

## ANALYSE BIBLIOGRAPHIQUE

Brigitte Le Roux and Henry Rouanet, *Geometric Data Analysis From Correspondence Analysis To Structured Data Analysis*, Dordrecht, Boston, London, Kluwer Academic Publishers, 2004.

L'analyse des données à la française a mauvaise presse en France, surtout chez les mathématiciens appliqués, c'est un lieu commun que de le rappeler, de sorte que ce livre a peu de chances d'être reçu, encore moins d'être lu, mais il ne fait guère de doute qu'il restera comme un livre témoin d'une époque, une synthèse très remarquable entre plusieurs des directions les plus innovantes de la statistique, et qu'on le redécouvrira d'ici quelques décennies et qu'on le commentera sagement dans les séminaires parisiens.

Le titre en égarera plus d'un, mais, qu'on ne s'y trompe pas, il s'agit d'un discours de la méthode statistique. Un parmi d'autres sans doute, mais un véritable, il n'y a pas l'ombre d'un doute à cet égard. Et comme la chose est finalement assez rare, il vaut la peine d'y aller voir.

De quoi s'agit-il ? De statistique évidemment, au sens classique ou historique du terme, cette « nouvelle science, qui s'est constituée peu à peu pour prendre rang dans les académies au début du XIX<sup>e</sup> siècle, c'est-à-dire comme le rappelle l'Académie des sciences de Paris en 1818 « une science de plus, formée d'un grand nombre de résultats positifs fidèlement représentés ». Depuis la même époque, à l'initiative de Fourier et de quelques autres, les « résultats positifs » sont présentés en « tableaux ». Donc, la statistique est la science des tableaux, dans tous les sens du mot, avec ce qu'il comporte de pictural et de descriptif et en même temps de représentation, d'explication, de conjecture et de détermination.

Ces tableaux, nous disent Brigitte le Roux et Henry Rouanet, on les retrouve sous des formes variées dans la statistique du XX<sup>e</sup> siècle, tableaux de contingence, tables des réponses individuelles à un questionnaire, caractéristiques numériques d'individus de tout genre, ... Les tableaux ne manquent pas, que l'on doit théoriser et étudier le plus fidèlement possible selon les vœux de l'Académie parisienne. Et c'est bien de cela qu'il s'agit.

Pour les auteurs, la première étape consiste à isoler un certain nombre de « variables actives » qui permettront de définir l'espace géométrique des données. Cet espace est intrinsèque aux données, on peut naturellement l'imaginer comme un nuage de points dans un espace de forte dimension, suivant l'usage, mais il faut également concevoir cet espace comme un espace géographique avec ses proximités, ses frontières naturelles, ses bassins, ses massifs montagneux et bien sûr ses cyclones, ses anticyclones, ses catastrophes. On ne peut se représenter l'espace des données que par le moyen de cartes. Les techniques désormais classiques de l'analyse spectrale sont là pour cela, et les auteurs nous les exposent dans leur nudité et leur généralité algébriques, suivant en cela l'exemple du maître Benzécri, et dans leurs réalisations informatiques les plus puissantes et les plus récentes. D'où l'on tire les interprétations des axes et les conclusions de l'analyse des données telles qu'on la pratique depuis les années soixante-dix. Mais, ce n'est que la seconde étape. Il y en a deux autres encore.

La troisième étape est peut-être la plus originale, elle consiste à «Structurer» l'espace géométrique des données, à l'aide de ce que les auteurs appellent des «Facteurs structurants». En général, il s'agira de variables supplémentaires non intégrées à la définition de l'espace et qu'on représente avec les mêmes coordonnées, sur les mêmes cartes. Cela revient à colorer l'espace, à tracer sur les cartes des frontières artificielles qui se marient plus ou moins harmonieusement avec les frontières naturelles des données, comme sur une carte géographique où les couleurs des différents pays suivent, ou ne suivent pas, le cours des fleuves et les lignes de crête. On peut alors déduire quantité d'informations supplémentaires sur l'espace ainsi structuré. On comprend assez que ce concept de structuration de l'espace s'apparente à des notions plus anciennes de l'analyse de la variance de Fisher ou de la sociologie de Bourdieu, mais l'usage qu'en font les auteurs est original et en tout point fascinant. Il suffit pour s'en convaincre de se reporter au chapitre IX qui propose des analyses de cas, où l'on voit en action les concepts fondateurs de l'analyse géométrique des données structurées. Prenons le second de ces exemples, l'analyse des élections législatives françaises de 1997. Le facteur structurant est ici le choix électoral, du Front national au Parti communiste, du blanc au rouge. Les variables actives qui définissent l'espace électoral sont diverses et variées et rendent compte des différences socioculturelles ou économiques de la population électorale française. On obtient ainsi des cartes de l'espace électoral avec leurs topographies propres, que l'on colorie avec le choix électoral. D'où l'on peut déduire quantité d'informations sur les réalités françaises, cette inadéquation presque totale des couleurs et des catégories politiques, le clivage droite/gauche ne se retrouvant guère sur les cartes électorales de la France. Nous pourrions reprendre aussi bien l'espace des surdoués de Stanford ou celui des parkinsoniens, mais nous ne voulons pas déflorer ce chapitre superbe.

Il reste une dernière étape, l'étape inférentielle. La population étudiée, son espace géométrique, sont des échantillons d'une population plus nombreuse sur laquelle on souhaite valider des hypothèses ou estimer des paramètres. Les auteurs montrent fort bien que l'analyse des données structurées n'est nullement incompatible avec les analyses inférentielles traditionnelles, tests de signification, intervalles de confiance, procédures bayésiennes, ces techniques s'insérant naturellement dans la méthode générale de la statistique proposée ici. On pourra contester tel ou tel point de celle-ci, mais sa cohérence et sa richesse ne laissent aucun doute. *Geometric Data Analysis* est un grand et beau livre qui fait honneur à ses auteurs.

Ajoutons que chaque technique présentée est accompagnée d'une justification théorique, d'une méthodologie complète et limpide d'application» préparation des données, structuration, interprétation, qu'on trouve rarement explicitée dans des ouvrages de mathématiques appliquées. Il y a là une somme d'expériences et de réflexions théoriques, enrichies par une pratique constante des sciences humaines et de leurs tableaux. Certains paragraphes poussent d'ailleurs plus loin encore l'analyse de la méthode statistique, et proposent des réflexions historiques et sociologiques de grande qualité sur la statistique anglo-saxonne et l'analyse des données française, leurs grandeurs et leurs petites. Si ce livre pouvait convaincre ici ou là que l'analyse géométrique et inférentielle des données structurées fait partie intégrante de la statistique mathématique, celle de Laplace et de Fisher, il aurait atteint son but.

Bernard Bru