

日本語版への序

Preface for the Japanese Version

Correspondence Analysis in Practice (Third Edition) の日本語版（『対応分析の実際—基礎・応用・展開—』）に序文を寄稿できることを嬉しく思います。翻訳を担った藤本一男さんの熱意と努力に感謝します。

対応分析（CA：correspondence analysis）は、1974年のMark Hillの論文「対応分析：放置されていた多変量解析法」*1以降、また、対応分析の創始者として知られているJean-Paul Benzécriのもと、パリで私が1970年代の博士論文を執筆して以降、長い道のりを歩んできました。私はこの手法のいくつかの開発に関与してきましたが、Jean-Paul Benzécriの「弟子」の1人として、Benzécriのデータ解析の見解を英語圏の研究者に公開した最初の研究者の1人に過ぎません。

ここ数年、対応分析に関する書籍の刊行や論文の発表で「（対応分析の）言葉を広めている」多くのBenzécriの弟子たちがいます。それは、Ludovic LebartやFionn Murtaghたちです。多くの研究者がいますが、ここでは2人だけあげました。また、Benzécriとフランス学派の他にも、Jan de Leeuwとオランダ学派として、Willem HeiserとPatrick Groenenがおり、多くの出版物を刊行しています。そしてまた、林知己夫による日本学派があり、（英語圏での）多数の書物を著している西里静彦もいます。Benzécri、Leeuwおよび林知己夫は、全く独自に理論化を行っていましたが、三人は統計分析に対する同じビジョン、哲学を持っていました。それは、分析は先験的モデルによって行われるのではなく、本質的にデータ主導であるべきだ、というものです。

林知己夫のケースは、特にこの日本語版に関連しています。大隅昇は林に対する献辞に「（林にとって）多変量解析はガウス分布に基づく不必要に複雑な算術演算に過ぎず、実用には限界があり、研究論文にしか適用できない」と書いています*2。林は（多変量解析ではなく）「多次元データ分析」という言葉を好んでいました。彼のアプローチは、Benzécriの5つのデータ分析原則と完全に一致し

*1 訳注：Hill, M.O. (1974) *Correspondence analysis: a neglected multivariate methods*, *Applied Statistics*, 23, 340–354. オンラインで読める。https://www.jstor.org/stable/2347127

*2 訳注：https://ifcs.boku.ac.at/site/lib/exe/fetch.php?media=newsletter_archive:ifcs-newsletter-24.pdf
日本分類学会会報 2003 に大隅が寄稿している「林知己夫先生を偲んで—分類からデータ科学に向けて—」も参照のこと。http://bunrui.jp/pdf/kaiho26.pdf

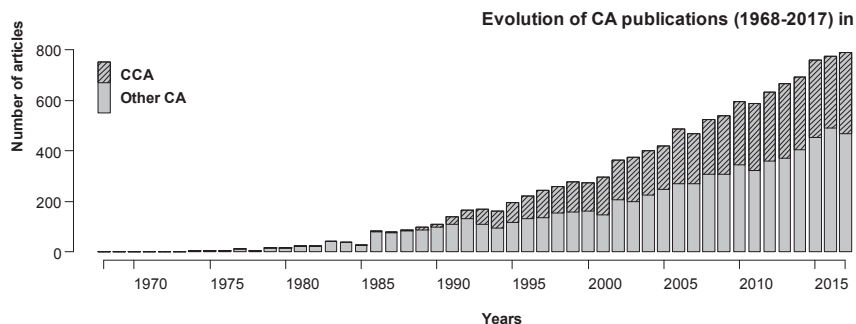
ています。

1. 統計は確率ではない
2. モデルはデータに従う必要があり、逆ではない
3. 情報は可能な限り最大の次元で扱われるべきである
4. 複雑なデータ、特に社会現象に関するデータを分析するためにはコンピュータは不可欠である
5. コンピュータを使用することは、自動計算の到来前に想起されたすべての技術を放棄することを意味する

本質的に同じ哲学に依拠したこれらの学派は、異なる言語、語彙および用語を使用していましたが「対応分析」という名称が他のものよりも長く持続していると思われますし、それは多次元データ分析の中心的技術として統計学文献に受け入れられてきました。林知己夫はそれを「数量化手法」と呼んでおり、それには番号付きバージョンとして、数量化Ⅰ類、Ⅱ類などがあります。対応分析は正確にはこの中の数量化Ⅲ類であり、このデータ分析「手法」の実用的なブランド名ではありませんが、林の方法論的な考えを示しています*³。Benzécriは、林の考えが彼自身のものに先行していると認めています*⁴。ただ、それが今日普及している多次元データ分析の考え方ではないことも指摘していました。同じ理論であるにもかかわらず、林は彼の理論の解釈は異なる」と主張しました。Benzécriは、分析結果の視覚化とそのデータの構造を可視化することを好みましたが、林は西里の「双対尺度法」アプローチの基本的な考え方と同じように、カテゴリカル・データの数値化というアイデアを優先しました。それは、オランダ学派の「非線形多変量解析」に近いアプローチでもあります。

以上のことは、歴史的で異文化間の非常に興味深い比較ではありますが、ここでは、現在と将来のCAを見てみましょう。CAの出版物の数は年間約7%の割合で増加しています（Web of Scienceの出版物の数はExhibit 0.0にプロットさ

Exhibit 0.0 ▶▶▶
対応分析関連出版物の進化



*³ 訳注：数量化Ⅰ類という名称は林本人ではなく鮑戸弘の命名による。

*⁴ 訳注：Hotelling, Hの主成分分析。

れています)。近年の、これらの出版物の大部分(約40%)は、Cajo ter Braakが1986年にEcology誌に発表し、現在、生態学者によって広く使用されている正準対応分析(CCA)を用いたものです*5。

いうまでもなく、これらの論文には様々な分野で対応分析の利用が記されており、この方法に関する応用知識は広範なものになっています。約15年前、私は(対応分析の)主題には理論的発展の余地はほとんど何も残っていないと考えましたが、これは正しくありませんでした。例えば、寄与座標(Contribution Coordinate)を使ってマップの解釈をしやすくするというアイデア(第13章)や、またCAと組成データ分析(第22章)を連携すること、また一定した(データ)ストリームから到来する巨大なデータの分析への挑戦など(Rのidmパッケージに実装されているAlfonso Iodice d'EnzaとAngelos Markosの最近の研究を参照)、と多くの興味深い展開がありました。そうしたことを踏まえれば、Exhibit 0.0の指標が示す良好な傾向から、CAとそのすべての変種はまだ理論的にも、実践的にも、長い寿命を持ち、幸せな未来がまっているということを期待できるでしょう。

今一度、藤本一男さんの日本語版への尽力に感謝を表明します。また、このプロジェクトに関心を持ってくださったオーム社に感謝します。そして、日本の読者がこの多次元データ解析という可能性に満ちた方法からのメッセージを受け取ってくださり、その研究と教義の恩恵に浴されることを願っています。

Michael Greenacre

アソス(トルコ)、2020年9月

*5 訳注: Ter Braak, C.J.F., (1986) *Canonical Correspondence Analysis: a new eigenvector technique for multivariate direct gradient analysis*, Ecology 67, 1167–1179. なお、Cajoのアイデアを基礎にしたRのパッケージがveganとして公開され開発が続けられている。第27章以降で触れられている。