

JSFS

Comptes rendus de lecture

Journal de la société française de statistique, tome 144, n° 4 (2003),
p. 107-110

http://www.numdam.org/item?id=JSFS_2003__144_4_107_0

© Société française de statistique, 2003, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

COMPTES RENDUS DE LECTURE

Methodology and Epistemology of Multilevel Analysis Approaches from Different Social Sciences

Daniel COURGEAU, éditeur
1 vol., 235 pages, Kluwer, 2003, ISBN 1-4020-1475-9

Pour un statisticien, la définition de l'unité statistique est la première clé de toute analyse. Si celle-ci s'impose à l'évidence dans plusieurs types d'applications, il en va tout autrement dans d'autres. C'est ainsi que les sciences sociales ont vu se concurrencer, voire s'affronter, diverses approches selon que l'unité de base est l'individu (l'être humain) ou le groupe social (à un certain niveau d'agrégation). L'analyse multiniveaux vise à dépasser cette opposition : le présent ouvrage en discute les divers aspects méthodologiques et épistémologiques. Si une bonne part des questions soulevées est commune à diverses disciplines, d'autres aspects sont plus spécifiques à l'une ou à l'autre. Le chapitre 1, par Harvey Goldstein, est consacré aux sciences de l'éducation, le suivant, par Daniel Courgeau, à la démographie. Dans le chapitre 3 Ana V. Diez Roux traite de santé publique et d'épidémiologie, tandis que Marc Tranmer, David Steel et Ed Fieldhouse traitent d'un problème de géographie humaine dans le chapitre 4. Le chapitre 5, par Bernard Walliser, s'occupe d'économie. Le chapitre 6, par Robert Franck, reprend les aspects philosophiques d'un point de vue plus général et fondamental. Ainsi chaque thème est-il traité par un spécialiste reconnu, ayant à son actif des apports significatifs au sujet, un des avantages des ouvrages collectifs. Mais il est remarquable que, contrairement à ce qui se passe souvent dans ce type d'ouvrage, l'ensemble présente une grande unité et très peu de redondances, sans doute à cause de l'important travail d'édition : c'est ainsi qu'une introduction et une conclusion générales substantielles, par Daniel Courgeau, donnent en début d'ouvrage une vue d'ensemble permettant de bien suivre les développements spécifiques des divers chapitres, en fin d'ouvrage des pistes de discussion et d'approfondissement valables pour l'ensemble des disciplines concernées (et au delà...). Par ailleurs, l'éditeur a écrit une introduction à chacun des chapitres le replaçant dans le contexte général. La bibliographie est donnée chapitre par chapitre, ce qui est ici sans doute la meilleure option, d'autant plus que l'ouvrage contient un index général des sujets et un index général des auteurs cités.

Sans vouloir être exhaustif, voici quelques éléments qui ont retenu plus particulièrement mon attention. En amont, il y a d'abord la perspective historique et une réflexion épistémologique, les deux indispensables pour comprendre où et comment interviennent la modélisation et l'analyse statistique : comment se structurent les divers niveaux d'agrégation ? Sont-ils emboîtés ou non ? Quid du facteur temps ? etc. Au plan plus technique, dans l'optique envisagée par l'ouvrage, on comprend que le modèle mixte (au sens le plus large) soit ici privilégié. Divers chapitres expliquent en détail pourquoi et comment celui-ci intervient, au plan général et sur des exemples concrets ; de ce point de vue, on peut dire que l'ouvrage complète, pour les sciences sociales, le numéro spécial récent de ce Journal « Modèles Mixtes et Biométrie » (vol. 143, n° 1-2, année 2002). Les questions abordées sont aussi discutées à la lumière de la nature même des données disponibles (le plus souvent il n'est pas envisageable d'expérimenter et l'on doit se contenter de données d'observation). Ainsi, au delà des éléments plus spécifiquement liés à l'ensemble des disciplines envisagées, l'ouvrage contient quantité de matière susceptible d'alimenter la réflexion de tout statisticien car, de façon générale, les questions abordées débouchent sur des perspectives intéressantes quant aux besoins de développer certains aspects de méthodologie statistique. Je recommande donc chaudement sa lecture à tous les statisticiens qui souhaitent mieux comprendre les tenants et aboutissants de leur discipline et/ou prospectent quelques voies de développement.

Henri Caussinus

Modèles et Algorithmes Markoviens

Bernard YCART

1 vol., 272 pages, Collection « Mathématiques et Applications »,
vol. 39, Springer, 2002, ISBN : 3-540-43696-0

Voici un excellent livre, un excellent manuel de simulation, qui profite de la grande expérience de son auteur et la résume. L'ouvrage est bourré de programmes et d'exercices bien choisis. Dès le chapitre 2, on se régale à voir l'auteur démystifier l'« importance sampling » page 29, à fabriquer des permutations aléatoires (2.5), avec les commentaires pratiques sur l'efficacité de telle ou telle méthode, à comparer des algorithmes de simulation des lois normales (2.6). J'ai même appris que $\mathcal{L}(X|Y \geq (1-X)^2/2)$ est $N(0, 1)$ quand X et Y sont iid et exponentielles de moyenne 1. Au chapitre 3, les chaînes de Markov (à temps discret, homogènes, simulables) d'espace d'états E sont définies en se donnant une loi ν sur les fonctions de E dans E et en considérant une suite iid (Φ_n) de v.a. de loi ν . La chaîne de Markov $X = (X_n)_{n \geq 0}$ gouvernée par ν est donnée par $X_n = \Phi_n \circ \Phi_{n-1} \circ \dots \circ \Phi_1(x_0)$. Ici $x_0 \in E$ peut être fixé, ou bien aléatoire indépendant de (Φ_n) . Le secret de la méthode de Propp et Wilson pour calculer la mesure stationnaire de la chaîne quand E est fini est d'étudier $W_n(x) = \Phi_1 \circ \Phi_2 \circ \dots \circ \Phi_n(x)$, qui devient une fonction constante si n est assez grand, si la chaîne est irréductible et apériodique et si ν est concentrée sur les fonctions contractantes (telles que $|\Phi(A)| < |A|$ si $A \subset E$ et $|A| > 1$). Dans ces conditions, si $T = \inf\{n; W_n = \text{constante}\}$ alors $\pi_i = \Pr(W_T \equiv i)$ est la mesure stationnaire. Ce chapitre 3 est très informatif, les exercices y changent de caractère (5 pages pour l'un d'entre eux, le jeu de Penney...). Quand à la phrase de l'introduction « ceci n'est pas un livre de probabilités » elle laisse rêveur quand on voit la somme de théorie d'équations différentielles stochastiques à maîtriser pour comprendre ce chapitre 3, le plus difficile pour le public de ce livre. Le chapitre 4 est tout simplement épatant : outre Propp et Wilson déjà cité, et Metropolis et ses applications au dénombrement approximatif, vous avez tout ce que vous avez toujours voulu savoir sur le recuit simulé, sur les algorithmes génétiques et sur MOSES sans oser le demander. Le chapitre 5 est plus classique : comment simuler les chaînes de Markov en temps continu avec du discret : c'est appliqué aux files d'attente, aux réseaux de Petri et de Jackson, plus aux systèmes de particules, avec de bonnes illustrations comme le modèle prédateur proie. Le chapitre 6 est une excellente introduction au logiciel Scilab, outil de simulation recommandé.

Je trouve quand même quelques défauts au livre : à force de dire implicitement au lecteur « ce n'est pas difficile », on finit par être vague alors qu'on aurait besoin de définitions claires, de bornes précises dans le contexte duquel sont écrites certaines phrases : est-on en train d'énoncer un théorème de probabilités, d'examiner les problèmes de simulation qu'il pose, ou d'estimer

des paramètres statistiques? L'usage des mots «estimer» et «estimation» (page 16-17-18) prête à confusion pour le statisticien et les commentaires de la fin de 2.1.4 n'arrangent rien. Certaines phrases emplissent le lecteur de perplexité : «Comment coder une fonction Random... est un problème que nous considérons arbitrairement comme résolu par Marsaglia...»; «les comportements de la chaîne en deux instants éloignés l'un de l'autre sont à peu près indépendants» (page 117). Ou celle-ci : «Une vraie suite aléatoire doit être uniforme. Mais peut-on considérer comme aléatoire toute suite uniforme?» (Page 51) (Ce qui suit toutefois au 2.8.3 est clair et bien fait. Mais fallait-il différer une bonne définition de l'expression «suite aléatoire»? Le lecteur qui resterait sur sa faim devra lire «What Is a Random Sequence?» de S. B. Volchan (2002) *Am. Math. Monthly* **109** 46-63). Il y a souvent confusion : «les seules notions de probabilité qui seront utilisées sont les suites de variables indépendantes (*fonctions d'appel de Random successifs*). . . » Au chapitre 4 on a des expressions obscures comme «état d'équilibre de la chaîne» (alors que le mot état est déjà pris) ou «état d'équilibre stable», expression indéfinie, ou «la chaîne n'a pas atteint son état d'équilibre». Comme l'auteur a un bon contact avec son public, le langage est plutôt parlé : «où est le piège?» «résolument pratique», «pour aller plus loin», «qu'à cela ne tienne», «rester vigilant», «le requin trucidé l'un de ses voisins».

Certains choix sont discutables : pourquoi ne pas donner les méthodes de simulation des lois usuelles? Où diable l'étudiant, c'est-à-dire l'utilisateur potentiel de ce livre, irait-il d'abord les chercher si ce n'est ici? Car c'est un manuel, puisque l'auteur ne donne pas systématiquement ses références comme on le ferait dans un ouvrage de recherche. J'ai trouvé que le théorème de Perron Frobenius aurait droit à une référence, car Feller ou Chung (tiens, le meilleur livre jamais écrit sur les chaînes de Markov n'est jamais cité) ne l'utilisent jamais dans leur exposé du sujet, ce qui veut dire qu'un étudiant aurait le droit de l'ignorer. Pas une seule fois le nom de von Neumann n'est cité : il l'aurait pourtant mérité dans les méthodes de rejet du 2.3.

Encore trois observations de pion : selon l'auteur «Il n'y a pas d'expression simple pour la fonction de répartition (de la loi de Poisson).» Mais

$$\sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} = 1 - \int_0^\lambda e^{-t} \frac{t^n}{n!} dt$$

mérite d'être mentionné. Ensuite : une «suite récurrente» sur E est une suite de la forme $u_{n+1} = g(u_n)$ avec $g : E \rightarrow E$. Une phrase comme «une suite récurrente sur un ensemble fini est périodique» est embarrassante. Et que faire d'une phrase comme «les valeurs de u_n et u_{n+1} sont indépendantes?» (page 50). Enfin : polynomial ne prend pas d'accent circonflexe, à la différence de polynôme, exactement comme omicron n'est pas oméga...

Ces remarques sont bien mesquines pour un ouvrage aussi original et important. Précipitez vous sur ce livre, et à vos claviers.

Gérard Letac
 Université Paul Sabatier