

## Un événement : le rapport de l'Académie des Sciences sur la statistique.

«La Statistique»

Académie des Sciences,  
Rapport sur la Science et la Technologie n°8  
Juillet 2000, Technique & Documentation,  
ISBN : 2-7430-0411-8

L'Académie des Sciences vient de consacrer un rapport de plus de 180 pages à la statistique, dans le cadre du rapport biennal sur l'état de la science et de la technologie demandé par le conseil interministériel du 15 juillet 1998. Pour notre discipline, il s'agit d'un événement important pour maintes raisons : la nouveauté, le prestige de l'institution, la qualité et la diversité des contributeurs, le diagnostic porté sur l'état de la discipline.

Ce rapport est l'œuvre d'un groupe de travail animé par Paul Malliavin et réunissant les membres suivants : Paul Deheuvels, Yves Escoufier, Edmond Malinvaud, Bernard Picinbono, Daniel Schwartz, Alain-Jacques Valleron et Marc Yor. Y ont également collaboré à divers titres : Henri Caussinus, Paul Embrechts, Nelly Hanoun, Nicole El Karoui, François Marcotorchino. Parmi les commentateurs d'un groupe de lecture critique, on notera les contributions de nos collègues Philippe Besse (au nom de la SFdS), Jean-Luc Besson, Dominique Collombier, Alain Monfort et Elisabeth de Turckheim.

De façon quelque peu inhabituelle, le rapport débute par neuf constats et recommandations, ce qui lui donne un aspect immédiatement offensif. En effet l'Académie des Sciences formule d'emblée le constat suivant : « *La France apparaît en retard dans cette discipline où historiquement elle a joué un rôle majeur... Aujourd'hui, il n'est pas exagéré de considérer la statistique comme une discipline émergente difficilement...* ». L'Académie dénonce « *l'absence de formation en statistique dans les collèges, les lycées et de vastes secteurs de l'enseignement supérieur qui entrave le développement économique et l'exercice des droits des citoyens* » et déclare « *un effort majeur est nécessaire. C'est une nécessité nationale urgente* ».

A l'appui de ce constat le rapport s'articule en quatre chapitres : présentation de la discipline, recherche méthodologique, interactions avec les autres disciplines, formations et métiers.

Le premier chapitre commence par le rappel indispensable de la distinction entre *la* statistique et *les* statistiques, c'est-à-dire entre d'une part la statistique méthodologique, et d'autre part les données et résultats. Tout statisticien appréciera la nuance, et on sait bien que ceux qui ne la font pas ne connaissent pas vraiment la statistique : à ce propos, on ne peut être que consterné de voir le nombre d'intitulés de cours, d'ouvrages et mêmes de diplômes parlant de statistiques avec un s, ce qui prouve que la statistique est trop souvent en-

seignée par des non-statisticiens ! Le chapitre 1 brosse ensuite un tableau des liens entre statistique, mathématiques et probabilités, et insiste sur les développements récents liés à l'usage de l'informatique, avec ce que l'on appelle « data mining ». L'emploi de titres comme « statistique sans probabilité » et « statistique et informatique : la nouvelle convergence » donne bien l'esprit de certains développements et insiste sur le fait qu'il ne s'agit pas d'une mode qui va passer mais d'une tendance lourde. Les conclusions de la dernière partie de ce chapitre insistent sur la nécessité de recherches coordonnées en statistique et informatique, avec l'adaptation des cursus universitaires et la création d'un Institut National de Sciences et Techniques Statistiques ; et elles évoquent, si rien n'est fait, la menace d'une domination encore plus forte des États-Unis. Si l'on peut juger que certaines parties sur la gestion de la relation-client sont trop détaillées, et d'autres sur les sondages un peu expéditives, ce chapitre pose clairement les problèmes de l'avenir de la statistique au sein de la société de l'information. J'ajouterai pour ma part, en suivant J. Friedman (1997), que la statistique a laissé se développer en dehors d'elle, et a largement ignoré, des pans entiers de la recherche appliquée actuelle, ainsi développés en dehors de la statistique conçue en une acception trop stricte. Voici ce point de vue quelque peu ironique de Friedman : « *One view recognizes that while the amount of data (and related applications) continue to grow exponentially, the number of statisticians is not growing that fast. Therefore our field should concentrate that small part of information science that we do best, namely probabilistic inference based on mathematics. This is a highly defensible point of view that may well turn out to be the best strategy for our field. However, if adopted, we should become resigned to the fact that the roll of Statistics as a player in the information revolution will steadily diminish over time. This strategy has the strong advantage that it requires relatively little change to our current practice and academic programs* ». L'autre stratégie consisterait à : « *First (and foremost) we would have to make peace with computing. It is here to stay ; that's where the data is. Computing has been one of the most glaring omissions in the set of tools that have so far defined Statistics. Had we incorporated computing methodology from its inception as a fundamental statistical tool (as opposed to simply a convenient way to apply our existing tools) many of the other data related fields would not have needed to exist. They would have been part of our field* » et de citer : la reconnaissance des formes, les SGBD, les réseaux de neurones, l'apprentissage automatique, les modèles graphiques, la chimiométrie, le traitement du signal...

Le chapitre 2 du rapport est consacré à un état de la recherche méthodologique en statistique en France avec deux contributions. Celle de Paul Deheuvels s'attache à montrer tout d'abord, à partir de comparaisons chiffrées, une exception française en ce qui concerne la population des statisticiens se déclarant tels : si les statisticiens français membres de l'IIS forment le deuxième groupe national après les USA avec 221 membres, on n'y trouve qu'un tiers d'enseignants et chercheurs et deux tiers de « statisticiens officiels » au contraire des autres « grands pays scientifiques », et le réservoir formé par les membres des sections de l'IIS (Société Bernoulli, IASC, etc.) est quasi nul. On peut l'expliquer par la faible appétence des collègues français à adhérer à

des associations, et considérer que cette situation est déplorable au pays de la loi de 1901 ! La faiblesse de cette participation a des conséquences nuisibles sur la représentativité des associations et l'audience de leurs membres. Paul Deheuvels analyse ensuite la production scientifique française en statistique, à partir des travaux de Christian Genest, et tire un constat inquiétant : si la France est au deuxième rang mondial en probabilités, selon le critère du nombre de pages publiées, et au troisième en nombre de pages par habitant, elle n'est qu'au sixième rang et au vingtième rang respectivement pour les mêmes critères, en statistique. De plus le nombre de chercheurs productifs est faible. Le rapport avance que « le nombre de chercheurs en statistique en France devrait être multiplié par un facteur minimal de 5 pour que le niveau d'activité soit comparable à celui des États-Unis ». Le rapport propose quelques explications au faible nombre de chercheurs actifs en France : tout d'abord la relative pénurie des formations de statisticiens professionnels et l'appel du marché du travail ont pour résultat que les meilleurs étudiants se dirigent vers des emplois dans les entreprises et les administrations plutôt que vers la recherche et l'enseignement. Dominique Collombier (p. 161) y ajoute la faible attractivité des carrières universitaires. Je serais tenté de rajouter que le manque d'enseignants de statistique entraîne que ceux-ci sont très sollicités pour intervenir dans de nombreuses formations, au détriment de leur activité de recherche. Yves Escoufier insiste d'ailleurs dans la deuxième partie du chapitre 2 sur la dispersion des statisticiens et le manque de lisibilité de la discipline (13 DEA seulement la concernent, en comptant large, souvent intitulés sans référence explicite à la statistique).

Le chapitre 3 est un panorama des principaux domaines d'application de la statistique montrant bien les enjeux et les apports de la statistique à l'agronomie, la médecine, l'économie, la finance, l'assurance et le traitement du signal. Chaque partie est écrite par un connaisseur et acteur du domaine et constitue une introduction aux problématiques. On note au cours de la lecture le manque de spécialistes dans certains domaines (génomique, actuariat).

Le chapitre 4 est consacré à la formation sous ses deux aspects : formation de spécialistes et formation à la statistique. Les deux premières contributions (Yves Escoufier et Daniel Schwartz) sont majoritairement consacrées à la formation du citoyen et des non-statisticiens en se référant à des travaux et expériences français et étrangers, que connaissent bien les membres de la SFdS (voir les diverses tables rondes pendant les journées annuelles). Ici l' enjeu est plus de former à la pensée statistique et au bon usage des méthodes que de s'appesantir sur les fondements mathématiques. Les deux auteurs insistent sur le nécessaire côté expérimental et l'utilisation de situations concrètes. Une dernière partie rédigée par Nelly Hanoune brosse un tableau, forcément rapide, des divers métiers de la statistique pour présenter les formations professionnelles : celles-ci sont diverses et existent à tous niveaux de Bac+2 à Bac+5. Le texte insiste sur le couplage nécessaire avec une bonne formation en informatique, pour conclure qu'actuellement les diplômés sont embauchés très rapidement, et que le développement de doubles compétences, très recherchées, se fait tant en formation initiale qu'en formation continue. On peut regretter

que ce chapitre ne traite pas de la formation des ingénieurs dans les grandes écoles, où les situations sont très diverses et pas toujours satisfaisantes.

Quelles sont donc les neuf recommandations de l'Académie des Sciences ?

Elles sont d'ordre général : « faire un effort pour rattraper le retard de développement de la France », « mettre en œuvre une réflexion concertée entre entreprises et établissements d'enseignement supérieur », « faire un effort de formation des professeurs des lycées et collèges », mais portent également sur des points précis : « constituer des structures de recherche et de formation doctorale », « recruter et former des enseignants et des chercheurs », « afficher clairement des recrutements de statisticiens dans les universités », « tenir compte des collaborations dans l'évaluation des chercheurs et enseignants-chercheurs », « mener des recherches coordonnées en statistique et informatique », « développer la recherche en actuariat avec les organismes professionnels pour aboutir à des thèses de doctorat ».

Au delà de ces recommandations, que certains pourraient juger trop vagues, il est clair que le constat (souvent provocant) de l'Académie des Sciences ne peut qu'être bénéfique pour le développement de la statistique en France. Il est clairement montré que la statistique est une discipline scientifique autonome d'intérêt fondamental, qui utilise, mais sans s'y réduire, mathématiques, probabilités, informatique. On doit donc se réjouir de cette publication qui devrait inciter autorités de tutelle, collègues d'autres disciplines, entreprises, à aider les statisticiens afin de mettre en place les réformes et de donner à la statistique toute la place qui convient. L'Académie des Sciences apporte en effet une caution et un soutien à des arguments trop souvent considérés comme *pro domo*.

Souhaitons donc que ce rapport soit lu par le plus grand nombre, statisticiens et non statisticiens.

Gilbert Saporta

## RÉFÉRENCES BIBLIOGRAPHIQUES

- FRIEDMAN, J. (1997). *Data Mining and statistics, what's the connection ?*  
<http://www-stat.stanford.edu/~jhf/ftp/dm-stat.ps>
- GENEST, C. (1997). *Statistics on statistics : Worldwide performance based on journal publication over the period 1985-1995*. The Canadian Journal of Statistics, 25, 427-443.
- GENEST, C. (1999). *Probability and statistics : A tale of two worlds ?* The Canadian Journal of Statistics, 27, 421-444.