

書 評

21世紀の統計科学III：数理・計算の統計科学

北川源四郎・竹村彰通 編，東京大学出版会，2008年

九州大学大学院数理学研究院 二宮嘉行

日本統計学会は、統計学・統計科学に関連する日本の学術団体の中で最大かつ最古の団体である。その日本統計学会の創立75周年を記念して発刊されたのが、三冊から構成されるシリーズ「21世紀の統計科学」であり、本書はその第三巻である。このシリーズが発刊された目的は、統計学・統計科学における最新の動向・議論が専門家に限らない多くの人に理解されることとある。第一巻「社会・経済の統計科学」および第二巻「自然・生物・健康の統計科学」がより応用的な視点で書かれている傾向があるのに対し、「数理・計算の統計科学」と副題を打たれた本書はより理論的な視点で書かれている傾向があり、第一巻や第二巻の基礎にもなっている。

さて、いきなりであるが、本書の中にある表現を用いて統計学およびその動向について書いてみることにすると……統計学の目的が「不確実なデータに基づいて、それらが得られた集団、あるいは、それらを発生させるメカニズムについて何らかの決定を行うこと（第三章）」であり、また「統計的推測の形式にのっとりることにより、推測の結果の精度自体が評価できるという点（第一章）」が統計学の大きな意義であったりすることは今も昔も変わらない。その中で、これまでの統計学においては、「Waldにはじまる統計的決定理論の枠組のような形式化が重視される傾向があった（第一章）」。また「主として小標本の場合に、精密理論、あるいは、標本数を大にしたときの漸近理論が展開され、様々な分野へ適応されてきた（第三章）」。しかし現在、「電子化された測定技術の進歩により科学の諸分野で、大規模なデータが蓄積され、それらの解析法の開発が重要な課題になっている（第三章）」。そこで必要とされるのが計算統計学である。計算統計学とは、「急速に発達した計算基盤および計算アルゴリズムを活用して統計科学の可能性を広げようとするものであり、本質は柔軟なモデリングの実現を目指すところにある（第二章）」。そしてそれは、「これまでの仮想世界における推論から現実世界における予測と知識発見へ一歩進めようとする挑戦である（第二章）」。ただし「大規模な階層ベイズモデルのような複雑なモデルでも、部品としては正規分布などの伝統的な部品が利用されることが多いように、計算機が無かった頃の数学者の計算の蓄積の中に、今後も役立つものは多いはず（第一章）」であり、過去の知識の再認識も必要である……などを得ることができる。私の文章に合わせるため、指示代名詞を補完したり記述をまとめたりして文章を若干変えてしまっている。また切り貼りもしているのでもともとの文章の意図がずれてしまっていることがあるかもしれず、そうであったら許していただきたいが、いずれにせよ、デー

データを解析する際の設定は現在急激に変化している。そして、今までとはデータのタイプが異なるがゆえに、新しい統計理論が必要となり、また計算機が高機能であるがゆえに、計算負荷は高くてもよいからより有効な結果を与える新しい計算手法が必要となっている。結果、現在統計分野においては新しい統計・計算手法の開発が急務の課題となっている。本書は、「数理・計算の統計科学」に関連するそのような11のホットトピックが11人の著者によって書き下されることによって構成されている。

本書の対象者についてであるが、学生や院生、本書の内容にあまり馴染みのない研究者、基本的な統計学を学びそれに興味を持った人、が対象の中にいることは間違いない。統計学を学ぶ者にとって、幅広い知識を持つことは深く理解することと同等に重要なことである。統計学の新しい流れにのっているたくさん話題について、基本的なところから丁寧に書かれている本書は、幅広い視野を確実に与えてくれるであろう。ちなみに、本書に書かれている統計的手法を実際に扱うことを検討している人にとっても本書は貴重なものとなる。というのも、新しい話題には、新しいが故にうまくまとめられた文献が少なく、特に日本語の文献になるとほとんど存在しないからである。

11人の著者はもちろんそのホットトピックの分野のスペシャリストであるが、その中には、その分野の研究を突き進めている若い統計学者だけではなく、日本の統計学界をまとめているような統計学者もいる。この本では、そのような著者が統計学について大きな視点から論じたりしている。実は上で記述を引用したのはその例を見せたかったからである。こういった記述の内容というのは、頭の中で漠然と感じていながらもなかなか思い切って断言したり書いたりできないことである。散りばめられているそのような表現を通し、統計学とはどのようなものであり21世紀の中でどう発展されていくべきか、ということを読者ははっきりとイメージしていくことができるであろう。

それでは各章について簡単にみていくことにしよう。第I部「統計数理と統計計算への誘い」を構成する第一章から第三章は、上で引用した何箇所かの記述を見ればわかるように、大きな視点から統計学そしてその未来を論じることを中心としている。その上で第一章では、数理統計学のクラシカルな研究が今後もやはり必要となるであろうということ、確率分布の話題を通して述べている。また第二章では、計算統計学について概観しているが、特に状態空間モデルの有用性を強く印象づけさせてくれる。第三章ではいくつかの文献を引用することにより、統計学の未来と挑戦的課題について記すことから始まっており、これは必見である。そして、小標本の高次元データに対する新しい漸近理論の話題を通じ、新しい統計理論の必要性を示唆している。

第II部「統計数理の展開と統計科学」では、ある特定の分野の動向を論じる形となっているが、第六章だけ若干趣が異なる。そこでは時系列分析という広いテーマを扱っており、その中で主流となってきた研究を、第I部のような大きな視点から非常にコンパクト

にまとめている。一方、第四章では、パネルデータなどに対する基本モデルである線形混合モデルを、官庁統計の分野などで需要が高い小地域推定で用いる話題を、第五章では、従属性があり正規分布ではない多変量分布を扱うための道具であり、生存解析や数理ファイナンスなどの分野で需要が高いコンピュータの話題を、第七章では、やはり数理ファイナンスなどで需要が高い、確率微分方程式で記述される連続時間確率過程の母数推定の話題を扱っている。これらは近年関連研究が多いと私も感じる話題であり、基本的なところから日本語でまとめられているものが他にはほとんどないという意味でも貴重な論文といえるであろう。

第III部「統計計算の展開と統計科学」では、計算統計学の基本といえる話題を扱っている。第八章ではデータのリサンプリングを繰り返すことで例えば予測や推測の精度を評価することのできるブートストラップ法を、第九章では不完全なデータに基づく推定や潜在変数を含むモデルの推定のために用いられ統計力学の分野で発展した手法とも組み合わせられて進化し続けているEMアルゴリズムを、第十章では複雑な高次元確率分布のシミュレーションを実現する方法であり例えばベイズ推定において重用されるマルコフ連鎖モンテカルロ法を、第十一章では非常に広いクラスの状態空間モデルによる解析を可能にした粒子フィルタの話題を扱っている。最初の三つについては、既に歴史がありこれまで大きく発展してきたものが現在さらに重要視されている話題であるのに対し、最後の一つは比較的最近発展しはじめたといえる話題であり、その意味で第十一章は第II部のものに近いかもしれない。

最後に本書における小さな不満を述べておこう。それは、21世紀の統計科学として取り上げられるにふさわしいホットトピックがまだまだ存在しているであろうにもかかわらず、それらに関する言及がないことである。例えば私が少し考えただけでも、変数選択の話題や多重検定の話題など、データの規模が膨大となったが故に近年開発が著しく、この本の内容に適合しているように感じる分野を思いつくのである。もちろんそのような分野をすべてうまく取り上げ、章を割くのは極めて困難なことであり、本書の監修者もあとがきにおいて「意識的に網羅的であることを避け」と述べている。ただ、シリーズでどのような話題を取り上げるかという議論がきっとあったはずである。本当は取り上げたかったものの断念したホットトピックの中でどのようなものがあったのかについて、いつか読者が知る機会があることを望むしだいである。ただ、逆に考えると、こういったことは統計学の分野にとってよいことであろう。なぜなら、本シリーズで取り上げられないほど、まだまだホットトピックがあるということだからである。そのことを意識しながら本書を読めば、21世紀において新しい統計科学が発展していくであろうことを、読者は一層強く感じることはできるのではないだろうか。