

STATISTIQUE ET ANALYSE DES DONNÉES

JEAN-MARIE LEGAY

Pour une biométrie

Statistique et analyse des données, tome 1, n° 2 (1976), p. 5-11

http://www.numdam.org/item?id=SAD_1976__1_2_5_0

© Association pour la statistique et ses utilisations, 1976, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

par Jean-Marie LEGAY (x)

Notre objectif premier n'est pas de créer une discipline nouvelle en mathématique ou en biologie, ni même une discipline apparemment mixte en biomathématique (1); il est d'abord de participer à l'élaboration d'une méthodologie nouvelle à la disposition des sciences expérimentales, c'est à dire dans les sciences expérimentales.

Notre objectif premier concerne l'articulation de la biologie, de la mathématique et de la logique de l'expérimentation et par conséquent notre situation est réellement, fondamentalement pluridisciplinaire. Notre position est d'accepter les problèmes tels qu'ils sont posés par la pratique de la biologie, de la médecine, de l'agronomie et à la limite de l'économie (dans la mesure où elle n'est pas absente de l'expression même des problèmes posés par les disciplines précédentes), et de tenter d'y répondre par une méthodologie qui fait appel aux techniques mathématiques en général, c'est à dire appartenant à toutes les branches des mathématiques.

Notre souci est de mettre en place la suite des raisonnements, des interventions expérimentales, des techniques de traitement des données et des traitements d'interprétation nécessaires à la résolution de ces problèmes. Et dans cette suite nous savons déjà que la place des modèles en tant qu'instrument de recherche (2) se révèle prépondérante, puisque c'est par leur intermédiaire que s'effectue pour la plus grande part la jonction entre biologie et mathématique; c'est la raison pour laquelle nous accordons à la modélisation une place privilégiée dans notre travail depuis plusieurs années.

Il peut se faire que secondairement nous participions à l'élaboration de théories, c'est à dire de systèmes cohérents de faits; il pourrait se faire que nous participions à la mise en place de branches nouvelles de certaines disciplines comme celle d'une biologie théorique (3); sa réflexion est nécessaire, dans le développement de la recherche, à la formulation des bonnes questions. Il est clair que nous ne nous interdisons pas non plus de trouver nous-mêmes à la suite de notre propre recherche des résultats nouveaux (spécialement dans certains domaines de la biologie moléculaire et cellulaire ou de l'écologie) et en mathématique (spécialement dans certains domaines de la statistique ou de l'informatique). Bien au contraire, car nous pensons que la mise au point de méthodes nouvelles ne se fait pas, ne peut pas se faire dans l'abstrait, mais bien à l'occasion de problèmes réels dans toute leur complexité et parfois même toute l'imprécision de leur formulation originelle.

Là où nous montrons que notre activité est fondamentalement méthodologique, c'est lorsque les données ne présentaient aucune difficulté d'acquisition ou lorsqu'elles étaient déjà à la disposition de tout le monde dans la littérature, mais n'étaient pas exploitées. C'est peut-être un des faits les plus frappants de la biologie que cet énorme gaspillage de faits, des faits acquis mais non publiés, ou publiés sous une forme inutilisable (par exemple quand les données brutes sont remplacées par les conclusions d'un mauvais traitement statistique), ou encore publiés sous une forme qui n'épuise pas toute l'information qu'elles contenaient (et qui par conséquent conduit à des interprétations au moins partiellement fausses).

x

x x

Les deux niveaux de la biologie dont le développement a été retardé pour des raisons à la fois techniques et conceptuelles sont le niveau moléculaire et le niveau écologique. Il pourrait bien se faire que ces deux niveaux apparaissent de plus en plus déterminants : le niveau moléculaire parce que c'est bien à ce niveau que fonctionne une cellule et même un organisme quel que soit le degré d'intégration auquel on s'intéresse; le niveau écologique parce qu'un être vivant ne vit jamais seul et que ce qui fonctionne en définitive à la surface de cette terre ce sont bien les écosystèmes. C'est à ces deux niveaux que se situent les articulations décisives avec les facteurs physico-chimiques du milieu. Ce n'est donc pas par hasard que les progrès les plus intéressants dont nous pouvons faire état relèvent de ces deux approches et que nous ayons pu participer à l'une comme à l'autre. Dans ces deux cas, on peut définir des structures éminemment mathématisables.

Au niveau moléculaire : non seulement les acides nucléiques, les protéines, les macromolécules d'une façon générale ont une structure remarquable, mais les relations entre ces molécules ou entre des fractions de ces molécules sont étroitement organisées.

Au niveau écologique, chaque population obéit à des lois de ~~structure~~ et de dynamique, mais en outre des relations entre ces populations à l'intérieur d'un système existent et évoluent de façon précise.

Nous n'avons pas les mêmes facilités, au moins pour le moment, au niveau cellulaire, faute de données systématiques. Il est presque incroyable de constater que nous ne possédons les quatre renseignements suivants : nombre d'ovules formés, nombre de spermatozoïdes, taille des ovules, taille des spermatozoïdes, que pour moins d'une vingtaine d'espèces animales ! Il s'agit pourtant de lignées cellulaires de première importance.

Mais il n'est pas venu à l'esprit des biologistes qu'il pourrait être utile de connaître simultanément pour un certain nombre d'espèces ces quatre informations. Par contre si l'on s'intéresse seulement à l'une d'entre elles, la taille des ovules par exemple, alors le nombre des espèces pour lesquelles on a cette mesure augmente considérablement. Si l'on devient plus exigeant et qu'on demande simultanément les deux dimensions principales de ces oeufs (longueur et largeur), alors le nombre d'espèces pour lesquelles on a ces deux mesures à la fois diminue de nouveau. Ainsi je les ai trouvées pour un peu plus de 350 espèces d'insectes sur les 800.000 espèces connues et soi-disant décrites dans cette classe d'animaux (4).

En fait toujours au niveau cellulaire, en dehors du cas particulier des gamètes, on manquait d'une bonne définition de "population cellulaire"; on vient de progresser dans ce domaine grâce aux recherches en biologie du développement et aux notions de compartiment, de clone et de polyclone (5).

De la même manière, au niveau organismique (6), nous ne possédons le poids de six organes chez les Vertébrés que pour quelques dizaines d'espèce dont seulement six espèces d'Amphibiens et de Reptiles; nous ne connaissons semble-t-il la fréquence des battements cardiaques et le poids du coeur que pour 23 espèces d'Oiseaux ! etc...

Je conclus pour ma part que la description quantitative systématique même la plus élémentaire, du monde biologique est à peine commencée; quant à une description que j'appellerais multivariée, elle n'est pas envisagée comme objectif de recherche, les renseignements que l'on possède sont anarchiques, proviennent de quelques individus seulement et sont rarement biologiquement simultanés; on peut dire que lorsqu'on les a, c'est par accident.

Chez l'homme cependant, où l'on manque aussi de données, on accumule actuellement sur une base clinique des données biochimiques simultanées; il serait évidemment intéressant de les organiser et de pouvoir les comparer à celles obtenues sur d'autres espèces. Les questions cruciales qui se posent concernent l'organisation et le stockage de ces données.

Or c'est précisément le recensement et l'étude des contraintes de structure et de fonctionnement, et par suite d'évolution, aux quatre niveaux du monde vivant (je veux dire moléculaire, cellulaire, organismique et écologique) qui devraient constituer aujourd'hui des objectifs prioritaires de recherche; dès que possible et en liaison immédiate, l'articulation entre ces différents niveaux devrait être abordée. On peut sans doute prédire que c'est de la maîtrise de ces articulations que naîtront les résultats biologiques importants de demain.

X

X X

En définitive, c'est le constat et l'étude de cette organisation du monde vivant, mais aussi sa mobilité, sa malléabilité qui domine notre façon de voir la réalité. Il y a donc lieu d'en arriver à une redéfinition de la biométrie, qui passe en France pour être officiellement (7) l'étude des variations biologiques, ou pour être une application de la statistique comme le pensent les anglo-saxons. La biométrie est engagée dans le monde réel, elle a ses buts, ses intérêts, des rapports déterminés avec la réalité. Elle n'est pas une théorisation de la réalité, ce qui ne signifie pas que parmi ses moyens il n'y ait pas cela aussi.

Pour nous, la biométrie est une science des données organisées et l'organisation en question est celle des êtres vivants. Et c'est en cela que la biométrie se distingue de la statistique qui n'a pas à se préoccuper des objets qu'elle manipule, c'est en cela que nous ne pouvons pas parler de hasard dans les termes où le fait cette branche particulière des mathématiques.

Peut-être est-il bon d'insister sur une autre différence. Ainsi l'intervention classique de la statistique en agronomie a consisté à proposer des plans expérimentaux dont l'interprétation était canalisée grâce à des techniques comme celles de l'analyse de la variance. Les données expérimentales étaient exactement, sauf erreurs ou omissions, celles prévues par les plans. Le cadre dans lequel l'information était acquise était donc, sauf accident, celui que permettait la technique, quels qu'en soient les raffinements. L'intervention actuelle des mathématiques dans les Sciences de la Nature et spécialement en Ecologie est totalement différente. L'objet de la recherche n'a pas été modelé à l'avance; les grandes lignes de l'expérimentation non plus; c'est au moment de l'acquisition des données qu'on met en jeu un certain modèle, lui-même lié plus ou moins étroitement à un modèle d'interprétation de ces données. Tout autre modèle, appliqué à la même réalité, eût éventuellement conduit à l'acquisition d'autres données. La construction du modèle d'acquisition des données fait donc partie intégrante du travail expérimental proprement dit.

Les êtres vivants sont des systèmes éminemment organisés (8), aussi bien vers l'intérieur - et jusqu'au niveau moléculaire - que vers l'extérieur - et jusqu'au niveau des écosystèmes. C'est cette organisation avec toute sa complexité et sa souplesse (ces deux phénomènes étant évidemment liés), c'est cette organisation dans son fonctionnement et son évolution (ces deux phénomènes étant également liés) que nous devons donner comme objectif d'étude à la biométrie. C'est cette organisation dans toute son ampleur.

x

x x

Plusieurs difficultés majeures, de notre point de vue, peuvent être mises en évidence.

La première qui peut paraître triviale est celle de l'acquisition des données; mais il faut la souligner dans sa dimension nouvelle qui lui est donnée par l'objectif à atteindre, qui n'est pas le constat d'une variation, mais celui d'une organisation, et qui implique la multiplicité et la simultanéité des mesures. La valeur de telles informations n'est sans doute pas contestable, puisque leur propriété pose des problèmes, jusqu'à l'échelon international.

La deuxième est celle de la complexité et de la particularité de l'organisation biologique. Les situations que la biométrie doit étudier et dont elle doit rendre compte ne sont pas quelconques. Tous les cas que la multiplicité des éléments d'un système permettraient d'imaginer dans une perspective de combinatoire ne sont pas biologiquement et physiquement possibles. Et il faut s'habituer à l'idée qu'ils ne l'ont jamais été à un moment donné quelconque. En outre une partie seulement des possibles s'est réalisé. C'est en cela que nous ne pouvons raisonner en géomètres; une partie de notre tâche consiste à expliquer cette liste des possibles et cette liste des réalisés. Nous avons à découvrir les contraintes limitantes, à les exprimer, à les interpréter. C'est en cela que la biométrie quels que soient ses moyens et ses méthodes est une science de l'expérience.

La troisième difficulté est celle des échelles; car l'organisation biologique se révèle à divers niveaux, non seulement à ceux que nous avons déjà soulignés (moléculaire, cellulaire, organismique et écologique), mais à l'intérieur même d'un niveau des degrés peuvent être distingués. Toutes ces strates ont leur organisation et tous ces plans d'organisation sont cohérents ou au moins compatibles. Nous avons à les démêler, à les dissocier par le jeu de l'expérimentation, mais aussi par celui de la procédure d'acquisition des données, et même par celui de l'interprétation. Nous avons à tenter d'en mesurer l'importance, les limites, la redondance. Nous avons aussi à nous poser quelques questions : ces différents niveaux sont-ils plusieurs apparences d'une même réalité? apparences que nous aurions formées, découpées par la diversité de nos approches techniques? ou encore les phénomènes relevant d'un certain niveau (organismique par exemple) se déduisent-ils mécaniquement de ceux relevant d'un autre niveau (moléculaire par exemple)? ou bien ont-ils une autonomie, même partielle, que nous ne saurions réduire, et dans ce cas quelles relations existe-t-il entre ces niveaux (ou ces strates de structure et de fonctionnement, quelles contraintes jouent-elles derrière le

voile des phénomènes immédiatement perceptibles ? des limites d'état, des relations de dépendance plus ou moins étroites, des interactions plus ou moins mécaniques, des articulations au sens des instructions d'un programme ou au sens des fournitures en matière ou en énergie ?

La quatrième difficulté qui peut apparaître plus technique provient du fait que l'expression de cette organisation ne semble pas se faire dans un système unique. En tout cas, nous ne l'apprécions pas, nous ne la percevons pas dans un même système. Tantôt des relations d'équivalence, tantôt des relations d'ordre, tantôt des relations fonctionnelles paraissent prévaloir, si bien que notre traduction en langage mathématique et le traitement des informations ainsi formalisées peuvent se révéler très différents selon les cas. Quelle est la part de notre intervention dans cette diversité ? Dans quelle mesure reflète-t-elle des sauts réels dans l'organisation de la matière vivante ?

Une dernière difficulté réside dans le fait qu'on ne sait toujours pas - malgré de nombreuses tentatives (9) - mesurer de façon satisfaisante un degré d'organisation, qu'on ne sait toujours pas comparer deux degrés d'organisation. Sans doute la question est-elle mal posée, de façon trop générale, et sans faire référence à un domaine et à des critères précis. Il reste que la biométrie au sens où nous la redéfinissons ferait des progrès décisifs, si cette question était éclaircie. Peut-être même l'un des objectifs de la biométrie pourrait-il être la mesure du degré d'organisation dans le domaine précis des êtres vivants.

Nous pouvons conclure brièvement : pour nous la biométrie a pour champ essentiel d'activité l'étude de l'organisation dans les systèmes biologiques.

Jean-Marie Legay
Laboratoire de biométrie
Université Cl. Bernard Lyon I

Références

- (1) Kashevsky N., Mathematical models and general mathematical principles in biology, Nuovo Cimento, suppl., Ital., 18, 2, 140-148, 1960.
- (2) Legay J.M., La méthode des modèles, état actuel de la méthode expérimentale, Informatique et Biosphère, 1973.
- (3) Waddington C.H., The basic ideas of biology, Towards a theoretical biology 1 Prolegomena, 1-32, 1968.
- (4) Legay J.M., Allometry and systematics : Insect egg form (en préparation).
- (5) Crick F.C.H. et Lawrence P.A., Compartments and polyclones in insect development, Science, 189, 4200, 340-347, 1975.
- (6) On consultera les ouvrages de données (quantitatives et qualitatives) suivants:
 Altman P.L. et Dittmer D.S., Growth, Feder. of Amer. Soc. for Exper. Biol. Washington, 1962.
 Altman P.L. et Dittmer D.S., Biology data book, Feder. of Amer. Soc. for Exper. Biol., Washington, 1964.
 Altman P.L. et Dittmer D.S., Respiration and Circulation, Feder. of Amer. Soc. for Exper. Biol., Bethesda (Maryland), 1971.
 Spector W.S., Handbook of Biological data, Saunders Co, London, 1956.
- (7) Alexander R.M.N., Animal mechanisms, Sidgwick and Jackson, London, 1968.
- (8) Alexander R.M.N., Size and shape, Edward arnold, London, 1971.
- (9) Schreider E., Biométrie, Encyclopaedia Universalis, 3, 306-307, 1968.
- (10) Von Bertalanffy L. Théorie générale des systèmes, Dunod, Paris 1973.
- (11) Atlan H., L'organisation biologique et la théorie de l'information, Hermann, Paris, 1972.