

À lire, à voir – À lire, à voir
À lire, à voir – À lire, à voir

À propos de l'ouvrage « Analyse statistique des données spatiales », de Jean-Jacques Droesbeke, Michel Lejeune et Gilbert Saporta, paru aux éditions Technip en 2006

 Christine Thomas-Agnan et Anne Ruiz-Gazen*

Ce livre offre un panorama relativement complet des méthodes statistiques utilisées en présence de données spatiales, c'est-à-dire pour des observations dont la position géographique est connue et que l'on rencontre en particulier dans des domaines tels que l'économie, l'écologie, la géographie et la géochimie.

Après un premier chapitre plutôt historique, l'ouvrage aborde tour à tour les grands domaines de la statistique spatiale en commençant par l'économétrie spatiale et en continuant avec les modèles du second ordre, les processus ponctuels et la géostatistique. Les derniers chapitres correspondent à des études de cas dans ces différents domaines. Les auteurs s'adressent à un public ayant de bonnes connaissances en statistique mathématique et souhaitant se former à l'ensemble des méthodes statistiques adaptées au traitement des données spatiales.

Le premier chapitre présente un historique intéressant sur la corrélation et la régression. On regrettera toutefois l'absence de lien avec les données spatiales. Une présentation du coefficient d'auto-corrélation de Moran dans un contexte historique aurait, nous semble-t-il, été pertinente.

Le chapitre 2, quant à lui, fait une bonne revue de la littérature sur les coefficients d'auto-corrélation les plus classiques que sont le I de Moran et le C de Geary. L'auteur de ce cha-

pitre va au-delà d'une présentation bibliographique détaillée (ce qui est fort utile, compte tenu du nombre de publications sur le sujet) et propose de nouveaux indices dits corrigés des indices usuels. Le chapitre est très formalisé et manque d'exemples concrets. En particulier, la compréhension des matrices de voisinage ou de poids (appelées de façon non standard « matrice d'information locale ») aurait été de beaucoup facilitée par la présentation d'exemples simples tirés d'applications concrètes.



Le troisième chapitre offre un catalogue détaillé des principaux modèles de régression spatiale utilisés en économétrie, où l'information géographi-

que attachée à chaque unité statistique est une zone géographique plutôt qu'un point. Il donne aussi une classification intéressante de ces modèles en termes de modélisation d'effet de débordement et d'externalités. La section concernant les différences entre économétrie spatiale et géostatistique est aussi digne d'intérêt.

Les chapitres 4 et 5 terminent la partie économétrie spatiale en présentant de façon très complète les tests de spécificité et les méthodes d'estimation (souvent délicates) dans les modèles de régression spatiale. Là aussi, on peut louer un gros travail de synthèse mais les exemples ou des références à des exemples d'application font défaut.

Les chapitres 6 et 7 présentent une variété de modèles de champs aléatoires indexés par une grille (ou réseau) finie ou infinie et régulière ou irrégulière. Pour le cas où le champ est à valeurs réelles, les modèles (dits du second ordre) ARMA (auto-

* Christine Thomas-Agnan est professeur de statistique à l'Université de Toulouse I. Anne Ruiz-Gazen est maître de conférences de statistique dans cette même université. Depuis plusieurs années, elles enseignent ensemble un module de statistique et économétrie spatiale en deuxième année de master professionnel statistique et économétrie des Universités de Toulouse I et III. Elles participent également à un module de statistique et économétrie spatiale en deuxième année de master recherche de l'École doctorale de mathématiques de Toulouse. Enfin, elles sont co-auteurs d'un package de R, nommé GeoXp, dévoué à l'analyse exploratoire des données géoréférencées.

regressive moving average) et CAR (conditional autoregressive) spatiaux sont présentés en développant les analogies avec les séries temporelles. Sont ensuite présentés, dans le cas d'un espace d'état plus général, les modèles de champs de Gibbs-Markov. Il est à noter que ces deux chapitres comportent une série intéressante d'exercices théoriques sur ces modèles mais pas d'applications concrètes.

Le chapitre 8 est une introduction assez complète à la théorie des processus ponctuels, données spatiales pour lesquelles la position est modélisée comme aléatoire. Le niveau d'abstraction est très bien ciblé. Les divers paramètres attachés à un processus ponctuel sont décrits et divers modèles sont introduits, avec quelques exercices. Pour les modèles introduits dans les chapitres 6 à 8, champs aléatoires sur grille et processus ponctuels, les méthodes de simulation par MCMC (Monte Carlo Markov Chains) sont abordées dans le chapitre 9 et les méthodes d'estimation statistique par la méthode de

pseudo-vraisemblance dans le chapitre 10.

Le chapitre 11 traite du problème du changement d'échelle sur données régionales et discute notamment du transfert des propriétés d'un processus spatial lorsque l'on agrège les unités géographiques.

Le domaine de la géostatistique est présenté dans les chapitres 12 à 15 mais on notera que cette partie ne comporte ni exercices ni exemples concrets de prédictions. Le chapitre 13 passe en revue les modèles courants de la géostatistique linéaire en introduisant succinctement les notions de fonction aléatoire stationnaire, intrinsèque d'ordre k , isotrope, le variogramme, la covariance généralisée, le krigeage et le co-krigeage. La prédiction de transformations non linéaires d'une variable régionalisée, comme par exemple le risque de dépassement d'un seuil critique, donne lieu au développement du krigeage disjonctif dans le chapitre 15. Enfin, la prédiction non linéaire peut se faire par simulation, c'est pourquoi

ce chapitre est entièrement consacré aux méthodes de simulation, conditionnelle ou non, d'un champ aléatoire gaussien par méthode spectrale ou par bandes tournantes.

Le volume se termine dans les chapitres 16 à 19 par quatre applications relatives aux différents domaines de la statistique spatiale. Ainsi, le chapitre 16 illustre notamment l'utilisation de méthodes d'économétrie spatiale avec une étude de l'auto corrélation spatiale de données de produit intérieur brut par habitant en Italie. Il aborde aussi brièvement les méthodes de classification avec contrainte de contiguïté et, de façon plus générale, fait le lien entre statistique et géographie. Le chapitre 17 concerne une étude de cas géostatistique et le chapitre 18, l'ajustement d'un modèle de processus ponctuels. Enfin, le dernier chapitre illustre une méthode récente (développée par les auteurs du chapitre) qui utilise une analyse en coordonnées principales de la matrice de voisinages pour une modélisation spatiale de tableaux d'abondances d'espèces en écologie. ■

Table des matières de l'ouvrage

- | | |
|---|---|
| 1. La corrélation et ses dérivés : le rôle de Galton dans leur histoire | 15. Prédiction non-linéaire par simulation conditionnelle |
| 2. Dépendance spatiale et autocorrélation | 16. Discontinuités statistiques et discontinuités spatiales : l'exemple des inégalités de richesse par habitant en Italie (1951-1991) |
| 3. Les modèles d'autorégression spatiale | 17. Validation d'un modèle géostatistique pour l'interpolation : application à un événement pluvieux |
| 4. Modèles autorégressifs : tests de spécification | 18. Poisson non-stationnaire et échantillonnage semi-raréfié : comment les cèdres du Lubéron envahissent-ils l'espace ? |
| 5. Modèles autorégressifs et problèmes d'estimation | 19. Quelles sont les échelles spatiales importantes dans un écosystème ? |
| 6. Modèles du second ordre | |
| 7. Champ de Gibbs-Markov sur un réseau | |
| 8. Processus ponctuels spatiaux | |
| 9. Simulation des modèles spatiaux | |
| 10. Estimation des modèles spatiaux | |
| 11. Données régionales et agrégation spatiale | |
| 12. Le cadre général de la géostatistique | |
| 13. Géostatistique linéaire, prédiction linéaire par krigeage | Bibliographie |
| 14. Krigeage disjonctif | Index |