

# まえがき

昨今「ベイズ統計」, 「ベイズモデリング」, 「ベイジアンネットワーク」などの、ベイズの言葉がついた専門用語を耳にする機会が非常に多くなってきた。またそれらを基礎技術とした製品やサービスが私たちの生活にどっぷりと入り込んできている。事実、米国の IT 大企業は、ベイズモデリングを専門とする数理科学・情報科学の大学教授を研究開発の統括者として引き抜き、ベイズモデリングの研究者を大量採用することで新製品・新サービスを次々と世に送り出している。

そもそもなぜ、ベイズモデリングが有用なのか？ また、きわめて簡単で古くから知られていたベイズの定理がなぜ、今、注目されるのか？ ずばり、コンピュータの計算性能向上、インターネットの発達、そして高精度センサーのコモディティ（日用品）化の3つが主たる理由としてあげられる。ベイズの定理の肝は、推定したい量を直接計算せず、間接的に積分計算（多くの場合、和に置き換えられる）でもって求めることである。ただし、この積分計算を実現するには、対象を計算機上で表現するために必要な大規模メモリーと高性能 CPU が必須であった。廉価な高性能計算機の登場なくしてベイズ統計の活躍はあり得なかったのである。

この積分の被積分関数には、事前分布と呼ばれる、対象の細かい情報をその不確実性を含めて数値化した確率分布が含まれる。事前情報を細かく与える作業は面倒に感じられる部分もあるが、データに基づいた事前情報の確度の評価がベイズの定理により実現できるメリットは大きい。事例が少ないがためにややもすると主観的な情報であっても、ベイズの定理によってデータに潜む細かい情報とひもづけられることで客観性が増すのである。これは、個人や“その状況”といった、極端に少ない事例に関する情報であっても、事前情報を1度与えれば、データからベイズの定理による学習により、客観性の高い情報に成長できることを物語っている。つまり、ベイズモデリングを用いれば、個人や状況に即した製品・サービス、つまり「個人化されたサービス」を実現できるのである。

インターネットは、個人の情報を網羅的に集め、同時に、個人をターゲットにした商品・サービスの提供を効率的に行える優れたインフラである。インターネットの発達が産業・生活に劇的な変革をもたらしたのは、別にベイズモデリング周

辺だけではないが、「個人化されたサービス」という、21世紀の産業の基幹をベイズモデリングが支えている事実はしっかりと認識してほしい。富を生む仕組みは、前世紀の「物質（モノ）を均質に大量に生産するシステム」から21世紀の「個人化された情報のサービス提供システム」に大きく変化したのである。

ベイズモデリングの隆盛を引き起こした3つ目の要因である高精度センサーのコモディティ化は、被積分関数を含むもう1つの確率分布である尤度関数<sup>ゆうど</sup>に関係している。興味ある対象を観測・計測して得られるデータは、計測精度の高いセンサーによってもたらされるべきであろうが、これまで予算面からその利用が制約されてきた。また、センサーを複数種類を組み合わせたほうが、対象に関する情報が多面的に得られるので望ましい。近年の高精度センサーの著しい低廉化は多様な尤度関数の利用を促してきた。このように、コンピュータの計算性能向上、インターネットの発達、高精度センサーのコモディティ化の3つがベイズモデリングの実用化の引き金となったのである。

では、ベイズモデリングの重要性は理解できたとして、もう少し基礎となる数学を理解したいという欲求が当然のごとく出てくる。ベイズモデリングに関連した専門用語を一般のビジネスマンなどにもわかりやすく解説する書籍は比較的手に入りやすくなってきた。ただし、読者が言葉だけの理解に終わらず、それらの基礎となっている数学をいざ勉強したいと思いついたときに、次のレベルの書籍が専門書や研究報告になってしまっている。この間の大きな隔たりをバトンタッチでき、基礎の数理科学・情報科学の教科書にもなるような書籍はこれまであまりなかった。本書の狙いの1つは、読み物や雑誌記事を卒業した読者が、次の段階として本書を読み、この本の読了後、専門書などに進むことができるようになることである。

本書のもう1つの目的は、読者自身の問題の具体的解決策を提供することである。基礎となる数学の理解だけにとどまらず、アルゴリズムを正確に習得し、自分の問題に対してベイズモデリングを行い、モデルに基づいて予測し、その予測結果を通じてモデルの評価を行い、必要があればモデルを改良する、この一連の情報処理の流れを学んでいただきたい。本書には、モデルを自分でつくれる、必要最小限の内容を含んでいる。よって、身近な問題があれば本書で得た知識をもとに自らモデリングを行い予測をしてみしてほしい。もし読者が何らかのプログラミング言語を知っているのであれば、モデルをプログラミングできるまで導いて

いくことが本書の目標である。

モデリングを学ぶには、具体的な事例に基づきながら基礎から応用にいたる解説があるべきであろう。世の中にはさまざまなモデルがあり、それらのモデルを使った成果がたくさん生まれている。本書のみで多様なモデルを解説するには限界があるので、筆者の経験が最も生かせる時系列モデルを扱いながらモデリングの解説を行いたい。時系列モデルをとり扱うほかの理由の1つとして、時間に依存しないモデルのとり扱い（たとえば、多変量解析）は多くの解説書などでとり扱われている一方、それらを学習した読者であっても、いざ時系列モデルになると、戸惑うケースを多々目にするからである。時系列モデリングにおいては、時系列データを一定の区間で分割し、その分割された区間に対して時間に依存しないモデルを適用してもよい結果は得られない。時間変動を生み出す動的な構造—ダイナミクスと簡単に呼ぶことも多い—をモデリングしない限り、モデルの評価と改善というサイクルがうまく機能しないからである。

本書がターゲットとする読者層は、大学の学部学生から修士課程の学生にとどまらず、企業の研究者や、ビジネススクールに通う学生なども含む。この本の読者層は理系であるかどうかを問わない。前提として必要とする数学的知識も、行列の積ぐらいである。本書を読むのに必要な微分積分や確率・統計の知識は本書内で解説しているので、内容として自己完結している。微分積分といっても、高校数学のレベルで十分である。確率・統計についても、もちろん、ガウス（正規）分布とは何か、平均や分散はどう求めるのかぐらいは知っておいてもらったほうがよいであろう。だが、必須知識ではない。とにかく、「自分でモデリングしたい」、「自分で予測したい」といった意欲をもった読者であれば、文系の学生であっても十分読み通すことができる。実際、そのことを念頭において本書は企画されている。

この本は3部構成になっている。前半の第1~4章は、モデリングの基礎となる数理や計算理論の解説を行う基礎編である。第1章では「予測とは何か」を、居酒屋レストランの売り上げデータを使いながら考えてみる。実際に役立つ予測モデルは第8章で記述されるので、ここでは単純なやり方では満足はいく予測はできないことを実感してほしい。第2章では、条件つき確率などの確率の基礎と、ベイズの定理を含めたベイズ統計の基礎を学ぶ。第3章では予測を行うために最も大切な、データを生成する統計モデルの紹介と解説を行う。第4章では、そのモデルがもつ特性を生かした、予測を実現するための計算理論を学ぶ。

中盤の第5～7章は、第2～4章で示された基礎理論をコンピュータ上で実際に展開する際に必要となる技術を解説する。第5章ではパラメータの具体的推定法とモデルの改善方策を学ぶ。第6章では粒子フィルタと呼ばれている、強力なアルゴリズムを解説する。第5、6章の説明時には、第3、4章で示された結果に立ち戻ることが多い。第7章は、データ生成モデルを使った解析と予測に必須の乱数の発生法を紹介する。乱数の発生法を知らずして、実践的予測の実現は難しいことを注意しておく。第7章の最後には簡単な時系列予測問題に対するアルゴリズムの詳細を示した。モデリングの未経験者には非常に役に立つはずである。

終盤の第8～10章は、それまで勉強したことの集大成と実践編である。具体的事例としては、第1章でとり上げた居酒屋レストランの売り上げデータのモデリングの結果が第8章で示される。売り上げデータの予測を精密に行うためには、既存の知識を総動員した手の込んだモデリングが求められることが明らかになる。さらに、シミュレーションとデータ解析を融合する技術（データ同化と呼ばれる）を第9章で、またロボット制御を第10章でとり扱う。ロボットの動作の原理を記述するモデルを知らなくても、経験とデータに基づいて“そこそこ”制御できる方策を示す。

このように第8～10章をのぞいて、章の内容は常にそれより前の章の内容を引用しているため、通読が必須である。“つまみ食い”は難しい構成となっていることをご了解願いたい。問題を見つけたときに、「自分でモデリングし、自分で計算を行い、自分で予測して、自分で制御する」意欲をもった人に向けた本である。ぜひ、通読してもらいたい。また、解説文のところどころに、コラムとして（印刷ではバックが灰色になっている）、解説の調子をややはずした、モデリングの先輩として大所高所からヒントを記した。解説とともに楽しんでもらえれば幸いである。

統計数理研究所・特任研究員の林圭佐博士には作図の手助けをしていただいた。ここに感謝の意を表します。また林博士と廣瀬慧君（九州大学・博士後期課程）および本橋永至君（総合研究大学院大学・博士後期課程）には、原稿を丁寧に読んでいただき多くの指摘を頂戴した。本書の出版にあたっては、講談社サイエンティフィク瀬戸晶子さんにはひとかたならぬお世話になった。感謝と御礼を申し上げたい。

2011年3月

樋口知之