

日本における「対応分析」受容の現状を踏まえて、EDA（探索的データ解析）の中に対応分析を位置付け、新たなデータ解析のアプローチを実現する

藤 本 一 男

概要

最初に、日本における「対応分析 CA」の受容状況を、ブルデュー『ディスタクシオン』の翻訳(1990年)をめぐった時代的制約という視点で整理した。計量手法(=CA)について注として参照されている文献の日本語訳が刊行されたのが1994年であったため、分析手法名の誤訳に止まらない訳語上の混乱があった。この混乱は、日本語版よりは少ないとはいえ、英語訳(1984)でも同様であったことを述べた。

これを受けて、「対応分析」の機能を理解する方策を、EDA(探索的データ解析)との関係でみていくことを提案した。それは以下の理由による。

- 1) CAは、行変数と列変数の関係を可視化する。これは、EDAの基本的操作である「可視化」(Tukey1977)の中に位置付けることができる。これによって、クロス表に対するカイ2乗検定の先に一步踏み込んだ解釈が可能となる。
- 2) さらに、EDAの原理を述べた Tukey1962で提示されているアプローチである計算機統計学的なアプローチを組み込んだ分析手法が関係する。ここでは、R.A.Fisherの伝統が再確認されているが、そのアプローチの具体的な形態が、CAやMCA、PCAを要素機能とする幾何学的データ解析(GDA)である。そして、それこそが、ブルデューが『ディスタクシオン』で採用した機能そのものであった。

こうした視点にたつて、GDAの普及が、ブルデュー社会学に止まらない、新たなデータ解析のアプローチを実現するものであることを述べた。

キーワード：対応分析(CA)、多重対応分析(MCA)、探索的データ解析(EDA)、幾何学的データ解析(GDA)、構造化、アブダクション

1 はじめに：「対応分析」認知の現状

1.1 対応分析は認知されているのか

社会学で「対応分析」(Correspondence Analysis: CA)といえば、ピエール・ブルデューの「ディスタンクシオン」での使用が想起される。しかし、こうした認知はまだ一部の研究者に限定されているし、ブルデュー研究者においても共通のものにはなっていない。

Le Roux と Rouanet は、次のように述べている。

「対応分析 (CA) は認知されるようになった。しかし幾何学的データ解析 (GDA) は発見されるのをまわっている。」(Le Roux & Rouanet 2010:4=2021:5)
幾何学的データ解析 (Geometric Data Analysis) とは、対応分析 (CA) や多重対応分析 (MCA)、そして主成分分析 (PCA) を基本的要素とした分析手法のことで、ブルデューやTベネットの調査分析において中心的な役割をになってきた Brigitte Le Roux と Henry Rouanet によって発展させられてきたデータ解析の考え方/手法である¹⁾。

対応分析には50年の、そして日本で発展してきた対応分析とは数理的に等価であると言われる数量化手法(とりわけIII類)にはすでに60年以上の歴史があると言われている。それではこの手法はLe Roux と Rouanet がいうように、十分に認知されているのだろうか。発見をまたれているのはGDAだけなのだろうか。

1.2 ブルデューの「ディスタンクシオン」の翻訳での対応分析をめぐる混乱

社会学と対応分析の関係が語られる時、ピエール・ブルデューと彼による「変数の社会学」批判が引き合いにだされる。それは、時に、次のような挑発的なフレーズとともに参照される。

「私は対応分析をよく用いますが、それは対応分析が関係論的な手続きであり、私にとっては社会的現実を構成するものを十全にあらわす哲学だからです。それは、関係論的に思考する手続きであり、私が界概念とともに試みていることです。(Lebaron 2010:102、磯 2020:130)」

ここに表現されている多重対応分析への注目²⁾は、必ずしも理解されてこなかった。確かに、『ディスタンクシオン』はブルデューの基本文献なのだが、そこでの計量手法、つまり対応分析と彼の社会学概念が切断されているから

である。

それは、ブルデューが用いた計量分析の手法(対応分析)が、「照応性の分析」と訳されていることに加えて、対応分析の基本的な概念、その意味では、ブルデューが「界」の概念と対応させている多重対応分析の重要概念である「慣性」「因子」の翻訳で失敗している。

対応分析において「慣性」とは伝統的統計学でいうところの「分散」であり、訳註で説明されるような「比喩」などではない³⁾。

さらに原文で「factorielle」とあるものが、すべて「因子」と翻訳されているが、これは「座標軸（もしくは主軸）」と訳されるべきものである。第5章の注の英訳は「This means that the first factor in the factorial analysis corresponds to the second dimension of the social space and the second factor to the third dimension.」となっている（Oxford版 p580 の(6)、Routledge版では p574 の8）。日本語訳では、「(8) このことは、因子分析の第一因子が社会空間の第二次元に、第二因子が第三次元にそれぞれ対応しているということを意味している」となっている。

原文は、以下の通りである。

「8. – Cela signifie que le premier facteur de l'analyse factorielle correspond à la deuxième dimension de l'espace social et le second à la troisième.」

つまり、日本語で「因子分析」とされている部分は、「l'analyse factorielle」であって、社会学や心理学で多用される（探索的）因子分析ではないのである⁴⁾。なお、この「因子」をめぐる解釈については、Lebart 2008 の 2. Principaux axes methods や大隅他 2021 の用語解説に詳しい。大隅他 2021: 162。

こうした基本的な分析手法についての混乱が解消されないまま 30 年近くが経過したために、ブルデューがおこなった、データの構造化、つまりある変数群で空間を形成し、それとは異なるサプリメンタリ変数（これは GDA では構造因子 structuring factor と呼ぶ）を投影する、という多重対応分析を用いた構造化が理解されてこなかった。そのために、多重対応分析の使い方をめぐっても誤解、混乱も生じている。北田 2017 でのこの点での混乱については、藤本 2019 で指摘してある。

なお、英訳版でも、原文の対応分析を意味する「L'analyse des correspondences」が、あるところでは「correspondence analysis」として訳され、あるところでは説明的に「Analysis of the correspondence」と訳されている。こうした訳語の揺れは、「L'analyse des correspondences」が「Correspondence Analysis」という一つの統計的処理手法を指していると認識していないことに起因しているだろう。

だが、こうした「無理解」「誤訳」には、時代的な制約という面もある。

1.3 英語圏、日本語圏への「対応分析」の伝播

フランスにおいて、「対応分析」は、J.P.Benzécri によって開発されその弟子たちによって発展させられてきた⁵⁾。

フランスで発展してきたこの手法の体系的な解説が英文で登場するのは、1984年である。Richard Nice による *Distinction* の英語訳の出版は、1984年であることを考えれば、その時点の英語圏で、*Correspondence Analysis* はまったくの無名であったといっても差し支えないだろう。

では日本ではどうか。1984年に刊行された2冊の英語文献のうち Lebart らによる “Multivariate descriptive statistical analysis” は、1994年に大隅昇たちによって翻訳されている(大隅・L. ルバル他1994)。なお、この英語版の原著は、1977年にパリの Dunod から出版されている L. Lebart, A. Morineau et N. Tabard. *Techniques de la description statistique. Méthodes et logiciels pour l'analyse des grands tableaux*, Paris, Dunod, 1977 であって、『ディスタクシオン』第5章の注の(2)にベンゼクリの文献とともに「分析方法については…」としてあげられているものである(普及版p931)。この英語版のサブタイトルは、*correspondence analysis and related techniques for large matrices* である。

『ディスタクシオン』の翻訳(この部分は5章なので、発行者が藤原書店にうつってから)が1990年であるから、流石にまにあっていない。不明な箇所は Richard Nice の英語版を随時参照したと「訳者あとがき」にはあったが、対応分析については英語圏でも認知がこれからという手法ゆえに、見落とされているのだと思う。最初の英訳版は Harvard 版1984で、Rutledge Classic で Tony Benet の解説がついたものが出たのが2010である。しかし、1984版はいざ知らず2010版でもこの Lebart たちによる文献の英語版は参照されていない。

1.4 時代的制約は克服される段階にある

以上は、時代的な制約ということで事情を理解することも可能かもしれない。しかし、2020年12月に刊行された『ディスタクシオン』普及版においても、上述の部分はそのままであった。つまり、こうした状況が、日本における「対応分析」認知のされ方を象徴していると言っているだろう。

残念ながら、冒頭に紹介した、Le Roux と Rouanet がいうような状況には至っておらず、GDA だけでなく対応分析も、発見されるのをまっているの

が2021年の日本の現実である。

しかし、2021年9月の時点で、対応分析についての基本的な文献がそろったとあってよい。列挙してみると以下ようになる。

翻訳年

1994年 Lebart 他『記述的多変量解析法』(原著 1984, 1977)

2015年 Clausen『対応分析入門』(原著 1998)

2017年 Benett、他『文化・階級・卓越化』(原著 2009, 2010)

2020年 Greenacre『対応分析の理論と実践』(原著第3版 2017)

2021年 Rouanet & Le Roux『多重対応分析』(原著 2010)

1.5 現状についての補足：対応分析に言及した論文数の推移

ここで、別の側面から対応分析への言及を確認しておきたい。

以下の確認は、国立情報学研究所の CiNii article をつかって、年次ごとにキーワードヒットをカウントしそれを Plot したものである。

それをみると、対応分析は使われていないわけではない。先に時代的制約と書いたことと矛盾するようだが、対応分析に言及した論文は、確実に増加している。ただし、ブルデューと対応分析の組みあわせでのヒットは非常にすくない。他方、対応分析単独でのヒットは、急速に増えている。しかし、これは、計量テキスト分析との関係で増えているのであって、計量テキストに関連しない対応分析のヒット数はすくない⁶⁾。

ここで検討した、ブルデューとの関係や計量テキスト分析との関係で対応分析に言及する、という使い方は、いわば、対応分析の応用領域である。この分野は、今日ではその分野の専門家には自明な使い方であっても、それ以外の人たちにとっては、縁遠いものでしかない。ここで縁遠いというのは、たとえば、クロス表の分析において、 χ 二乗検定を知らない社会学者はいない。しかし、対応分析は、なにか遠いものである。そういう意味での縁遠さに、認知のされ方の現状が表れている。

2 CAは数量化による探索的データ解析 EDA の拡張を行う

続いて、CAをEDA(探索的データ解析)に位置付けながら、その機能を確認していきたい。その理由は、第1に、CAは、行変数と列変数の関係を可視化する。これは、EDAの基本的操作である「可視化」(Tukey1977)の中に位置付けることができる。これによって、クロス表に対するカイ2乗検定の

先に一步踏み込んだ解釈が可能となるからである。さらに、第2にEDAの原理を述べたTukey1962で提示されているアプローチである計算機統計学的なアプローチを組み込んだ分析手法が関係する。そこでは、R.A.Fisherの伝統が再確認されているが、そのアプローチの具体的な形態が、CAやMCA、PCAを要素機能とする幾何学的データ解析(GDA)である。そして、それこそが、ブルデューが『ディスタンクオン』で採用した機能そのものであったからである。

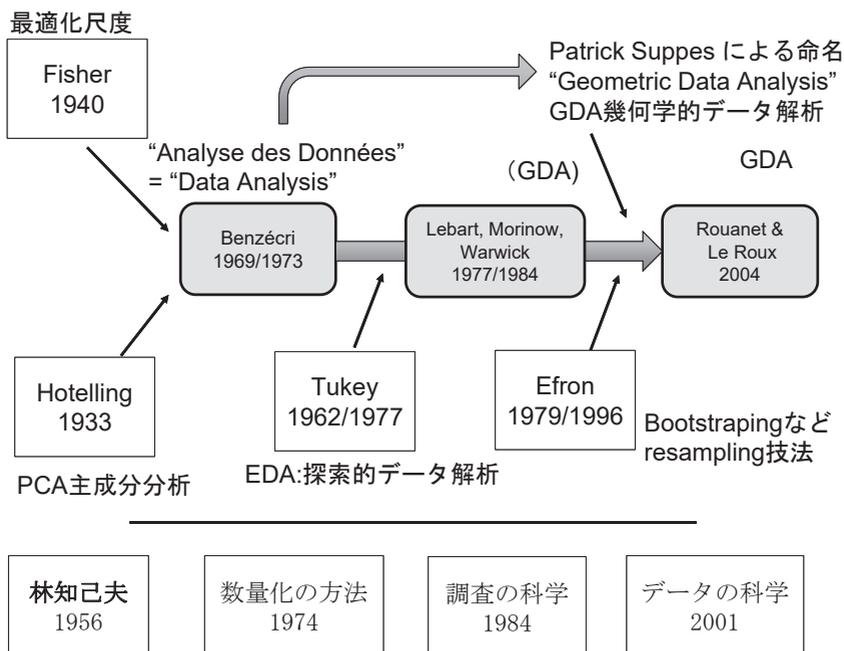


図2 「対応分析」がBensécriによって開発されてからそれを用いた方法にGDA（幾何学的データ解析）という名前がつくまでの概略図。下部に示したのは、林知己夫の数量化手法とデータの科学までの関係。

2.1 EDAの基本ツールに位置づけられる対応分析

対応分析の基本的、基礎的機能を理解するために、EDA（探索的データ解析）の中に位置づけられた対応分析をみていく。

参照するのは、Michael. Freindlyの“Discrete Data Analysis”CRC、2016⁷⁾で

ある。EDA (探索的データ解析) とは、J.W.Tukey が提唱したデータ解析の手法であり、CDA (確証的データ解析：仮説検証アプローチ) と対比させて説明されることもある。データ・サイエンス分野では、データ解析に入る前にデータの分布を確認すべし、という文脈で紹介される (ブルース 2018=2018、ヒーリー 2019=2020、など)。

Tukey の EDA に関する文献の中心にあるものは、Tukey 1977 での EDA であるが、そこでは、幹葉図 (Stem-Leaf Plot)、箱ヒゲ図 (boxplot) などのデータを可視化するためのいくつかの技法が提示され可視化の側面が強調されている。

Friendly は、Friendly 2016 で離散データ (カテゴリカルデータおよび計数データ) の可視化とモデリングを扱っているが、この第二章が、「2元分割表」(いわゆるクロス表)、「Mosaic plot」⁸⁾、そして「対応分析」という内容で、探索的、仮説検定手法というタイトルである (Exploratory and Hypothesis-Testing Methods)。

なぜ、ここに、クロス表と Mosaic plot と対応分析がならんでいるのだろうか。それは、対応分析の原理を考えると納得がいくことである。

対応分析の原理は、クロス表に表現された二つの変数を、行もくしは列からの視点で、それぞれを多次元ベクトルのポイントとして位置付け (プロファイル)、それらの距離を χ 二乗距離として計量し、行ポイントであればその列数分の次元を持つベクトルを、次元縮減によって低次元で近似していく手法である。

対応分析の「対応」⁹⁾ とは、行に対して行う処理と列に対して行う処理が相互に浸透していること、そして、関係が対称であることによっている。こうして得られた新たな座標軸上の座標からポイント間の関係を分析していく。(シンプル) 対応分析が、2変数のクロス表を入力データとするのに対して、多重対応分析は、行に個体、列にカテゴリカル変数を配置した集計表を分析する。手法についての入門的な説明は、Clausen 1989=2015 を参照されたい。クロス表に対して、カイ二乗検定を行うところまでは、どの教科書にも書いてあるところである。しかし、帰無仮説が棄却され、変数同士が無関係ではなさそうだ、ということまでは言えても、その変数のカテゴリの関係はわからない。そこにアプローチするのが対応分析の役割である。

こうして、クロス表、 χ 二乗検定、Mosaic Plot と合わせて、対応分析が、カテゴリカル・データ分析の基本ツールとして位置付けられることになる。

2.2 数量化手法としての対応分析からみた EDA

対応分析は、カテゴリカル・データに、分析対象であるデータ系が体现する分散が最大化するような係数を各変数カテゴリに割り当てる。つまり、主成分分析と同様の目的で、分散最大化という基準に基づいて係数の付与＝数量化がおこなわれる。この機能を使えば、5点法で取得されたアンケートデータをそのまま、5、4、3、2、1とアприオリな「数量化」処理を経て、回帰分析や因子分析にかけるといふ乱暴なことをしないですむ。このアприオリな数量化に対して、林知己夫、西里静彦は、警鐘をならし続けているが、いっように省みられることはない。

以下は、満足度調査でよくある5点法での選択肢をめぐって林知己夫によって表明されている疑問である。

「回答は、非常に満足している、満足している、どちらともいえない、満足していない、非常に満足していない、…これで回答をとるのは一つの物差しで測ったことになる。…一応ここはそのままにして回答をとった時の物差しはあくまでも当該項目に反応したということであって、満足、どちらともいえない、不満足が、いわゆる物差しのように等間隔の数量的表現をもつものではない。さらにこれらが、1次元的なもので、どちらともいえないが、満足と不満足の間にあることさえも疑問である。…こうした基礎的なことの処理も社会現象を取り扱うとき忘れてはならない。…いまも無反省な数量化ではこうしたことが行われている」(林 1993:7)

林によるこの「無反省な数量化」の問題点は、西里 2007 でも同じように論じられている。(西里 2007:39「ライカート方式、SD方式の弊害」)

このように、対応分析のもっとも根底にある機能は、「数量化」機能である。これは、林知己夫の数量化手法と原理的には同じアプローチであるが、それと EDA について林は以下のように述べている。

「トウキー (Tukey) の exploratory data analysis (1962 年に論を発している) の思想は、われわれの考え方に近いが、自然科学のデータを念頭においているせいか、なまの数量そのものを取り扱うことを考えているので窮屈である。しかし、さすがに核心をついており、われわれの狙いに近いことが行われている。数量化の考えを使えばもっと楽になるのだが、と

私は考えている。」(林 1986=2004:13)

ここで、林が対比しているのは、『調査の科学』『データの科学』で要約したデータ分析手法である。ここでは『行動計量学序説』(林 1993)の叙述をもとに整理しておく。

「通常は、まず単純集計を行い、これを読んで次に構造分析をするというのが常道のようなものであるが、知見を得るという上では、どうも逆のような気がしている。もちろんデータそのものはしっかりと吟味されてあ上での話である(このために単純集計を用いることは多い)。単純集計は実に内容ある情報を示しているのであるが、いくらこれをながめてもその意味を取り出して知ることは困難である。そこで、全問について、あるいは、質問群を用いて、数量化Ⅲ類を用い、どのようなデータ構造が示されているかを把握し、それから単純集計を分類しながら、見ていくと、多くの知見が得られてくるのである。」林 2001:105(3.6 単純集計と構造分析)

こうした視点は、先に紹介した Michael Friendly による「離散データ解析」(Friendly and Meyer 2015b)にもシンプルに表現されている。

2.3 EDA をカテゴリカル・データにまで拡張するために必須となる対応分析

カテゴリカル・データに対しては、対応分析を使って数量化による EDA アプローチが可能である。そこでは、5 件法で採取されたデータが、CA による数量化を行ってみたら、5 と 4 が近接、また 2 と 1 が近接。その中間に 3 という実質 3 件法として解釈したほうがいいものもあるし、また、どちらでもない「3」が、5 と 4、2 と 1 の中間ではなく、別の次元に位置している、ということが明らかになることもある。「どちらでもない」が、中間ではなく、そのようなことは考えてない、気にしてない、という反応を記録した場合など。

こうしたデータに対する事前の精査をおこなった上で、しかるべき、統計処理(モデリング)などをおこなわなければ、取得したデータの構造を歪めてしまっていることにしかならない。

このように、整理すると、カテゴリカルデータに対する基礎的な分析に対して CA が提供するのとは以下のようなこととなる。

- ・ カテゴリカルデータに対する数量化（これによって、適用可能な統計手法は格段に広がる）
- ・ カテゴリカルデータに対する、EDA アプローチの実現

3 CA/MCA を通して見えてくる、もう一つの分析パラダイム

3.1 EDA（探索的データ解析）と GDA（幾何学的データ解析）

このように、EDA の中に位置付けることで、対応分析の基本的機能が明らかになる。次に、EDA と幾何学的データ解析（GDA）の関係から対応分析の位置をみていこう。これは、先に引用した林 2001 の流れにそったものである。

Tukey1977 は、EDA の実践的なマニュアル的な要素が大きく、それだけを見ると、EDA とは、統計処理に入る前の「記述的分析」としての「可視化」である、と理解される。その側面はあるものの、実は、Tukey は EDA の考えを述べた Tukey1962「データ解析の未来」で、更に重要な提案をおこなっている。それは、当時すでに、統計学の標準となっていた、Neyman-Pearson 流の仮説検証アプローチ¹⁰⁾ に対する、帰納的なデータ解析の提示である。EDA とこの CDA（確証的データ解析）が対比的に説明されるのはこの帰納的なアプローチゆえである。

この提案では、1960 年代という企業情報システムでの MIS（経営情報システム）の展開に象徴されるような社会的普及期に入ったコンピューター技術を用いることによって、仮説と演繹によるデータ解析（CDA）ではなく、手に入ったデータと向き合い、それにリサンプリング手法を組み合わせ、母集団を推定していくことが現実的に可能になった、ということを背景としている。

ところで、多重対応分析を中心にした GDA 幾何学的データ解析は、以下のようなステップで構成される。つまり、GDA は Tukey が Tukey1962 で述べた「データ解析の未来」の形の一つの具体例なのである。

「データ表」 → 「ポイント・クラウド」 → 「構造化」 → 「統計的推測」
 構造化データ解析 帰納的データ解析

ブートストラッピングを開発した Efron が、「21 世紀のフィッシャー」(Efron 1996) で、Fisher の理論（の大半）は、リサンプリングという思想を基礎にしていると述べている。我々に馴染みのある手法でいえば、「フィッシャーの

正確性検定」がその実例である。カテゴリ数がふえると、計算時間とメモリの消費が増えるわけだが、これは今日のコンピューター (PC) の能力の拡大からすれば対処できない問題ではなくなっている。

3.2 対応分析 (CA) と多重対応分析 (MCA)

多重対応分析とは、最も基礎的には、2変数の関係を分析する (シンプル) 対応分析の多変数拡張として説明される。対応分析は、2変数のクロス表を処理対象とするが、多重対応分析は、行に個体、列に変数 (多変数) を配置したいいわゆるケースフォームデータを分析対象とする。その意味では、多重対応分析は、対応分析の多変数版である。ここでも、数量化機能は実現している。しかし、注目すべきはその先である。

対応分析は、その原理からサプリメンタリ変数を扱うことができる。これは、質量 (周辺度数) を持たないポイントでも、マップに射影することができる、という対応分析における行と列の相互浸透という特性による。データは、構造を生成するポイント (アクティブ・ポイント) と、その構造の生成には寄与しないが、位置することができるポイントに区分される。これが、分析対象の構造設計である。どのポイントをアクティブに指定し、どのポイントをサプリメンタリに指定するかは、当然分析の目的による。こうして、多重対応分析を用いて構造化データ解析の準備が整う。

3.3 構造化データ解析 (SDA)

あるポイントで構造を生成し (空間の生成)、あるポイントをそこに位置づける。これを構造化データ解析 (SDA) と呼ぶ (Le Roux 2004, 2010=2021)¹¹⁾ こうした構造化によって Le Roux 達が構築しているのは (そして、それがディスタンスシオンのパラダイムであり、トニー・ベネットによる多重対応分析の使用である)、幾何学的データ解析 (GDA) という手法である。

3.4 分散 (慣性) の分解技法としての対応分析

この手法は、以下のように要約できる。実験計画が不可能な観察データに対して、分散分析で実現される分散 (対応分析では慣性と呼ぶ) の分析を実現する。ここで、分散の分解単位は、群としてのサブクラウドである。そして、その群内の分散と全体分散の関係から、その群の全体への寄与 (効果) が評価される。そこでは、交互作用を表現する合成変数 (サプリメンタリ) を投入することで、交互作用の評価も行う。対応分析は、系全体の慣性 (つ

まりは分散)を分解する技法であるが、それにサプリメンタリ・ポイントという空間形成には寄与しないが位置はもつポイントを構造因子として投入することで実現される、もう一つの分散分析(慣性分析)と呼びうる技法なのである。

こう整理することで、もう一つの側面が見えて来る。それは、分散分析というフィッシャーのアプローチの観察データに対する適用という側面である。

3.5 構造化データ解析から帰納的データ解析へ

つまり、EDA(探索的データ解析)と対置されるCDA(確証的データ解析)とは、それぞれ異質なアプローチではあるものの、行為としての分析を考えた場合には、相互に浸透しあうものである。そこにみいだされるものは、EDAが帰納的、CDAが演繹的とすれば、アブダクション(仮説構築/更新)アプローチである。

このアプローチでは、MCAによる記述的統計処理が、従来の「記述統計」vs「推測統計」とは異なる意味付与がされている(分析過程における仮説更新)。

4 結語

本稿の最初に確認したとおり、残念ながら、対応分析(CA)は、回帰分析や主成分分析、因子分析に比べて認知されてはいないのが現実である。しかし、ブルデュー研究者の中には、計量的手法として対応分析を駆使して研究をする人たちが現れているし、また、計量テキスト分析の普及のなかで、対応分析への関心も高まっている。

また、本文で述べたように、まだまだ少ないとはいえ、理論的、実践的、そして事例となる基本的な文献も翻訳され、普及の条件は整ったと言えると思う。

こうしたことは、ひとり対応分析限界のできごととして限定されるものではなく、2016のAPA会長声明としてだされた「p値問題」への一つの手がかりをも提供している。Efronは、先にふれた「21世紀のフィッシャー」論文(Efron 1998)で、有名なEfronの三角形を提示している。その底辺には、頻度主義とベイジアンが位置し、その両方に対峙する形で、フィッシャー主義者(Fisherian)が位置している。頻度主義は、Neyman=Pearson型の統計分析

パラダイムである。つまり、 p 値問題を頻度主義 vs ベイズアプローチで考えるのではなく、Fisher の伝統も考慮せよ、というのがこの Efron の三角形の意味するところである。

21 世紀の今日、20 世紀初頭に確立された統計手法への見直しがはじまっている、と考えてもよい。対応分析、多重対応分析、そして幾何学的データ解析は、そこで一つの実践例を提示している。

CA および MCA は、確かに、1) 一分析手法ではある。しかし 2) これは、EDA (探索データ解析) をカテゴリカル・データも含めて拡張するツールである。さらに 3) MCA 多重対応分析を基礎に構成される GDA 幾何学的データ解析は、実験可能な環境で構築された分散分析のアイデアを、観察データにまで拡張した構造化データ解析、帰納的データ解析を実現する。

そして、この GDA のアプローチは、他ならぬ、林知己夫たちが開拓した「データの科学」が意味するものと同じ、データ解析手法、手順、データへのアプローチを意味しており、それは、狭い意味での統計処理には収まらないがゆえに、データ「の」科学なのである。

謝辞

本稿は、拙訳本、『対応分析の理論と実践』を検討素材とした「対応分析研究会」(磯直樹先生主催)での発表、報告したことを基礎にまとめたものである。一ヶ月に一回というペースでの研究会に参加していただき、さまざまな意見を頂戴できたこと、参加者の皆様に感謝いたします。

また、本稿は、日本社会学会第 94 回全国大会での「文化・社会意識(5)」部会での筆者による発表「『対応分析』が提示するデータ分析の視点—探索的データ解析(EDA)の重要な構成要素としての CA と「ディスタンクション」パラダイムから GDA へ—」をもとに加筆したものである。大会での発表に対して質問、ご意見をいただいたことに感謝いたします。

なお、本研究は、JSPS 科研費 JP20K02162「データの幾何学的配置に注目したカテゴリカルデータ分析手法の研究」の助成をうけたものです。

注

- 1) 考え方、手法としたのは、この GDA とは回帰分析や因子分析、そして対応分析や多重対応分析のような多変量解析の方法という意味ではなく、データ分析の枠組みについてのフレームワークであるからである。本稿では、個々の解析手法として CA や MCA を扱うときは、狭義の CA、MCA とし、データ解析手法、枠組みを表すときは、広義の CA、MCA と呼んでいる。なお、ブ

ルデューの調査手法については、磯 2020:125 を参照のこと。

- 2) 同様の発言は、ネット上に公開されている「調査票の幾何学的分析 ブルデューの『デイスタンクシオン』から学ぶ」<http://bit.ly/2zWLeCH> でもふれられている。
- 3) 訳註の原文は以下の通り「※図 11・12に見られるように、支配階級の趣味を示すさまざまな指標を座標の中に位置付けるとき、それらの支配階級内集団との対応関係はほぼ一定しており、外力が働かないかぎり変化しないので、ここではこの持続的・慣性的関係を支配する原理を比喩的に「慣性」としてとらえている。以下、()によって示されているパーセンテージは回答率ではなく、それぞれの指標がこの「慣性系」における各集団の位置決定因子にもたらしている寄与率を表す。」

この訳註にあるように比喩的に「慣性」として捉えているところは翻訳者のこじつけでしかないが、おそらく、このあとのパーセンテージについては、ブルデューに質問したのだと思う（「訳者あとがき」からの推察）。そこで言われたことが後半の文として残っているのだと想像している。ここは正しい。つまり、各集団は、多重対応分析における下位集団として定義されて、それが分析対象から生成される（因子と訳されてしまっているが）座標軸（位置を決定する座標軸）への貢献度、つまり寄与率、として計量されていることなのである。それにしても、寄与率のパーセンテージをどうやって比喩的な把握から導くのだろうか。いずれにしても、この訳注で、多重対応分析と界の概念の結び付きは切断されてしまった。

- 4) こうした理解のねじれによって、ブルデューが用いた計量的手法が「因子分析」である、という誤解、混乱が生じた可能性もある。宮島・藤田 1991 の第 8 章は、日本での調査データに対して因子分析をおこなってブルデューの展開として展開している。その分析を結果をうけるかたちで執筆された『文化的再生産の社会学』の初版(1994)はいざしらず、2017に刊行された増補版でも分析方法についてはそのままである(宮島・藤田 1991, 宮島 2017)。
- 5) これと等価な手法は、オランダと日本でも独自に開発、発展してきたのだが、名称を含めその側面はおいでなく。Greenace2016=2020 の「日本語版への序」を参照。）
- 6) グラフは、南アフリカ、ケープタウンで開催された CARME2019 での発表資料に掲載したものである。公開資料の URL は以下の通り。
<https://bit.ly/33CSp5w>
- 7) Michael Friendly は、SAS、R での `vcd: visualizing Categorical Data` の開発者であり、同書には、R の `package, vcd` の成果が凝縮されている。
- 8) Mosaci Plot については、藤本 2018 参照。ただし、日本語表示方法については、`package showtext` を用いて、`showtext_auto(TRUE)` を指定するだけで、可能になることがわかっている。`base` の `graphics` を使う時は、FLASE にする必要があることに注意。
- 9) なお、`correspondence analysis` を「対応分析」と名付けたのは、大隅 1994 である。大隅他 2021:17 第 1 章の訳註 12

- 10) 柳川 2018:41 第 5 章「統計的推論と統計的判定：真の検定をもとめて」の冒頭で「統計的検定には大きくわけて二つの考え方がある。一つは P 値に基礎をおく Fisher 流の考え方、もう一つは、Neyman-Pearson 流の考え方である。前者は統計的推論、後者は統計的判定に焦点をあてている」と要約している。
- 11) この構造化データ解析という名称を用いているかどうかを別にすれば、大隅たちによる翻訳「記述的多変量解析」1994 でも、こうした解析方法は紹介されている。LeRoux & Rouanet 2004,2010=2021 でもふれられているが、幾何学的データ解析とは、ベンズクリが対応分析をデータ解析と同義につかっている（原語で“Analyse des Donnees”）を直訳すると「データ解析」になってしまうので、スタンフォード大学の Patric Suppes によって幾何学的データ解析 (GDA) と命名されたそうした経緯をみると、当初、ベンズクリによって「データ解析」と呼ばれたデータ分析過程は、Rouanet&LeRoux らによって、GDA として整理発展してきた。それは、対応分析と数量化 III 類の数理的な共通性をさらに超えて、林知己夫の「データの科学」とも通じるものである。

参考文献

- Clausen, Sten-Erik, 1987, *Applied Correspondence Analysis An Introduction*, SAGE, (=2015, 藤本一男『対応分析入門』, オーム社)
- Efron · Bradley, 1998, *R. A. Fisher in the 21st Century* [Statistical Science] 13 (2): 95–114.
- Friendly, M, Meyer, D, 2015, *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data*, CRC press
- Greenacre, M.J, 2017, *Correspondence Analysis in Practice Third Edition*, CRC press, (=2020, 藤本一男『対応分析の理論と実践』, オーム社)
- Henry Rouanet, Werner Ackermann & Brigitte Le Roux, 1998, 2004, *THE GEOMETRIC ANALYSIS OF QUESTIONNAIRES: The Lesson of Bourdieu's La Distinction*, <http://www.math-info.univ-paris5.fr/~lerb/publications/LessonDistinction.html> : 「調査票の幾何学的分析：ブルデューの『デイスタンクシオン』から学ぶ」(仮訳) <https://bit.ly/2XbW7j0>
- Hill, M. O. *Correspondence Analysis: A Neglected Multivariate Method. Applied Statistics* 23, no. 3 (1974年): 340. <<https://doi.org/10.2307/2347127>>.
- Lebart Ludovic, Alain Morineau, Kenneth M Warwick, 1984, *Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices*, Wiley series in probability and mathematical statistics. (=1994, 大隅昇・L. ルバール・A. モリノウ・K.M. ワーウィック・馬場康維『記述的多変量解析法』日科技連出版社)
- Lebart, Ludovic, 2008, *Exploratory multivariate data analysis from its origins to 1980: Nine contributions.*, Electronic Journ@l for History of Probability and Statistics Vol 4, n°2; Décembre/December 2008 <http://www.jehps.net/Decembre2008/Lebart_ang.pdf>

- Le Roux, Brigitte, and Henry Rouanet. 2004. *Geometric data analysis: from correspondence analysis to structured data analysis*. Dordrecht: Kluwer Academic Publishers.
- Le Roux Brigitte・Henry Rouanet, 2010, *Multiple correspondence analysis*, Sage Publications, (=2021, 大隅昇, 小野裕亮, 鳩真紀子. 『多重対応分析』. オーム社)
- Tukey, John W. 1962. *The Future of Data Analysis*. The Annals of Mathematical Statistics 33 (1) :1-67. doi: 10.1214/aoms/1177704711.
- Tukey, John Wilder. 1977年. *Exploratory data analysis*. Reading, Mass: Addison-Wesley Pub. Co.
- 磯直樹・竹ノ下弘久, 2018, 「現代日本の文化資本と階級分化 —1995年SSMデータと2015年SSMデータの多重対応分析—」『2015年SSM調査報告書8意識I』.
- 磯直樹, 2020, 『認識と反省性: ビエール・ブルデューの社会学的思考』法政大学出版局.
- 大隅昇, ・L, ルパール・A. モリノウ・K.M. ワーウィック・馬場康維, 1994, 『記述的多変量解析法』日科技連出版社.
- 岡田謙介, 2017, 「ASA 声明とこれからの統計学の使われ方」『社会と調査』19号
- 北田暁大・解体研, 2017, 『社会にとって趣味とは何か: 文化社会学の方法規準』, 河出書房新社.
- 林知己夫, 1986, 「数量化理論のできるまで」『オペレーションズ・リサーチ』vol31, Np12, 日本オペレーションズ・リサーチ学会, 『林知己夫著作集3』所収 p13
- 林知己夫, 1993, 『行動計量学序説』, 行動計量学シリーズ, 朝倉書店.
- 林知己夫, 2001, 『データの科学』朝倉書店.
- 林知己夫, 2011, 『調査の科学』筑摩書房.
- 宮島喬・藤田英典・志水宏吉, 1991, 「現代日本における文化的再生産過程 —一つのアプローチ—」『文化と社会』, 有信堂, 所収(第8章)
- 宮島喬, 2017, 『文化的再生産の社会学 —ブルデュー理論からの展開—』, 藤原書店
- 西里静彦, 2007, 『データ解析への洞察: 数量化の存在理由』関西学院大学出版会
- 西里静彦, 2010, 『行動科学のためのデータ解析: 情報把握に適した方法の利用』, 培風館
- 西里静彦, 2014, 「行動科学への数理の応用: 探索的データ解析と測度の関係の理解. *行動計量学* 41, no.2 (2014年): 89-102. <https://doi.org/10.2333/jbhmk.41.89>.
- 柳川堯, 2018, 『P値: その正しい理解と適用』近代科学社
- 藤本一男, 2019, 「Supplementary 変数 からMCA(多重対応分析)を考える —幾何学的データ解析(GDA)と多重対応分析(MCA)—」『津田塾大学紀要』, 51 p155-67. <http://id.nii.ac.jp/1234/00000205/>
- 藤本一男, 2020, 「対応分析は〈関係〉をどのように表現するのか —CA/MCAの基本特性と分析フレームワークとしてのGDA—」『津田塾大学紀要』52 p169-84. <http://id.nii.ac.jp/1234/00000247/>