

ブラックボックスでない 分析手法を求めて

村上 隆

中京大学 文化科学研究所 特任研究員

私は計量心理学者で、これまでの時間の多くを、分析手法の理論的検討とプログラミング、データ分析作業と結果の解釈に充ててきた。読書にかけた時間は長くないが、それでも仕事に行き詰まるたびに開き、まさに韋編三絶となった書物は少なくない。それを3冊に絞るのは難しかったが、ここでは、私がおの人の人柄や仕事ぶりに直接触れる機会があり、ロールモデルとして尊敬している方たちの著作であること、加えて現在も入手が容易で、若い読者にとっても有益であると信じられることを基準にして選んだ。

日本の多変量解析の黎明期

『多変量データ解析講義』

水野欽司 著

朝倉書店 1996 (2018年刊行の新装版あり)

本書は、1960年代から著者が勉強会や大学院レベルの講義用の資料として作成してきた、ブルーコピーや謄写版印刷の資料にもとづく。私が勉強したのもそうした未公開資料であり、韋編三絶どころかボロボロになって四散した。資料は当時から「幻の多変量解析」と呼ばれており、それを知った編集者の熱意と努力によって刊行に至ったものである(その経緯は、編集に当たった後輩たちによる「あとがき」に詳しい)。

今日のスマホにも遠く及ばないほど低い性能の「電子計算機」、そのソフトウェアも自作するしかないという環境で、多変量解析の手法がどのように導入されていたのかという、いわば先達の労苦が本書からは読み取れるが、ここにあるものはそれにとどまらない。すなわちこれは、かつては存在したプログラマーと実務家を兼ねた研究者の手による、一点一画もゆるがせにしない分析手法の解説なのである。

私はまず学生として、後に部下として、著者から実に多くの叱責や忠告をもらった。自作の因子分

析のプログラムを走らせて得意になっていたときにかけられた「君には因子分析が本当にはわかっていない。使い込んでいないから」という言葉は、いまだに耳に残っている。

初等的な数理による独創的な業績

『Least squares optimization
in multivariate analysis.』

Jos M.F. ten Berge 著

DSWO press 1993

(ウェブからダウンロード可能 (2019年9月現在)
<http://kiers.webhosting.rug.nl/>)

最小二乗法によって定式化される9つの多変量解析の手法を「統一的なフレームワーク」で扱ったテキストである。その特徴は微分を一切用いず、ベクトル関数や行列関数に対して一般化された「平方完成」によっているところにある。著者の膨大な研究論文も、ほぼ常にこの原理に基づいており、ここに複雑な方法をより初等的で見通しのよいやり方で読者に理解させようとする本人の強い意思が感じられる。私自身、本書程度の内容にはすでに十分習熟しているつもりで読み始めたのであるが、慣れ親しんでいるはずの線形代数も実は「使い込んで」いなかったことを痛感させられたのであった。私は本書が出版された直後の1995年、長期在外研究員として、著者が教授を務める Groningen 大学に滞在し、その仕事ぶりを身近に見ることができ、気鋭の若手であった Henk Kiers を加えた3人で論文を執筆する機会を得た。その論文の基本となった命題の証明を、金曜日の午後中かけて下手な英語で何とか書き上げ、月曜日に出勤してみると”Dr. Murakami’s proof is really brilliant, but…”で始まる、はるかにエレガントな(もちろん初等的な)証明が書かれたA4のペーパーが1枚、メールボックスに入っていた。

その後の交流も含め、私の思考のスタイルはここ



で固められたと思っている。それは確かに自分の強みとなったと思うが、他方で限界にもなったようだ。

理論と実際の橋渡し

『医学・薬学・健康の統計学
—理論の実用に向けて』

吉村 功・大森 崇・寒水孝司 著
サイエンティスト社 2009

第1著者の吉村さん（ご本人は「先生」と呼ばれることがお嫌いだった）とお会いしたのは、1980年代の名古屋大学大型計算機センターでのことである。当時はまだカード・デッキを持ち込んでのバッチ処理が中心で、センターには自然発生的な文理融合コミュニティとでもいうべきものが成立していた。その頃、センターの広報委員であった私が、恐る恐る講習会の講師をお願いしたのが最初の接点だったと記憶する。

吉村さんには、60年代に執筆された『数理統計学』という名著があり、（私に理解できる範囲で）お世話になっていたが、その一方で、統計学の誤用を指摘して公害、薬害の企業責任を明らかにしたいいくつかの論文にも大きな感銘を受けていた。つまり、当時の私には雲の上の人だったのであるが、その後、私の所属する部局の非常勤講師をお願いしたり、学会の大会委員長に就任してもらったりと、個人的な接点ができた。私がオーガナイザーを担当

した学会のシンポジウムで登壇していただいた時の演題が、まさに「理論と実際の橋渡し」であった。

本書は、90年代に東京に移られてからのお弟子さんたちとの共著だが、吉村さんの本領が発揮された好著である。第1章「統計学の役割」で複数の実データを示しながら、解明されるべき問題と統計手法の必要性を説明した上で、第2章～第5章では、確率変数を用いたモデル化、確率分布、統計的推定、そして、あくまでも頻度主義の立場を堅持した統計的検定に関する要領を得た解説がなされる。続いて第6章～第8章、「よく使われる手法の理論的解説」、「理論の実用に向けて」、「実際のデータ解析」へと進む。基礎的理論の章では、特に初学者が躓きやすい部分について、通常のテキストには見られない絶妙の説明がなされる。これまでの「もやもや」が晴れる思いがする読者も多いことだろう。

だが、何とんでも圧巻は第6章～第8章である。たとえば、誤解・誤用が多いため批判されることの多い p 値について、検定論の基本に戻ってその意味と誤用の理由、さらには p 値の健全な用法までがていねいに説明される。多数の検定を同時に行うことによって検定が「甘く」という多重性の問題もそれと関連する。統計学のプロは保守的な安全策を提案しがちであるが、それがしばしば公害企業の責任をあいまいにしてきたことを知る著者は、状況によってはより大胆な基準の使用を提案する。このあたりの論述は、スリリングでさえある。なお、出版社のウェブサイトにも、各章末の演習問題の解答例と、詳細な正誤表がある。

近年、統計学とその応用分野には急激な変化が起こっている。その変化のうちには真の進歩と呼べるものもあるであろうが、単なる流行にすぎない部分も混在しているように思われる。この時点で、上記3冊のように論理的に一貫し、かつ血の通った著作の存在意義は決して失われてはいないものと信じる。