



「予測にいかす
統計モデリングの基本
改訂第2版
ベイズ統計入門から応用まで」
樋口知之 著
講談社, 2022年7月
175頁, 3,080円(税込)
ISBN 978-4-06-528570-1

ベイズの定理は不思議である。事象 A を所与としたときの事象 B の条件付き確率から、事象 B を所与としたときの事象 A の条件付き確率が求まってしまう。アマチュア数学者のベイズがこのことに気づいたのは18世紀中頃のことであったが、ベイズ統計が徐々に受け入れられ始めるまで約200年を要した。気象学でおなじみのデータ同化がベイズの定理に基づいていることは、ご存じの方もおられるであろう。現在ではベイズ統計はさまざまな分野で広く利用されており、その背景には、汎用的な計算法であるマルコフ連鎖モンテカルロ法や粒子フィルタの登場と、それらを動かすための計算機の高性能化がある。

本書は、ベイズ統計に基づく時系列モデリングの格好の入門書である。時系列解析では、現在までのデータに基づいて、過去・現在の状態の推定や未来の状態の予測を、統計モデルによって行う。モデリングのための数式化や主なアルゴリズムが、多くの図を使ってわかりやすくかつ的確に説明されており、著者の学識のなせる技であろう。改訂版は色刷りになったので、読みやすさが増した。読者層としては、大学の学部から修士課程の学生、ビジネススクールの学生、企業の研究者などが想定されており、プログラム言語を知っている読者が時系列モデルをプログラミングできるところまで導いてくれる。

本書の初版は2011年に出版された。この間に起きた大きな出来事といえば、なんといっても深層学習によるモデリングの驚くべき性能向上であろう。ただし、ブラックボックス化していることが大きな問題点とされている。「統計学の強みは、生成モデルの構築に関する、様々な知見とノウハウの蓄積、またモデルに基づく意思決定の綿密な評価にある。ある種、モデリングに関する匠の技ともいえる暗黙知に、統計学の存在感が増していくであろう」というのが、前書きに書かれた著者の主張である。これは、数値天気予報など物理

法則に基づくモデリングにも通じる考えであろう。

本文は基礎編(第1～4章)、発展編(第5～8章)、実践編(第9～12章)の3編からなるが、第8章までは順に読んでいく必要がある。第1章は、あるレストランの売上高の予測を考えることから始まる。ここでは統計モデルのニューラルネットに対する利点として、逆推論機能の提供、結果の可読性、学習の効率性が挙げられている。そのあと確率や統計モデルの話があり、第4章から第8章で計算アルゴリズムが解説される。過去の状態の推定に使われる平滑化アルゴリズム、統計モデルのパラメータの推定法、粒子フィルタなどが取り上げられている。特に粒子フィルタの実装については詳しく、システムノイズや観測ノイズの確率分布の作り方で載っている。

実践編の前半では、モデリングの前に時系列データをよく見ること、解析者の知識と経験を生かしてモデリングすることの重要性が説かれる。東京の年平均気温のトレンド解析と、第1章で触れたレストランの売上予測の実例が示されていて、大変参考になる。データ同化にも1章が割かれている。ただし、「ほとんどのデータ同化では『状態ベクトルの“一部の”変数を直接観測できること』を前提としている」とあるが、少なくとも気象・海洋学のデータ同化では、数値モデルの格子点と観測点の位置が一般に異なるので、これは当てはまらない。粒子フィルタというと物体追跡がよく知られているが、最後の章ではお掃除ロボットの移動制御が紹介される。いろいろな状況のもとでロボットに同一のコマンドを与えたときの行動の結果を、確率分布として知識化しておけば、粒子フィルタによって簡単にロボットを制御できる、というのが面白い。

著者の長年の経験に基づく統計モデリングの勘どころが、本文中やコラムに散りばめられているのはうれしい。確率変数間の依存関係を把握するためのグラフィカルモデルの利用、性質の悪いパラメータの推定法、計算機が生成する乱数を用いる際の注意点など、参考になることが多々ある。既出の数式や図を参照する際に、それらの掲載ページが記されているのも親切である。数式の多い書物を読んだことのある方なら、このありがたみはお分かりであろう。

ノーベル賞物理学者のファインマンは、長く務めた教壇を1988年に去るとき、黒板に次の有名な言葉を残した。What I cannot create, I do not understand. 本書を読まれた方は是非ご自分で計算機プログラムを作って、あれこれといじって頂きたい。

(気象研究所 露木 義)