「特集 空間統計モデリング：理論と応用」について

村上 大輔¹・瀬谷 創² (オーガナイザー)

地理情報システム (Geographic information system; GIS) や各種測位システムが発展し、大規模データを扱う基盤も日進月歩する中、自然科学から人文・社会科学に至るまで、社会・経済・環境・生物・疾病・交通・気象といった幅広い情報が、位置情報と紐づけられた(時)空間データとして高頻度かつ高解像度で観測され、日々活用されている。

一般に、空間データには空間従属性 (spatial dependence) や空間異質性 (spatial heterogeneity) という性質が見られることが知られている。空間従属性とは、データが近隣のデータとより強い従属関係を持つという性質であり、空間異質性とは、場所毎にデータの特性が異なるという性質である。Tobler (1970) は前者を地理学第一定理 (first law of geography) 「Everything is related to everything else, but near things are more related than distant things」として位置づけ、後者はしばしば地理学第二定理として位置付けられるなど、古くから空間データの統計モデリングを行う上での重要な性質として認識されてきた。

空間データを対象とした統計モデル研究の起源は、概ね 1950 年代まで遡れるようである。Moran (1950) は空間従属性の代表的な診断統計量である Moran I 統計量を提案した。また、Krigé (1951) が鉱物資源の分布を評価する目的で開発し、フランスの数学者の Georges Matheron 氏により拡張・体系的に整理された空間予測手法は、今日ではクリギング (kriging) と呼ばれ広く用いられている。その後、1980 から 90 年代にかけていくつかの標準的なテキストが刊行され、空間データを対象とした統計モデルの研究はその裾野を広げてきた。例えば、自然科学を主な適用対象に発展してきた地球統計学 (geostatistics) (Cressie, 1991)、社会科学における空間計量経済学 (spatial econometrics) (Anselin, 1988)、医学における空間疫学 (spatial epidemiology)、生態学における空間生態学 (spatial ecology) 等である。今日では、それらの分野の垣根はどんどん曖昧になっており、空間データを取り扱う統計分野は総称して空間統計学と呼ばれることが多くなっている。

空間統計学では、モデルの代表的なパラメータにナゲット、レンジ、シルがあり、共分散の要素を与える関数はコバリオグラムと呼ばれる。かつては、これら特有の概念が参入障壁となっていた部分があるが、近年ではクリギングがガウス過程 (Gaussian process) の空間予測に特化した (2-4 次元の) 応用例と位置づけられ、手法の見通しも格段によくなってきている。現在、空間従属性や空間異質性を柔軟にモデル化することのできる統計モデルの開発や理論的な整理が急速に進んでいる。他方、空間統計モデリングのためのソフトウェアの整備も着実に進んでいる。空間統計学分野では、研究で実装した手法を R のコードとして公開することが一般化しており、多数の有用なパッケージが存在する。また近年では機械学習分野との融合が進んでいることから、Python のコードとして公開される例も増加している。以上より、空間統計学は、学問分野としてある程度成熟してきたといえる。

¹ 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

² 神戸大学 大学院工学研究科：〒657-8501 兵庫県神戸市灘区六甲台町 1-1

その一方で、空間データを取り巻く状況は近年目まぐるしく変化している。具体的には、観測技術やWeb情報などの自動収集技術の進展に伴い、空間データの大規模化、超高頻度化、多様化が進んでいる。多様化については、低質なデータ(画像やテキストなどの未成形の情報からなるデータ等)の普及も顕著である。また、社会科学分野では因果推論が重要視され、空間データモデリングも、因果推論への貢献を求められつつある。このような中、データの大規模化・超高頻度化・多様化に対応するような新たなモデルがベイズ推論や機械学習を駆使しながら盛んに開発され、空間データの因果推論に関する研究も進みつつある。

以上述べたような最新の動向を踏まえ、本特集号「空間統計モデリング：理論と応用」では、空間統計モデリングに関連する7編の論文を掲載する。最初の2編は、地球統計学と空間計量経済学の研究動向についてのレビューであり、続く2編は近年の興味深い手法についてのよりフォーカスした紹介となっている。残り3本はそれぞれ賃貸物件市場、COVID-19、観光客の動向に着目した手法開発と応用に関する論文となっている。以上のように、空間統計学に関する文献レビューから理論・応用までをカバーする幅広い内容となっている。空間統計学に関する和文文献は、特に最近の研究動向に関するものは少ないと思われる。本特集号が空間統計学の学習や情報収集、今後の研究や実務の参考として役立てば幸いである。

最後に、本特集「空間統計モデリング：理論と応用」の執筆ならびに査読にご協力いただいた皆様にこの場を借りて感謝を申し上げます。

参 考 文 献

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.
- Cressie, N. (1991). *Statistics for Spatial Data*, Wiley, New York.
- Krige, Danie G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand, *The Journal of the Chemical, Metallurgical and Mining Society of South Africa*, **52**(6), 119–139.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena, *Biometrika*, **37**(1/2), 17–23.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region, *Economic Geography*, **46**, 234–240.