

まえがき

近年、計測技術の進展により各分野で大規模データベースが構築され、研究対象は複雑化・大規模化するとともに、情報を効果的・効率的に収集・集約し、革新的な科学的手法により知識発見や新たな価値を創造する高度な解析が求められています。こうした環境下、統計科学の主な課題は、線形モデルによる解析から非線形な解析、時点や空間を固定したデータの解析から時系列・空間的な解析と複雑化し、各分野固有の知識を必要とする、より高度なモデルや手法の開発へと移行しています。

リモートセンシングデータや地域データのような空間的位置を持つ空間データの構造を解析するためには、データを二次元、三次元的に表示するデータの可視化が有効です。地理情報システム (Geographic Information System, GIS) や最新のコンピューター機能は、こうした空間データ解析の強力なツールとして使用されていますが、空間データ構造の客観的な表現は困難であり、それらの解釈は主観的です。

こうした問題に対して、エシェロン解析 (echelon analysis) は、空間的な位置を表面上のデータの高低に基づき分割し、空間データの位相的な階層構造を系統的かつ客観的に視覚化するために開発されました。echelon とは、もともとは航空機などの梯隊ていたいのことで、同位相を持つ空間データの階層が隊列をなしていることに由来しています。エシェロンは、位置情報を持つ観測データの相対的に高い値と低い値を持つ領域に基づき、同じトポロジー構造を持つ領域全体の表面上の値を集約し、領域を階層的に表現します。エシェロンは、領域の表面を最上部から最下部までスキャンするとき、ピーク、スロープ、サドル等によって発生します。

エシェロン解析の最大の長所は、次元を問わず隣接情報が与えられた空間データに対して、その構造が客観的に階層化できることです。データの分布や構造を調べるために、ヒストグラム、箱ひげ図、幹葉図、散布図、散布図行列を使用するのと同様に視覚的に記述することができます。さら

に、本書で紹介したエシェロン構造を用いた圏、有意に高い値を示す地域（ホットスポット）の検出をはじめ、多次元時空間データへの拡張、多変量時空間データへの拡張などより高度な時空間データ解析への展開が数多くあるということです。

本書では、エシェロン解析の基礎となる考え方やアルゴリズム、さらに、エシェロンの構造を利用した応用例として、圏による地域の分類、リモートセンシングデータの分析、ホットスポットの検出、そして、エシェロン解析のためのソフトウェアを取り上げました。本書で取り上げた以外のエシェロンの構造を利用した空間データの分析法については、別の機会に取り上げたいと思います。

さて、本書はエシェロン解析に関する世界で最初の著書ですので、筆者らのこれまでの取り組みについて簡単にご紹介したいと思います。エシェロン解析は、1999年に栗原が故松下嘉米男先生からご紹介いただいた米国ペンシルバニア州立大学統計生態学および環境統計センターのG. P. Patil先生の研究室に文部省在外研究員として出張しているとき、ゼミナールにおいてW. L. Myers先生から紹介され初めて出会いました。帰国後も、エシェロン解析に関する国際共同研究は続きましたが、2002年ニューヨークでのJSM (Joint Statistical Meeting)、2005年モンリオールでのESA2005 (Ecological Society of America)での招待講演は良い経験でした。特に、年末年始にインドの統計学者が参集し行っていたインド、ハイデラバードでの国際学会では、年末に講演を行い、初日の出を飛行機の中で迎えたのはいまとなってはよい思い出です。また、石岡は2007年にPatil先生からの招きで米国ペンシルバニア州立大学で開催されたHotspot Geoinformaticsに関するワークショップに参加し、当時から開発を始めていたエシェロン解析のソフトウェアなどについて講演を行いました。Myers先生にはその内容について大変興味を持っていただき、“I encourage putting the R-based echelon software in the CRAN library of contributed modules.”というメッセージをいただいたことはその後の研究活動の励みとなりました。それから苦節十数年を経て、現在、RパッケージechelonがCRANで公開されています。

エシエロン解析については、これまで多くの共同研究者とともに研究を推進してきました。ホットスポット検出では、水藤寛先生（三次元環境汚染データのホットスポット検出）、韓相勲先生（韓国地震データのホットスポット検出）、富田誠先生（ゲノムデータのLDブロック同定）、小田牧子先生（パッチに基づく森林の分類）、多変量空間データへの応用では、洪韓杓先生（ボロノイ空間）、文勝浩先生（主成分空間）、エシエロン解析の応用では、羅明振先生（最適配置）、金秀珰先生（空間データの外れ値検出）、湊真一先生、水田正弘先生、川原純先生（格子の隣接ブロックの網羅の数え上げ）などがあげられます。ソフトウェアの開発は、石岡および梶西将司先生が中心となり、その研究開発を進めてきました。なるべく早く公開しようと考えていたのですが、もう少し機能を増やしてからと思っているうちに、あっという間に20年が経ってしまったという感じです。

最後になりましたが、本書を作成するにあたり、防衛医科大学の小田牧子先生、岡山大学統計学教室の梶西将司氏（当時）、竹村祐亮氏、神原あん氏には、本書の原稿を丁寧にご確認いただきました。また、閲覧者の先生方には原稿を隔々まで確認するとともに的確なご指摘をいただき、本書の内容の改善に役立ちました。今まで共同研究を行ってきた、上記の諸先生、もちろん、Patil先生、Myers先生、さらに、当時エシエロン解析の研究を進めるきっかけとなった在外研究を快く認めていただいた岡山大学の田中豊先生、垂水共之先生、故大竹正徳先生には深く感謝いたします。

また、統計学One Pointシリーズ編集委員長の鎌倉稔成先生には、本書の執筆を勧めていただいたことに感謝いたします。出版にあたっては、共立出版株式会社編集部の菅沼正裕氏には、約2年前に原稿の依頼を受けて以来、長期にわたり執筆を始めるまでお待ちいただくとともに、原稿が遅れがちな筆者を励まし見守っていただきました。さらに、編集制作部の大久保早紀子氏には編集作業を丹念に担当していただきました。ここに記してこれらの方々的心から感謝の意を表したいと思います。

2021年3月

栗原考次、石岡文生