur note lucidement : « … les géographes demeurent souvent rétifs à la notion même de loi. L'ajustement d'une distribution observée à une distribution théorique […] implique à l'inverse l'acceptation *a priori* que des lois régissent la répartition géographique des phénomènes [60]. »

C'est donc contre les trois obstacles observés jusqu'ici – négation de la *variation*, négation de l'existence de *lois* de variation, négation de la *diversité* des lois de la variation – qu'un enseignement de la statistique doit agir. On notera à cet égard que la réalisation trop exclusive de simulations « où toutes les issues ont des chances égales d'apparaître », comme semble le prescrire le document d'accompagnement du programme de seconde, risque de conforter la propension à nier la variation réglée par une loi qui peut elle-même changer avec la grandeur considérée. Par contraste, il convient au contraire de développer très tôt la capacité à simuler des distributions *non uniformes*, point sur lequel le programme de seconde gagnerait, nous semble-t-il, à être clairement amendé.

La notion de *distribution* d'une grandeur aléatoire, la notion corrélative de *loi de variation* doivent certainement se situer au cœur de la culture statistique diffusée à l'École. De telles

[56] Jaspard 2003.
[57] Besse Desmoulières 2003.
[58] Vigneron 1997, pp. 11-12.
[59] En géographie, les « individus statistiques » sont en général des *unités spatiales* (*op. cit.*, p. 13).
[60] *Ibid.*, p. 50.

lois ont eu, au XIXe siècle, un infatigable champion : Quetelet. Prenant notamment appui sur le fait que les moyennes des distributions qu'il étudie ont une relative stabilité dans le temps, celui-ci s'efforce de montrer l'existence de véritables lois *qui s'imposent aux sociétés*. « L'" inexorable budget du crime ", prophétisé par Quetelet, note justement Desrosières [61], annonce le déterminisme sociologique d'Émile Durkheim et de ses successeurs. De même que chaque individu est doté d'une taille et d'un poids, il l'est aussi d'une " propension " à se marier, à se suicider ou à tuer autrui. » Il y a là un fait essentiel, celui de la montée d'une « juridiction sociale » face à la « juridiction » de l'État : « La " statistique ", qui, à son origine allemande du XVIIIe siècle, était synonyme d'État, s'émancipe maintenant par rapport à celui-ci, et fonde l'existence d'un *social autonome* », commente Desrosières. Les lois de l'État ne sont plus les seules ; s'y ajoutent des « lois » que la sociologie va s'efforcer d'énoncer et de faire reconnaître. On tient ici une raison d'être forte de l'étude scolaire de la statistique – non pas tant pour en devenir un utilisateur avisé en tel domaine d'activité particulier, mais pour se rendre capable d'accéder à cette connaissance « laïque » du monde naturel et social qui s'exprime dans la connaissance de ses variabilités et de leurs lois.

La position préconisée paraît être la seule possible dans le cadre de l'enseignement général, sauf à laisser le curriculum s'atomiser en une poussière de cas particuliers ou se dissoudre dans des généralités sans véritable portée. Mais son établissement rencontre des difficultés propres à la conceptualisation statistique. Pour le faire mieux voir, un ultime exemple, qui s'apparente au *paradoxe de Saint-Pétersbourg* de Nicolas Bernoulli (1695-1726), nous sera utile. Supposons qu'une personne ait le choix entre une situation $A$ (gagner *un* million de francs *sûrement*) et une situation $B$ (gagner *vingt* millions de francs avec *une chance sur dix*). La comparaison des espérances conduirait à choisir $B$, mais beaucoup de gens choisiront prudemment $A$… Pourquoi cela ? « Le calcul en termes d'espérance », commente Desrosières [62], « n'est vraiment plausible que dans une perspective fréquentiste : si ce choix (fictif !) était répété de nombreuses fois (dix fois ? vingt fois ?), il deviendrait évident de retenir $B$. » Le choix de $B$ apparaîtrait plus pertinent, de même, si le choix portait *sur des petites sommes* (1 euro sûr, ou 20 euros à une chance sur dix). L'exemple a le mérite de suggérer que, plus généralement, l'intérêt – au double sens du terme – pour une vision « fréquentiste » du monde suppose *une certaine position dans la structure sociale*, qui permette un certain point de vue sur les jeux de la vie des hommes, « celui du juge au-dessus des plaideurs, celui du banquier ou de l'assureur gérant des risques nombreux, ou même celui de l'individu isolé confronté à des micro-décisions l'engageant peu (le choix entre des petites sommes) [63] ». Cette position « en surplomb », qui est toujours relative (un enfant de dix ans préférerait sans doute recevoir sûrement 10 euros plutôt que 200 euros avec une chance sur dix…), apparaît comme un facteur clé de l'émergence d'un rapport à la variabilité conforme à la problématique statistique. S'il est vrai qu'on ne peut être pleinement citoyen sans travailler à développer en soi ce « rapport statistique » à la variabilité du monde, inversement la posture citoyenne vis-à-vis du monde est, nous semble-t-il, la seule qui puisse permettre à *tous* de se situer dans un surplomb adéquat par rapport à *toute* question digne d'être étudiée à l'École.

## REFERENCES BIBLIOGRAPHIQUES


Belorizky, É. (1998) *Probabilités et statistiques dans les sciences expérimentales*, Paris : Nathan.

Besse Desmoulières, R. (2003) Six femmes meurent chaque mois sous les coups de leurs conjoints, *Le Monde* édition du 9 août 2003.

Chevallard, Y. (2002) Organiser l'étude. 3. Écologie & régulation, in Dorier, J.-P. et al. (eds) *Actes de la 11e école d'éyé de didactique des mathématiques*, Grenoble : La Pensée sauvage, pp. 41-56.


---

[61] Desrosières 2002, p. 5.

[62] Desrosières 1993, pp. 65-66.

[63] *Ibid.*

Darmois, G. (1957) *Statistique et applications*, Paris : Armand Colin.

Desrosières, A. (1993) *La politique des grands nombres. Histoire de la raison statistique*, Paris : La Découverte.

Desrosières, A. (2002) Adolphe Quetelet, *Courrier des statistiques* **104**, 1-8.

Desrosières, A. et Thévenot, L. (1996) *Les catégories socio-professionnelles*, Paris : La Découverte.

Droesbeke, J.-J. et Tassi, P. (1990) *Histoire de la statistique*, Paris : PUF.

Jaspard, M. (ed) (2003) *Les violences envers les femmes en France. Une enquête nationale*, Paris : La Documentation française.

Klatzmann, J. (1996) *Attention, statistiques !*, Paris : La Découverte.

Le Bras, H. (2000) *Naissance de la mortalité. L'origine politique de la statistique et de la démographie*, Paris : Seuil & Gallimard.

Malliavin, P. (ed) (2000) *La statistique*, Paris : Technique & Documentation.

Maillard, R. et Millet, A. (1954) *Arithmétique*, Paris : Hachette

Morellet, A. (1822), *Mémoires sur le dix-huitième siècle et sur la Révolution*. Paris : Mercure de France ; rééd., 2000.

McGee, V. E. (1975) *Principes de statistique. Approches traditionnelle et bayésienne*, Paris : Vuibert.

Pressat, R. (1987) L'enseignement de la statistique en France à ses débuts (ca. 1850-1939), *Journal de la Société de statistique de Paris* **128(1)**, 18-29.

Reeb, G. et Fuchs, A. (1967) *Statistiques commentées*, Paris : Gauthier-Villars.

Rude, N. et Retel, O. (2000) *Statistique en psychologie*, Paris : In Press.

Schwartz, D. (1994) *Le jeu de la science et du hasard. La statistique et le vivant*, Paris : Flammarion.

Schwartz, D. et Lazar, P. (1978) *Éléments de statistique médicale et biologique*, Paris : Flammarion.

Utts, J. M. (1999) *Seeing Through Statistics*, Pacific Grove : Duxbury Press.

Vigneron, É. (1997) *Géographique et statistique*, Paris : PUF.

Volle, M. (1984) *Le métier de statisticien*, Paris : Economica.