

はじめに

重力波やニュートリノなどの電磁波以外でも天体の情報が得られるようになつたが、宇宙からの情報は主に電磁波によってもたらされる。より暗い天体、より遠くの天体を観測するために、大望遠鏡が建設され、最先端の技術を取り入れた観測装置が開発されている。そのような最新の技術をもってしても、天文学における最初の発見は、しばしば非常に暗くて少ないサンプルの中から生まれる。雑音に埋もれたデータや、少数のデータから重要な情報を引き出すためにはデータに潜む偶然雑音や系統雑音の正しい理解に基づく統計解析が不可欠である。

また、装置の大型化によって一度に得られる画像データは飛躍的に増え、今後もデータはますます増えていくことが予想される。一方、世界の天文台で得られた高品質のデータが大量にアーカイブされている。生画像からカタログデータに至るさまざまなアーカイブデータを用いて研究を行う機会も多いが、データそのものに内在する誤差やアーカイブするまでの過程で行われる情報処理の内容をほとんど知る機会がない。

本巻では、著者の専門分野である光と赤外線で画像として得られる情報とノイズ、さらに画像から抽出される天体情報の統計解析について解説する。宇宙からの情報には、望遠鏡で受けるまでにさまざまな雑音が混入する。望遠鏡に入ってきた光子は CCD や赤外線センサーによって電気的な信号に置き換えられるが、この過程においても雑音が混入する。このように雑音に埋もれた天体からの情報を統計的な手法によって高い精度で引き出すことが求められる。

本書の前半では統計学の基本を天体観測のデータによる具体的な例を挙げながら勉強し、後半では得られたカタログデータから個々の天体やその集団としての統計的性質を導くためのさまざまな手法を解説する。さらに天体からの情

報が画像としてディジタル化されるまでの信号と、そこに混入する雑音の理解、その画像から天体の情報を引き出してカタログを作成するまでの情報処理について解説する。特に、観測データに潜むさまざまな誤差について解説された教科書はほとんどなく、筆者が長く装置開発に携わり、得られたデータを使って研究を進めてきた経験に基づいて解説する。

本書の前半では統計解析向けの R 言語を用いている。一方、天文学や宇宙物理学の分野では固有のツールが Python をベースに開発されていることが多く、天文学向けの解析用パッケージも多数利用できる。付録に具体的なコーディング例を添えて、これらの言語を用いた解析方法を紹介する。高度な解析をしたい読者は是非、C 言語等を用いて、自分でコーディングして欲しい。R や Python はインタプリタ言語なので実行が遅い。特に、すばる望遠鏡などの大きな画像や大量のデータ、複雑なモデルの解析には C や FORTRAN などのコンパイル言語は圧倒的に速い。ソースコードが公開され、信頼の高い解析パッケージも多数あるが、自身でコーディングすることにより、解析手法の正しい理解の助けになり、また公開されているパッケージの不備や応用の限界を知ることもある。

本書では観測天文学を勉強する学部学生や修士課程の学生を対象として、天文学におけるデータ解析、統計処理に関する基本的な解説に留めた。これから天文学のデータを扱う学生の入門書としていただけたらありがたい。各専門分野についての解説はそれぞれの教科書や文献等にゆずる。なお本シリーズ第 1 卷「数理統計学の基礎」(尾畠伸明著) に記述されている定義、定理、証明を多く引用しているので、合わせて参照して欲しい。

最後に、本書を「クロスセクショナル統計シリーズ」の 1 冊に加えていただいた編集委員の諸先生方、および原著を丁寧に読んでいただき、多くの貴重なコメントをいただいた広島大学の植村誠氏に深い感謝の意を表したい。また辛抱強く、原稿の完成を待っていただき、また励まし続けていただいた共立出版編集部の山内千尋氏に感謝したい。