LIVRES

Introduction to the Mori program
KENJI MATSUKI
Universitext, Springer-Verlag, 2002. 478 p. ISBN : 0-387-98465-8. 80,20 €

Le livre de Kenji Matsuki est une introduction à un domaine de la géométrie algébrique souvent perçu comme étant d'une difficulté intimidante : le Programme Minimal ou Programme de Mori. L'objet de ce programme est l'étude des variétés algébriques (projectives complexes, même si diverses généralisations sont possibles hors de ce cadre naturel), en particulier en dimension supérieure ou égale à 3. Un premier objectif est l'obtention d'une recette permettant, à partir d'une variété donnée, de sélectionner dans la même classe birationnelle une autre variété « plus jolie que les autres ». Cette variété privilégiée est appelée *fibration de Mori* ou *modèle minimal* suivant une première grande dichotomie dans la théorie. Dans un deuxième temps se pose la question d'étudier la géométrie de ces modèles privilégiés.

Depuis le début des années quatre-vingt cette théorie a été développée par de nombreux mathématiciens dont Kawamata, Mori, Reid, Kollar, Shokurov, et a connu de nombreux succès spectaculaires. En particulier en dimension 3 on dispose à présent de résultats assez complets (quelques énoncés cruciaux restant à l'état de conjecture à partir de la dimension 4).

Je discute maintenant le contenu du livre, plus ou moins chapitre par chapitre.

L'objet de la théorie de Mori étant essentiellement l'étude des variétés projectives de dimension plus grande que 3, on pourrait s'étonner que l'auteur consacre un premier long chapitre d'une centaine de pages au cas des surfaces. Pourtant le pari me semble gagné : ce chapitre permet d'intégrer en douceur les concepts de la théorie en les confrontant aux résultats « bien connus » (ou en tout cas qui le deviennent à l'étude de ce chapitre...) de la théorie des surfaces.

J'aime aussi beaucoup les deux chapitres informels en début d'ouvrage ; d'une part l'introduction joliment intitulée *The Tale of the Mori Program*, d'autre part le chapitre 3 *Overview of the Mori Program* qui trouve tout naturellement sa place après qu'ait été traité le cas de la dimension 2.

Une idée centrale dans la théorie est d'utiliser la dualité, au travers de la forme d'intersection, entre les fibrés en droite et les courbes (précisément les 1-cycles) d'une variété. Un fibré en droite joue un rôle majeur : c'est le fibré canonique, associé à une forme volume holomorphe. Ainsi l'idée centrale du programme de Mori est que les courbes qui sont d'intersection négative avec le diviseur canonique peuvent être « éliminées ». Par ailleurs dans de nombreuses situations il est utile de considérer des perturbations du diviseur canonique, c'est ce qu'on appelle travailler dans le cadre logarithmique. C'est une qualité du livre de nous convaincre qu'en effet « le cadre logarithmique est agréable et naturel », Kenji Matsuki revenant avec insistance sur ce point, dans le chapitre 2 puis tout au long du livre.

Cependant le livre n'est pas exempt de défaut : d'ailleurs un livre qui débute par une liste de notations s'étendant sur onze pages (!) m'inspire aussitôt une méfiance instinctive. Il me semble que ce détail est assez typique de la manière de rédiger de Matsuki : à force de vouloir rendre son sujet abordable il finit parfois par noyer le lecteur sous une profusion d'informations vagues et redondantes... J'ai parfois eu la sensation à la lecture du livre d'être face à l'un de ces orateurs par trop enthousiastes qui vous assène sans discontinuer de nouveaux arguments chacun plus incompréhensible que le précédent...

Je ne suis ainsi guère convaincu par le chapitre 4 concernant les singularités. Après avoir introduit de manière aride les multiples définitions nécessaires à la théorie (singularités terminales, canoniques, log-terminales, purement log-terminales...) l'auteur ne nous présente pour exemples que le cas de la dimension 2.

Il existe deux ensembles de techniques assez distinctes dans ce que l'on appelle génériquement « la théorie de Mori » : le point de vue (dû à Mori précisément) d'étudier une variété *via* les déformations des courbes qu'elle contient, en particulier à l'aide du principe fameux du « bend and break » ; et d'autre part les techniques cohomologiques qui s'appuient sur des théorèmes d'annulation. Le point de vue du livre est résolument le deuxième. Durant cinquante pages l'auteur nous amène du théorème d'annulation de Kawamata-Viehweg, qui généralise le théorème d'annulation de Kodaira, jusqu'au théorème du cône, *via* des résultats techniques : théorème « base point free » (je suis incapable de trouver un équivalent français satisfaisant), théorème de non-annulation de Shokurov, théorème de rationalité. Les résultats sont clairement motivés, et de plus l'auteur prend la peine d'énoncer des versions non-logarithmiques de ces théorèmes ce qui donne une chance au lecteur de ne pas se noyer aussitôt dans la technicité.

Si les motivations et les énoncés sont agréablement rédigés, on peut cependant souvent regretter un manque de concision dans les preuves techniques proprement dites, ainsi qu'un usage abusif du « copier-coller » qui nuit à la cohérence du texte. Un exemple typique : la figure 7.2.3, illustrant la preuve du théorème de rationalité, est manifestement une copie de la figure 1.3.3, illustrant le même théorème dans le cas des surfaces ; malheureusement l'auteur a omis de changer le S de surfaces en un X de variété de dimension quelconque... De façon générale, le livre comporte beaucoup de coquilles qui ne sont pas toutes aussi faciles à rectifier.

Les chapitres 8 à 11 sont quatre chapitres très (trop ?) courts, où l'on a la sensation que l'auteur se rend compte qu'il a presque épuisé le nombre de pages qui lui était alloué alors qu'il lui reste encore une montagne d'informations à faire passer. Du coup le chapitre 8, qui est un chapitre d'exemples, nous laisse un peu sur notre faim : même en considérant que le lecteur doit mettre la main à la pâte il me semble que certains exemples (comme ceux concernant les fibrations en surfaces del Pezzo) sont mentionnés de manière trop rapide pour pouvoir être éclairante. Le chapitre 9, sur les flips, fait l'impasse sur le problème de l'existence (ce qui est tout à fait raisonnable) et se donne pour ambition de démontrer l'unicité et la finitude. Était-ce vraiment possible en seulement sept pages ? En particulier pour la question de l'unicité, il est tout d'un coup demandé au lecteur une maîtrise certaine du « Hartshorne » qui jusque-là n'était pas aussi cruciale. Le chapitre 10 concerne un aspect de la théorie de Mori occulté dans tout le restant du livre, à savoir la technique dite du « bend and break ». L'idée générale est de produire des courbes rationnelles en déformant une courbe intersectant négativement le diviseur canonique, l'argument nécessitant un aller-retour entre caractéristique nulle et caractéristique positive. Le lecteur intéressé par cet aspect de la théorie de Mori trouvera dans [Debarre, *Higher dimensional algebraic geometry*, également paru dans la collection Universitext] une agréable introduction ; en général les livres de Debarre et Matsuki se complètent d'ailleurs très bien. Le chapitre 11 consiste en une redite des principaux résultats dans le cadre logarithmique.

Les chapitres 12 et 13 sont consacrés au problème de l'étude des produits du programme de Mori, à savoir les modèles minimaux et les fibration de Mori. Le chapitre 12 est consacré au problème de la non-unicité des modèles minimaux, ce qui contraste avec le cas de la dimension 2. Deux modèles minimaux dans une même classe birationnelle sont isomorphes en codimension 1, et on peut passer de l'un à l'autre par une suite de transformations appelées « flops ». L'auteur discute également la finitude du nombre de modèles minimaux dans une classe birationnelle donnée.

Le chapitre 13 concerne le programme de Sarkisov, qui explique comment deux fibrations de Mori contenues dans une même classe birationnelle sont reliées par une suite d'opérations élémentaires. L'auteur reprend essentiellement un article de Corti [*Factoring birational maps of threefolds after Sarkisov.* J. Algebraic Geom. 4 (1995)], apportant quelques précisions et simplifications bienvenues. Par contre l'application au théorème de Jung, qui est un sujet que je connais bien et où je pense que les idées de Sarkisov devraient être particulièrement éclairantes, me paraît rédigée d'une manière extrêmement confuse.

Le premier chapitre sur les surfaces se présentait comme une mise en bouche ; le dernier chapitre quant à lui se veut une récréation dans le monde des variétés toriques. Ce but n'est sans doute pas pleinement atteint, et il est à craindre que ce chapitre technique ne semblera une agréable pause que pour les spécialistes des variétés toriques.

Moi-même novice dans ce sujet j'ai récemment co-animé un groupe de travail dont l'objectif était de comprendre les grandes lignes de la théorie de Mori. À cette occasion le livre de Kenji Matsuki a été pour moi une porte d'entrée privilégiée dans ce domaine. Ainsi les quelques critiques que j'ai formulées ci-dessus ne sauraient entacher la principale qualité du livre, qui est de nous amener, à partir d'un bagage initial réduit au strict minimum, au cœur d'une théorie difficile et fascinante encore en pleine évolution.

Stéphane Lamy, Université de Lyon I

A Course in metric geometry, Graduate studies in mathematics

D. Burago, Y. Burago et S. Ivanov

Graduate Studies in Mathematics, 33. Amererican Mathematical Society, 2001.
415 p. ISBN : 0-8218-2129-6.44 $.

La « géométrie métrique » s'apparente bien sûr à la géométrie riemannienne. Elle remonte aux travaux de l'École russe, qui a développé avec succès une théorie des surfaces convexes non lisses.

De quoi s'agit-il ? En simplifiant à peine, c'est de la géométrie riemannienne où l'on abandonne complètement le point de vue infinitésimal. Le tenseur de courbure est un « *monstre d'algèbre linéaire* » (M. Gromov) qui rend compte de l'écart à l'ordre 2 entre une métrique riemannienne et la métrique euclidienne de l'espace tangent (sachant que pour une métrique lisse il n'y a pas d'invariants infinitésimaux d'ordre 1). Certaines propriétés ont une interprétation métrique. Tout le monde sait par exemple que le segment qui joint les milieux de deux côtés d'un triangle (euclidien !) a une longueur moitié de celle du troisième côté. Si pour les (petits) triangles géodésiques d'une variété riemannienne, cette longueur est inférieure (resp. supérieure) à celle du troisième côté, cela signifie que la courbure (plus précisément la courbure sectionnelle) est négative (resp. positive).

On trouve dans la littérature la terminologie de « CAT(0) space ». CAT est l'acronyme de Cartan-Alexandrov-Toponogov, et les espaces CAT(0) sont les espaces métriques à courbure positive ou négative au sens précédent. Les auteurs ont abandonné cette terminologie pour celle d'espaces d'Alexandrov (à bas les acronymes). Quel que soit leur nom, leur théorie est l'un des fleurons de l'École russe.

Le livre commence par des résultats de base sur les espaces métriques, suivi d'un exposé systématique sur les espaces de longueurs. Ce sont les espaces métriques pour lesquels les propriétés des distances que nous avons évoquées se prêtent à une exploitation raisonnable. Cette terminologie, popularisée par M. Gromov (cf. [6], ch. 2) remplace la terminologie tradionnelle de métrique intrinsèque ou géodésique. Une autre vertu de ces métriques est de se prêter très naturellement aux constructions usuelles de la topologie, quotients, recollements, cônes, suspensions. Vient ensuite un exposé détaillé sur les espaces d'Alexandrov à courbure inférieure ou supérieure à k (leur définition est la même que pour $k = 0$, sauf que l'on prend comme espace de référence une sphère ou un espace hyperbolique de courbure k). L'un des aboutissements de la théorie est le passage du local au global. Un espace sera dit globalement de courbure $\leq k$ ou $\geq k$ si le résultat de comparaison mentionné plus haut vaut pour *tous* les triangles géodésiques. Alors :

Un espace simplement connexe complet à courbure $\leq k \leq 0$ est globalement à courbure $\leq k \leq 0$ (Hadamard-Cartan-Gromov) ; un espace complet à courbure $\geq k$ est globalement à courbure $\geq k$ (Alesandrov-Toponogov-Perelman).

Les auteurs ont fait le choix délibéré de ne faire intervenir que très tard, au bout de 200 pages, la géométrie riemannienne. C'est confirmé, les variétés riemanniennes à courbure sectionnelle inférieure (ou supérieure) à k sont bien des espaces d'Alexandrov ! On aura déjà vu auparavant tout ce que l'on peut faire uniquement avec la distance.

Cette théorie métrique va d'ailleurs bien au-delà du riemannien. Un bel exemple d'espace d'Alexandrov est l'ensemble des points à l'infini d'un espace symétrique non compact muni de la métrique de Tits (cet exemple ne figure pas dans ce livre, suffisamment riche par ailleurs, sans doute vu l'existence de [1] et de [3] ; ce dernier livre est d'ailleurs écrit dans le même esprit). Un autre exemple est celui du graphe de Cayley de certains groupes de type fini. Ce thème de théorie géométrique des groupes (voir par exemple [5]) est par contre bien abordé dans le livre.

Mais la géométrie métrique aujourd'hui, ce n'est plus seulement la théorie des espaces d'Alexandrov, si belle soit-elle. Inspiré par la distance de Haussdorf sur les parties compactes d'un espace métrique, M. Gromov a inventé une distance sur les classes d'isométries d'espaces métriques compacts qui s'est avérée extraordinairement féconde (voir [6] ch. 5 et 8, ainsi que [7] et [4] pour des développement récents). L'exposé détaillé des fondements de la théorie qui est donné au chapitre 7 est particulièrement bienvenu. Un exemple éclairant est celui du cône asymptote d'un espace métrique (X, d). Il s'agit de la limite, si elle existe, de $(X, \lambda d)$ quand λ tend vers zéro. Si X est $\mathbf{Z}^2$ muni de la métrique des mots pour les générateurs les plus simples, on obtient $\mathbf{R}^2$ muni de la métrique de Manhattan. Cela n'a pas l'air très sérieux. Mais c'est par ce procédé, appliqué à la métrique des mots d'un groupe à croissance polynomiale, que M. Gromov a montré qu'un tel groupe est virtuellement nilpotent (cf. [8]).

Cet ouvrage se présente comme un manuel. Mais quel manuel ! Il est unique sur le marché. Partant de la définition d'un espace métrique, il nous conduit à des théories en plein développement. Il ne ménage ni les exemples, ni les motivations ou les exercices (parfois faciles, parfois ardus, toujours intéressants ; les plus faciles renouvelleraient avantageusement le stock d'exercices de topologie de licence). Il est parsemé de discussions sur des points « élémentaires » mais fondamentaux, et rarement explicités. Je pense par exemple aux ennuis causés par les points fixes d'une action de groupe, au lien entre la lissité des géodésiques d'une métrique de Finsler et la stricte convexité des boules pour la norme, ou encore aux différents volumes sur un tel espace. Des résultats importants et significatifs, parfois avec quelques simplifications techniques bienvenues, arrivent suffisamment tôt pour tenir en haleine.

Ce livre est un formidable pied à l'étrier pour aborder des théories profondes et en plein développement. Il est abordable avec profit par tout non spécialiste un tant soi peu curieux. Son étude est un impératif (catégorique ?) pour quiconque veut étudier sérieusement la géométrie riemannienne (et sans doute aussi pour qui s'intéresse à la théorie géométrique des groupes).

Références

[1] W. Ballmann, M. Gromov & V.Schroeder – *Manifolds of non positive curvature*, Progress in Maths., Birkhauser, 1985.

[2] M. Berger – « La Géométrie métrique des variétés riemanniennes », in *Élie Cartan et les mathématiques d'aujourd'hui Lyon 25–29 juin 1984*, Astérisque, Soc. Math. France, Paris, 1985, supplementary issue, p. 9–66.

[3] M. Bridson & A. Haefliger – *Metric spaces of non-positive curvature, Grundlehren der mathematischen wissenschaften*, vol. 319, Springer, 1999.

[4] S. Gallot – « Volumes, courbure de ricci et convergence des variétés (d'après T.H. colding et Cheeger-Colding) », in *Séminaire Bourbaki 1997-1998*, Astérisque, vol. 252, Soc. Math. france, 1998, p. 7–32.

[5] E. Ghys & P. de la Harpe – *Sur les groupes hyperboliques d'après Gromov*, Birkhaüser, 1990.

[6] M. Gromov – *Metric structures for Riemannian and Non-Riemannian spaces*, Birkhaüser, 1998.

[7] P. Petersen – *Riemannian Geometry*, Graduate Texts in Mathematics, vol. 171, Springer, 1998.

[8] J. Tits – « Groupes à croissance polynomiale (d'après M. gromov et al.) », in *Séminaire Bourbaki 1980-1981, exp. 572*, Lecture Notes in Mathematics, vol. 901, Springer-Verlag, 1981, p. 176–188.

Jacques Lafontaine, Université de Montpellier II

L'enseignement des sciences mathématiques – Statistique et probabilités
Sous la direction de J.-P. Kahane
Odile Jacob 2002. 284 p. ISBN : 2-7381-1138-6. 22 €

Ce compte rendu critique porte sur le contenu du chapitre 2 « Statistique et probabilités » et ses annexes, du rapport « L'enseignement des sciences mathématiques » réalisé par la commission de réflexion sur l'enseignement des mathématiques, sous la direction de Jean-Pierre Kahane. Cette partie du rapport sera désignée dans la suite par « rapport SPSM ». Certaines parties de ce texte ainsi que des annexes et des comptes rendus de sessions non publiés sont disponibles sur internet sous forme de rapports d'étape, à diverses adresses dont le site de l'IREM (Institut de recherche sur l'enseignement des mathématiques) `http://www.ac-grenoble.fr/irem/sobre/kahane.htm`. Ces documents donnent une vision plus large et explicative du travail de la commission. On pourra en particulier y consulter des textes sur le travail préparatoire de la commission, absents du rapport final, qui peuvent en éclairer la lecture. Je recommande toutefois (en particulier pour le corps du rapport SPSM) la lecture de la version finale imprimée, récemment publiée aux éditions Odile Jacob, qui a été retravaillée et remise en forme. Le style et certaines expressions, dues sans doute à la difficulté d'une synthèse pourront cependant surprendre : le lecteur en jugera par lui-même.

Vision d'ensemble du rapport

Il convient de noter dès le départ que ce rapport porte plus sur l'enseignement des statistiques que sur les probabilités. Ainsi, le travail de la sous-commission sur « la statistique et les probabilités » s'inscrit dans le prolongement du 8e rapport sur la science et la technologie, « La statistique » de l'Académie des sciences. Ce dernier qui dresse un panorama très clair de l'état de la recherche en Statistique et décrit les besoins de formation en France, y est abondamment cité. La conclusion du rapport SPSM en reprend en particulier les recommandations 2 et 9 à savoir :

– la nécessité de développer la recherche en statistique en France ;

– la nécessité de favoriser la formation initiale et continue des professeurs de lycée et des collèges, l'objectif 2) pouvant et devant contribuer à l'objectif 1).

Après avoir rappelé les buts du rapport de l'Académie des sciences, le rapport SPSM, essentiellement destiné au public des professeurs des collèges et des lycées se propose :

(1) d'une part de rendre compte des débats sur « la place des statistiques et des probabilités (de l'aléatoire) au sein des mathématiques » ; je ne suis cependant pas sûr que les multiples avis exprimés dans ce rapport rédigé par un collectif de personnes, qui y expriment des visions très différentes (complémentaires ou opposées) de la statistique soient toujours complètement transparents à des mathématiciens n'ayant pas de formation statistique. Aussi, je reviendrai dans un premier temps sur ces visions de la statistique et leur lien aux outils mathématiques.

(2) d'autre part de contribuer à l'élaboration du contenu des enseignements. Plutôt que de dresser une liste d'objets mathématiques utiles aux statistiques et à

l'aléatoire, les auteurs ont préféré illustrer la « démarche statistique » par l'exposition de très nombreux exemples, soit purement illustratifs et à visée pédagogique (voir en annexe les textes de C. Robert et J. Treinier ou Y. Escoufier) soit plus développés dans le cadre d'applications spécifiques. En particulier, les annexes proposent divers textes des membres de la commission qui donnent quelques exemples sur la manière d'aborder des problèmes statistiques concrets, dans des domaines du *credit-scoring*, de l'analyse sensorielle, la biophysique ou le traitement de données textuelles.

Le rapport met particulièrement bien en évidence l'importance prise par les « statistiques » tant dans notre vie courante que dans les domaines de recherche les plus divers, ne serait-ce que par leur rôle social et civique (voir ces fameux « sondages d'opinion » dont on se demande toujours, surtout lorsqu'on est statisticien, ce qu'il disent réellement) que par leur intervention de plus en plus importante dans les sciences physiques, naturelles ou économiques et sociales. Comme le souligne le rapport, cette omniprésence a tendance à donner de la statistique une image un peu simpliste qui peut se retourner contre elle, notamment s'il s'agit de convaincre des mathématiciens d'introduire de la « statistique » dans les programme du primaire et du secondaire, ou tout simplement d'inciter à étudier les probabilités et les statistiques à des étudiants attirés par les mathématiques et leurs applications. Les nombreux exemples donnés dans le corps du rapport SPSM ont une fonction essentiellement pédagogique. Cependant je ne peux m'empêcher de penser à sa lecture que l'utilisation trop systématique d'exemples simples, pour expliquer mieux à des non-spécialistes a le même revers, en ce qu'ils font perdre le fil de l'exposé et n'est peut être pas la meilleure façon « de guider des choix de contenus ». La dernière partie sur la formation et les acquis du secondaire laisse pour cela un peu sur sa faim. Dans une partie développée, le rapport propose pour les formations professionnelles ou supérieures non-scientifiques une approche de la statistique en contexte, la moins mathématique possible et reposant sur des considérations de bon sens, mais ne donne que peu de suggestions voire de solutions pour une véritable réforme de l'enseignement dans les grandes écoles ou le supérieur, pas plus que pour le secondaire. La question posée par Y. Escoufier dans le rapport de l'Académie de sciences « Que faut-il enseigner en statistique et à qui ? » me semble donc encore en suspens, mais ce point, il est vrai, est délicat.

La place de l'aléatoire dans les mathématiques

Comme le souligne le rapport SPSM, la statistique se situe à la fois « dans les mathématiques », dont elle utilise de très nombreux outils (la géométrie, l'analyse, le calcul, l'algèbre) tout en créant également ses propres objets mathématiques et « en dehors des mathématiques » dans le sens où son application a de très nombreux champs (la biologie, la physique, l'économie, la sociologie, et « la vie publique ») relevant de l'induction et de l'interprétation, nécessite une connaissance profonde (problématiques, concepts) du domaine d'application. C'est cette place ambiguë qui pose souvent problème dans les débats sur la place de la statistique dans les mathématiques. Si on ne peut nier cette double appartenance, s'y greffent également des questions vaines de hiérarchie implicite entre les mathématiques pures, les impures (les mathématiques appliquées) et les intouchables (les statistiques), débat très sensible dans la France cartésienne mais beaucoup moins dans les pays anglo-saxons.

Apprentissage contextualisé ou décontextualisé ?

Le rapport accorde beaucoup d'importance à un apprentissage de la statistique en contexte tout en insistant sur le fait que les statistiques doivent être enseignées par des mathématiciens. Le rapport justifie l'introduction des statistiques dans des cours de mathématiques essentiellement pour les raisons suivantes :

(i) l'omniprésence des statistiques dans la vie quotidienne et les sciences (la statistique permettant alors d'ouvrir l'horizon des mathématiciens...)

(ii) la nécessité de donner aux futurs citoyens une culture statistique minimum pour aiguiser leur sens critique face aux interprétations douteuses.

(iii) l'existence d'une « culture statistique et mathématique commune » sous-jacente, par delà les différences de modèles et d'applications.

Les points i) et ii) ne me paraissent pas entièrement convaincants et posent le problème de l'enseignement contextualisé sur lequel je reviendrai. On peut également imaginer que le point ii) puisse intéresser la philosophie et l'épistémologie ou l'éducation civique.

Seul le point iii) me semble de nature décontextualisée et donc indiscutable. En particulier, une des idées fortes du rapport SPSM est que l'introduction des probabilités dans le secondaire doit fortement contribuer à une formation à l'aléatoire et aux statistiques. Les auteurs proposent là encore plusieurs exemples pour illustrer leurs propos mais les visions différentes de la statistique qui sous-tendent ce rapport rendent le texte parfois contradictoire, les probabilités étant déclarées parfois comme indispensables et parfois inutiles, ce qui peut paraître un peu surprenant.

Pour expliquer ces apparentes contradictions, je rappellerai, comme le fait également L. Birgé lors de son intervention devant la commission, que le mot « statistique » recouvre lui-même au moins trois sens qu'il me semble fondamental de distinguer :

(a) la statistique en tant que « représentation et collecte de données ». C'est une phase complexe car elle mêle tant les définitions des objets, la constitution des catégories, les choix des nomenclatures qui vont déterminer le champ et les limites de l'analyse (la nature ontique de toute science), que la collecte, le nettoyage (le redressement, les repondérations) et le stockage de données d'expérience ou d'observations, qualitatives et/ou quantitatives.

(b) la statistique « exploratoire » ou « descriptive ». Elle travaille sur les données brutes pour essayer d'en dégager du sens, des structures, des régularités, des lois, etc.

(c) la statistique inférentielle reposant sur la notion de modèle probabiliste. Elle développe des outils mathématiques qui vont permettre de confronter un modèle scientifique et des hypothèses, aux données d'expérience ou d'observation, compte tenu de leur caractère supposé aléatoire.

Ceci explique en particulier pourquoi il existe tout un continuum de métiers de la statistique. Chacune de ces composantes de la statistique requiert des outils mathématiques spécifiques, mais contrairement à ce qu'affirme le rapport, la connaissance des probabilités me paraît indispensable dans ces trois aspects.

Les statistiques et les probabilités

En ce qui concerne a), il est rare que l'on dispose de données exhaustives pour étudier un phénomène : rappelons par exemple que même le recensement exhaustif de la population française (qui ne l'était pas tout à fait, puisque certaines catégories de personnes notamment les personnes sans domicile fixe lui échappaient) n'existe plus depuis cette année. Le premier travail du statisticien « appliqué » est de savoir comment les données sont construites (les définitions, concepts implicites ou explicites), constituées (collectées) et stockées ou de décider comment elles doivent l'être. Cette phase conceptuelle se révélera indispensable à toute interprétation postérieure. Les dérives que l'on a pu constater en analyse exploratoire viennent toujours du fait que cette étape n'est pas maîtrisée.

Les méthodes de planification (sondages, les plans d'expérience), le contrôle des biais de sélection, des censures (monnaie courante en médecine, toxicologie, économie, finance), les troncatures (très fréquentes en astronomie)[1], le contrôle des erreurs de mesure (voir dans les annexes) utilisent les probabilités. Comme le souligne plus tard le rapport, dans le chapitre sur la formation, les sondages sont minoritaires dans les enseignements d'école spécialisée comme l'ENSAE (École nationale de la statistique et de l'administration économique), mais sont surtout, en France, pratiquement absents de la recherche académique et même de la plupart des enseignements de statistiques.

Dans la confrontation des points b) et c), on entre dans ce qu'on appelle l'opposition des deux cultures statistiques, l'une qui serait probabiliste et l'autre pas, opposition que reflète d'ailleurs partiellement le rapport SPSM, dès son introduction. C'est un point qui me paraît discutable et somme toute assez secondaire vu les développements des méthodes statistiques dans les vingt dernières années. Même si l'on oublie complètement comment ont été constituées les données, les techniques exploratoires de données de très grande dimension (le fameux « data mining » ou « l'apprentissage statistique ») font appel aujourd'hui pour en comprendre le fonctionnement profond à des techniques mathématiques et probabilistes sophistiquées (inégalité de concentration, problèmes de transports de masse, théorie des processus empiriques dans des Banach non séparables, théorie de l'approximation non-linéaire). On peut aussi introduire des probabilités pour expliquer les techniques classiques d'analyse des données (analyse en composante principale, classification, etc.), on en comprend alors mieux les limites. Enfin, la statistique inférentielle c) est par nature probabiliste et utilise de nombreux outils mathématiques, souvent les mêmes d'ailleurs que ceux cités précédemment pour b).

---

[1] On parle de données censurées lorsque les données ne peuvent être observées entièrement, soit que le phénomène n'ait pu être observé dans son intégralité sur une période de temps (par exemple l'effet d'un traitement qui serait interrompu par le decès du patient), soit qu'il y ait des contraintes techniques empêchant cette observation (par exemple en toxicologie, il existe des limites, dépendant du matériel, des conditions de l'expérience, en dessous desquelles, il n'est plus possible de déterminer la quantité d'une substance chimique dans un corps). La censure (ici le temps, là la limite de détection) apporte néanmoins de l'information dont il faut tenir compte : éliminer les observations incomplètes peut s'avérer catastrophique pour l'interprétation. Dans la troncature, une partie de l'information est irrémédiablement perdue (par exemple si l'on envoie sur une étoile de nombreux signaux, et que seulement une petite partie de ces signaux revient). Les problèmes de censures, de données incomplètes et leur prise en compte pratique ont donné lieu à d'importantes recherches méthodologiques dans les vingt dernières années.

Le véritable dénominateur des différentes formes de la statistique me semble donc bien être les probabilités. Introduire et développer l'enseignement des probabilités dès le secondaire, c'est déjà faire un grand pas dans la formation à l'aléatoire et ne peut que fortement contribuer au développement de la statistique. Par ailleurs, les statistiques font appel à un bagage mathématique de base. L'exemple 3 donné page 58 sur l'importance de la vision géométrique me semble en particulier important, parce qu'il est aussi commun à b) et c) (et même à a), les techniques de redressement pouvant s'interpréter comme des méthodes de projections). Les notions d'espace vectoriel, d'espace de Hilbert, de projections orthogonales ou le théorème de Pythagore sont indispensables pour faire comprendre la statistique sous toutes ses formes, l'analyse exploratoire, la régression, les séries temporelles, l'estimation semi-paramétrique, les problèmes inverses etc... Cependant, contrairement à ce que conclut ce paragraphe page 58 (en contradiction flagrante avec son introduction), je pense que ce bagage fondamental, que peut seul apporter un enseignement solide des mathématiques en secondaire, est indispensable à tout statisticien théoricien ou appliqué et me semble dans une certaine mesure plus important que l'introduction des histogrammes ou des boîtes à moustaches dans le secondaire.

Statistiques et champs disciplinaires

Comme le souligne le rapport, l'apprentissage de l'informatique et de la simulation peuvent également contribuer à familiariser les étudiants avec l'aléatoire. Une formation initiale à la statistique par « l'expérimentation informatique » qui ne nécessite finalement que peu de bagage théorique pourrait se faire dès le collège. Le chapitre sur la formation dans le secondaire propose quelques pistes pour relier l'enseignement des statistiques aux vécus des étudiants. Dans cet apprentissage de la statistique, l'informatique est un outil remarquable qui permet de voir et d'entendre, souvent sous une forme ludique. Au-delà de l'apprentissage, notons par ailleurs que les modes de stockage optimaux (hiérarchiques ou non) qui dépendent de la taille des données observées sont généralement du ressort de l'informatique (avec ses propres outils mathématiques) mais peuvent également avoir un impact non négligeable sur le choix des méthodes statistiques à mettre en œuvre, en raison des contraintes qu'elles imposent (traitement en temps réel ou pas par exemple). On lira à ce sujet le chapitre « Statistique et informatique : la nouvelle convergence » du rapport de l'Académie des sciences.

Cependant, il y a là une dérive possible, qui est patente dans les réformes des enseignements actuels. Vouloir faire des statistiques (ou des mathématiques) « une science expérimentale » sous le prétexte que les statistiques et les mathématiques sont utiles aux sciences expérimentales, c'est tout confondre. L'outil informatique peut aussi donner des intuitions fausses et brider l'imagination (Einstein n'avait pas d'ordinateur et si le modèle de la relativité générale n'a pas été construit en aveugle, il a fallu un certain temps avant d'avoir des données qui permettent de le tester... Cela pour rappeler qu'on peut aussi construire des modèles peu intuitifs et extrêmement porteurs sans données ni simulations). Le rapport insiste à juste titre sur le fait que les statistiques utilisent l'ordinateur mais que l'informatique n'est qu'un outil privilégié ne pouvant en aucun cas se substituer à la compréhension des concepts sous-jacents. Je doute même, comme il est dit page 66, qu'il puisse permettre « d'appréhender la nature des preuves statistiques ». Travaillant depuis de nombreuses années sur une méthode, « le bootstrap » qui utilise l'ordinateur de manière intensive, je peux en témoigner et montrer à ceux qui en douteraient encore, que la preuve statistique et mathématique peut être en totale contradiction avec ce que montre l'ordinateur.

Le rapport insiste, à juste titre, sur ce que peut avoir d'enrichissant le contact entre les mathématiques et les autres sciences. Les exemples retenus bien qu'intéressants et révélateurs, me paraissent souvent peu attrayants. Si le but est de sensibiliser des adolescents avec la statistique et les probabilités, pourquoi ne pas plutôt évoquer les problèmes statistiques liés au génome, aux interrogations de sites web et aux files d'attente, à l'image ou au traitement du signal (déconvolution, codage, compression, reconnaissance de forme) ou l'évaluation et la gestion des risques (environnementaux — OGM —, alimentaires, financiers), qui ont actuellement des répercutions médicales, techniques et parfois éthiques considérables ?

Dans l'intéressant chapitre sur la géométrie, la commission souligne l'importance des arts graphiques dans la perception géométrique. On peut en statistique et en probabilités, en trouver une analogie dans celui de la musique, art de la combinatoire s'il en est ou la poésie (cf. Mille Milliards de poèmes de Queneau et l'Oulipo). De nombreux musiciens, depuis Mozart (voir les fameux menuets K294d qui utilisent des matrices de transition et d'anachroniques chaînes de Markov) jusqu'aux compositeurs contemporains, J. Cage ou Xenaxis, ont eu recours à l'aléatoire dans leurs œuvres. Expliquer comment les probabilités peuvent également jouer un rôle dans les arts, comment des techniques de « sampling », de nettoyage de bande son, de codage en mp3 ou en divx sont directement liés à des problèmes de probabilités et statistiques, peut aussi susciter des discussions, des intérêts ou des vocations... et donner une image de la statistique plus vivante et contemporaine que « des pourcentages », « les colonnes de chiffres » ou des régressions douteuses[2]. Vouloir donner un sens critique aux citoyens vis-à-vis des statistiques et de toute science me semble important mais convaincre de l'utilité et de l'importance de ces techniques l'est au moins autant.

Formation à la modélisation ?

Plus que les techniques probabilistes et mathématiques, le point sans doute le plus difficile de la statistique et celui qui, à mon avis, pose le plus de problèmes au niveau de l'enseignement est celui de la *modélisation*. L'acte de modéliser est un acte qui requiert une connaissance profonde des phénomènes que l'on observe et nécessite de l'imagination. Il est difficile justement parce qu'il introduit la notion de la pertinence et de potentialité pour le domaine d'application, qui sont des notions

[2] Le mot « régression » vient des prétentions eugénistes de M. Galton, qui voulait montrer que la taille de la « race ouvrière » décroît de génération en génération et est vouée à la régression. Si le but est de démontrer qu'on peut faire n'importe quoi en statistique (comme dans toute science), cet exemple historique me semble plus intéressant. Ce n'est pas la technique qui est en cause dans l'exemple page 57 mais bien la pertinence de l'interprétation.

subjectives et non plus mathématiques. Je pense que seul un enseignement contextualisé à un niveau supérieur peut permettre de réellement développer cette faculté, ne serait-ce que parce qu'elle nécessite de nombreux acquis dans des domaines très différents et un investissement total du champ d'application. Le rapport n'apporte malheureusement pas vraiment de solutions pour développer cette approche dans les cycles supérieurs et n'aborde que peu le problème de la modélisation, ce qui était déjà le cas du rapport de l'Académie des sciences.

Si l'on veut développer dans le secondaire cette faculté de modélisation, des exposés d'applications spécifiques en situation et des enseignements spécifiques peuvent y contribuer. La collaboration entre enseignants mathématiques et enseignants d'autres disciplines (physique, biologie, génétique, économie, philosophie) sous forme de cours communs serait sans doute encore plus intéressante mais actuellement difficilement envisageable au niveau national par la refonte totale du système qu'elle impliquerait. De telles collaborations constitueraient sans doute un exercice délicat, mais très formateur pour les élèves. De telles tentatives existent déjà dans les pays anglo-saxons, alors pourquoi cette perspective est-elle rarement évoquée en France ? Peut-être n'est-ce encore qu'une question de moyens accordés à l'éducation et aux enseignants ? Qui peut en France se charger de tels enseignements ? Le rapport évoque dans sa dernière partie la formation des professeurs impliqués par l'introduction des probabilités et des statistiques et évoque la formation des 60 000 professeurs, toutes matières confondues. On se demande dans le contexte actuel de pénurie d'enseignants et de chercheurs et même de professionnels de la statistique en France, qui pourrait bien enseigner à 60 000 personnes... Les statistiques et les probabilités ne s'enseignent pas en quinze jours. Si l'on ne commence pas d'abord à former des formateurs qui puissent assurer ce rôle, ce qui prendra, même avec des moyens suffisants un certain temps, il restera toujours l'autoformation...

Patrice Bertail, CREST-Laboratoire de Statistiques