

大塚 淳 著

## 『統計を哲学する』

名古屋大学出版会、2020年、A5版並製、248ページ、3,200円+税

「統計学と哲学」を敬遠してしまいそうな読者には、少し長いが「データサイエンティストのための哲学入門、かつ哲学者のためのデータサイエンス入門」がより適切なタイトル、と著者が説明していることを指摘しておこう。またPC（計算機）上で中身の理解なしにあるソフトウェアのどのボタンをクリックしたらよいか、だけに飽き足らない若手データサイエンティスト、大学で履修した「統計学」の授業では習わなかった「ベイズ」、「因果」、「RCT」、「深層学習」などに関心のある読者には、それほど分かりにくくなく統計学の主要な哲学的話題を扱う本邦初と思われる文献、と紹介したい。

本書の内容を要約しておこう。全体の内容を短く概観した序章「統計学を哲学する？」に続き、第1章では「現代統計学のパラダイム」、第2章では「ベイズ統計」、第3章では「古典統計」、第4章では「モデル選択と深層学習」、第5章では「因果推論」と続き、終章は「統計学の存在論・意味論・認識論」となる。記述統計の説明から始まる第1章では確率モ分布と推測統計の考え方をダイジェスト的に説明している。最後に18世紀の哲学者ヒュームの「自然の斉一性」、自然種（natural kind）と確率種という哲学用語がさらなる展開を示唆している。第2章と第3章では統計学では古くから続いている論争的な、主観確率に基づくベイズ派統計と頻度確率に基づく古典派統計（非ベイズ）を巡

る幾つかの論点を扱っている。事前分布と事後分布に基づく前者を内在主義的認識論、さらにNyman-Pearsonによる仮説検定論と応用を主要な題材として、後者を外在的認識論との理解を示している。つまり異なる哲学的立場として統計学アプローチの違いを説明しているのである。第4章では現実モデルと予測モデルによる統計モデル選択と深層学習論の哲学的問題、プラグマティズム哲学との関連性を議論している。第5章では哲学からみた反実仮想（counterfactual）モデルなど、近年での応用では重要な統計的因果推論を議論している。終章では統計学の理論の意味と有用性について哲学的基礎をまとめている。

普段は全く哲学（Philosophy）とは縁遠い存在と考えている本評者には本書の内容は全体として新鮮である。若手のデータサイエンティストにはそれほどなじみがないかもしれないが例えば「ベイズ統計」と「非ベイズ統計」については古くから繰り返し議論されてきている問題である。もし今まで全く学んだことがなければ、ほんの少し立ち止まって考える良い機会かもしれない。（例えば一昔前の次の文献は簡単にネットからすぐに手に入れられる：鈴木・国友編（1989）「ベイズ統計学とその応用」、東京大学出版会、<http://www.kunitomo-lab.sakura.ne.jp/Bayes1987.pdf>。）

なお、著者はこの統計学を巡って異なる立場

---

---

「ベイズ統計と非ベイズ統計」という古くて新しい問題について、「…論争に決着をつけることで、またそこに新たな薪をくべることでなく…」と述べている。その理由として哲学者ヒューム以来、現代までの哲学で問題としている内容と符合する、という著者の指摘は新鮮だろう。本書の主要な論点は特に最近になり「データサイエンス」、「ベイズ統計」の重要性が叫ばれるようになってきている現代の日本でも重要な指摘と思われる。近代の統計学の基礎をつくった著名人 Thomas Bayes (1702-1761), (David Hume (1711-1776)), Francis Galton (1822-1911), Karl Pearson (1867-1936), Jerzy Neyman (1894-1981), Ronald Fisher (1890-1962) は皆、英国で活躍した学者であることに改めて気づかされるが、これは偶然ではないのである。

本書の後半では統計学・統計分析のより最近の話題について哲学的に考察されている。因果推論が取り上げられるのは当然としても、「モデル選択」や「深層学習」が取り上げられているのはやや意外であるが、真理の世界と予測の世界の相違について、プラグマティズム認識論や徳認識論と云った、哲学の門外漢には理解しづらい観点から解釈を試みている。こうした議論が成功しているのか否か、正直のところ評者には十分理解できなかった部分が少なくない。しかしながら、いずれにせよ現代の統計学でのホットな話題の解釈や応用につながる事項ばかりであり、その評価は当然にも論争的、哲学的な問題として取り上げる試みは敬服に値しよう。

本書を一読して評者のこれまでの経験から気がついた幾つかの関連する論点を挙げておこう。まず細かな論点として F. Galton (1886) の図の回帰直線の説明は正しくない(注:これは中西寛子・統計数理研究所特任教授の指摘)。楕円の長軸を固有ベ

クトルと解釈すると、長軸は主成分に対応するが回帰線には対応しない。(よく間違えることがある) 評者を含めて注意深さが要求されよう。第2に Ramsey-Savage 流の事前分布と主観確率の説明に関連して「ダッチブック」の説明が興味深い。実は現代ファイナンス (finance) 分野を学ぶとすぐに分かることであるが、1970年以降に本質的に全く同じ概念が世界中の金融市場で実際に利用されている。ここで長い間、統計家や哲学者が「ダッチブック」として議論している確率は金融市場での日常的な「裁定機会のない確率測度、不確実性下の公平な価格」に対応し、毎秒のように計算されているのである。第3には因果性に関連して1950年代~1980年ごろ米国の計量経済学 (econometrics, 例えば雑誌 *Econometrica* を見られたい) 分野での主要な話題として causal ordering (因果順序), simultaneous equation (同時方程式), endogeneity (内生性), exogeneity (外生性), pre-determinedness (先決性), instrumental variables (操作変数), Granger-Causality (グレンジャー因果性)などを巡る議論が盛んであった。近年ではやや下火であるが、最後の例を除きなぜ下火になったのか経済系の統計家には興味ある話題だろう。

なお、最後の章で著者は「データ解析に携わる人にちょっとだけ哲学者になり、哲学的思索を行う人にちょっとだけデータサイエンティストになってもらう」ことが本書の目的と説明している。様々な分野において統計データに関わる多くの読者が「データサイエンス時代の統計学の意味」を考察する1つのきっかけとなればと考え本書の一読を推奨したい。

評者: 国友 直人・くにとも なおと  
(東京経済大学研究員、東京大学名誉教授)