

「Supplementary」変数から 多重対応分析（MCA）を考える

—— 幾何学的データ解析（GDA）と多重対応分析（MCA） ——

藤本 一男

概要

本稿では、多重対応分析（MCA）の応用手法である幾何学的データ解析（GDA）という方法を「計画的要素」に着目して説明する。そこでは、分析対象が active 変数と supplementary 変数に区分され、前者によって分析対象が「場」として構造化され、後者は分析関心としてそこに射影される。これは、分散分析が実験計画とセットになってこそ要因の効果を環境要因から分離できるということを、実験（計画）が不可能な観察データにおいて実現する手法である。

多重対応分析の数理的な核は林知己夫の数量化 III 類と同じである、と言われることがあるが、それは土台の共通性にすぎない。問題はその先にある。この手法（GDA）においては、なにを supplementary 変数として選択するか（同時に分析対象の構造をどの変数をもって生成するか）という点に、重要な「計画的要素」が存在するのである。

1.0 はじめに

1.1 多重対応分析における supplementary 変数の重要な機能

社会学における対応分析の使用ということではブルデューによる『デイスタンクシオン』があげられる。しかし、翻訳者の多重対応分析に対する理解

の問題もあり、ごく一部の研究者の営みを除いて、ブルデューの方法の全体像が把握されてきたわけではなかった。

それが、ここ数年、新しい展開を見せている。2016年に北田たちによってブルデューの多重対応分析使用への疑問が提示され、2017年には、ブルデューの手法を批判的に継承しイギリス社会に適用した『文化・階級・卓越化』の翻訳が刊行された。

他方、分析環境としてパーソナルコンピュータの高性能化と統計処理言語Rの普及にともなう対応分析、多重対応分析の利用の容易化が進行し、データ分析の実行が用意になってきているという側面もある。

本稿では、こうした変化を念頭におき、対応分析（CA）／多重対応分析（MCA）を、分析手法として用いていく際のポイントになるのが「supplementary」変数の扱いであるという視点から、GDA（幾何学的データ解析）という手法を説明する。

1.2 北田によるブルデューの多重対応分析への疑問

北田は、北田（2017）の第二章で、ブルデューによる多重対応分析の使い方疑問をなげかけている。（*1～*3..は引用者によるマーク。）

「生活様式空間と相同的な関係を持つとされる社会空間（生活様式空間と社会空間の相同性）は、これまでの「関係論的」視座からすれば、それ自体、（多重）対応分析において考慮されるべき事項であるように思えるが、ブルデューは、「性」「学歴」「所得」といった基本的なフェイスシート項目を、社会空間の構築にさいしては変数として考慮していない*1。多重対応分析であれば、こうしたフェイスシート項目は、当然分析の中に組み入れることは可能だし、また一般的にもそれが通常の使用方法である*2。しかし、ブルデューは、「趣味どうしの関係性」を純粋に数学的に取り出すという目的を設定するためか、対応分析の時点では、こうしたフェイスシート項目は加味していない（が、実際は社会階級×特性の（多重）対応分析がなされている部分もある）。つまり、人口学的項目は「追加変数」として「生活様式空間」の図に射影されている*3（Broudeiu 1979=1990:12）。」

この北田による多重対応分析での人口学的変数の扱いへの疑問は（その意図に反して）、そこでの多重対応分析の使われ方を正確に読みとっている。

*2で、「フェイスシート項目を分析の中に組み入れることは可能」という。それはその通りである。しかし「一般的にもそれが通常の使用方法である」というが、それは違う。それは、多重対応分析をどのように用いるかという分析の手法にかかわることで、フェイスシート項目、つまり人口学的変数（デモグラフィック変数）を主軸の生成に寄与させる使い方もできるし、*3として読みとっているような使い方、つまり、生活様式空間を（active変数である）趣味変数によって構築し、そこに質量を持たない supplementary 変数として人口学的変数を射影する、という使い方もあるのである。ブルデューは、後者を意識的に用いている。この手法は、GDA（幾何学的データ解析）と呼ばれている。

本稿では、ブルデューの側からではなく、多重対応分析の側からこの問題にアプローチしてみたい。

2.0 幾何学的データ解析 (Geometric Data Analysis) という方法

さて、社会学において対応分析の応用例といえば、まずブルデューによる使用があげられる（Clansen1998:37=2015:38）。その意味ではブルデュー研究において、彼の基本概念である「界」「文化資本」「社会空間」を理解するためには、用いられた数理的な手法である対応分析の理解が不可欠であるはずであるが、残念ながら手法としての対応分析への認知が低いため、社会学の主要な学術誌に掲載された論文に限定してではあるが、対応分析と結びつけてブルデューの立論を検討する論考は（近藤のものを除いて）現れていない。

近藤（2010）が述べているように、ブルデューの方法を日本に適用した事例もすくなく、また、そもそも、多重対応分析をブルデュー的に適用した例もほとんどないのが現状である。

それでは、対応分析は用いられていないのかということ、そうでもない。図1にみるように、一定数の論文が生産されている。多くはマーケティング分野であり、この間急激に増加しているのは、テキスト・マイニングに関連するものである。他方、ブルデューや近藤の対応分析の適用は、大規模な社会調査データである。

そのため、多重対応分析を用いるのは、大規模（社会調査）データ、という理解があり、他方、ソフトの使い方の解説しかないために、この手法、対応分析の機能が理解されている小規模なサンプルデータを用いたものとはいえない状況が続いている。

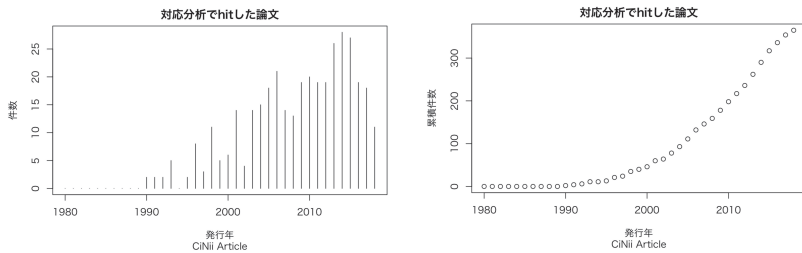


図1 CiNii でヒットする「対応分析」の論文

2.1 (多重) 対応分析はデータ分析の土台である

Brigitte Le Roux は、「多重対応分析を用いればブルデューのような分析ができるわけではない」と述べている (Rouanet, Le Roux 2004)

対応分析 (CA) や多重対応分析 (MCA) の理解が普及していない现阶段では、はじめに引いた例のように、この手法に対しては、誤解が生じている。そのため、この章では、幾何学的データ解析 (GDA) という視点を踏まえて CA、MCA の特徴について整理しておきたい。

詳細は、他書にゆずるが、対応分析の基本的な特性を整理しておこう。

2.2 対応分析、多重対応分析、主軸の生成、supplementary 変数

まず、対応分析 (Correspondence Analysis : CA) という手法について概説する。この手法は、クロス表に表現されたカテゴリカル変数を、表頭の変数と表側の変数という 2 つの変数の変数カテゴリーの関係をマップと呼ばれるものに図示する。この 2 変数を 3 変数以上に拡張したものを、多重対応分析 (Multiple Correspondence Analysis : MCA) と呼ぶ。

多くの解説書が、(多重) 対応分析は、林知己夫の数量化 III 類と、また、西里静彦の双対尺度法などと数理的に等価である、という言い方がされる。～と等価である、と言うとそこでわかった気になるのかもしれないが、問題はその先である。なお、手法のコアの部分はそうであっても、相違点がある、という展開を村上が行なっているりが、本稿で扱うのは、この側面ではない。

2.3 「supplementary 変数」とはなにか

「追加変数」「補助的変数」という訳語があてられる supplementary Variable であるが、対応分析の応用を考える上で重要な位置をもった用語である。

本稿のテーマである幾何学的データ解析 (GDA) は、まさにこの機能を用いて展開される。

CA (そして MCA) においては、データはプロファイル・ポイントの集合として分析の対象となる。そこから次元縮減が行なわれ、データの座標軸 (主軸) の生成、各ポイントの座標の決定が行われる。主成分分析で行われるように、分散が最大になる方向に座標軸が設定される (数理的には、残差行列に対する特異値分解による)。データの各ポイントはこの座標軸の生成に寄与している。その寄与は、クロス表に表現されたデータの周辺度数に対応する「質量」(mass) の大きさに関係する。質量をもったポイントが、座標軸を生成するのだが、supplementary 変数とはこの「質量」をもたない変数のことである。質量をもたない変数であっても、対応分析の (行変数と列変数の間の) 遷移公式 (Transition formula) によってマップ上に位置を占めることができる。

この特徴を生かして、分析手法に整理したものが GDA である。

2.4 サプリメンタリ変数の使い方

GDA の説明に入る前に、もうすこし、この supplementary 変数についてみておきたい。というのも、GDA 的に用いるのが supplementary 変数の唯一の使い方ではないからである。そのことを、Greenacre 2017 の第 12 章は、この変数の使い方を以下のように整理している。

第 1、分類上別々に扱ったほうがいいものを分離する、第 2、質量が少ない外れ値をあてはめる、第 3、ポイントのグループや区分を表示する。(Greenacre2017:90-93)。

第 1 の用い方は、座標軸の生成 = 構造構築に用いる変数の属性をあわせておかないと、生成される軸の意味を解釈しづらくなるからである。第 2 は、質量が少ないポイントは、データの重心 (平均を意味する) から遠く離れたところに位置することになるのだが、それが生成される軸に大きな影響をあたえてしまうことになる。それを防ぐための処置である。軸の生成には寄与させない。しかしマップには位置をもつ。第 3 は、変数をグルーピングしてその関係を見る場合の処置である。これの詳細は次の GDA の説明の中で触れていく。

以上に共通しているのは、座標軸を生成するポイント (active ポイント) と、座標軸生成には寄与しないポイントにわけてデータを分析するというアプローチである。第 1、第 2 の用い方では、supplementary variable,

supplementary point を「追加変数」「追加ポイント」と呼ぶのも違和感はないが、第3の方法では、active point と同等の位置を付与されている比較対象としてのポイント群を指している。そのため、Illustrative point (説明ポイント)とも呼ばれている (Le Roux, Rouanet 2004)。

3.0 GDA というアプローチ

やっと GDA の説明に入るが、ここで更に前置きを加える。

CA や MCA を用いると必ず GDA になるというわけではない、ということである。それは、Le Roux たちが強調するように、「MCA を用いればブルデューのような分析ができるわけではない」ということでもある。ブルデューは意識的に MCA を GDA として用いている²⁾。また、GDA としてつかわなければ、CA や MCA で分析する意味がない、ということでもないのである。

3.1 GDA を考える前提の整理

データと分析ツールの関係を考えるにあたって、幾層かのレイヤーを考えてみたい。一番下には、取得したデータが位置する。統計的手法がその上にのり、さらに、その上に解釈が乗るようなイメージが一般的である (図2)。

しかし、データと統計手法の関係は、(統計手法によって)自動的に形成されるわけではなく、「データ取得前提」とでもいうようなレイヤーを考える必要がある。

GDA を構成するレイヤーには次のものがある。データは、CA や MCA、そして PCA によって分析される。その際に、データは、構造化されている。その構造化は、先に触れた、active variable という機能を使うことによって実現する。

「active」「supplementary」になにを割り当てるかという「デザイン」によってデータは構造化される。そして、active 変数によって構築された主軸が形成する空間に、このサプリメンタリー変数による空間が射影され、分析されることになる。

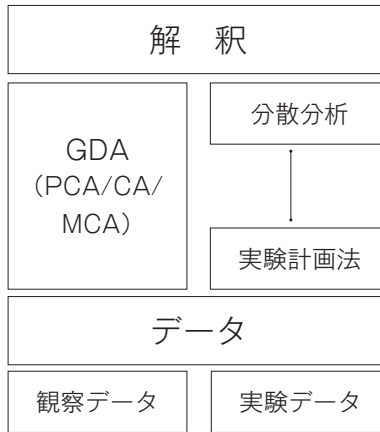


図2 データ分析のレイヤー

3.2 実験計画法 / 分散分析と GDA（幾何学的データ解析）

こうした構造化された分析を、Le Roux たちは、実験計画法と分散分析を例に説明している（Le Roux, Rouanet 2004, 2010）。

分散分析は、得られたデータから、誤差成分を分離して要因の効果を抽出する手法である（この手法についての解説は多数あるが、例えば三中 2018 参照）。しかし、それには前提があり、測定しようとしている要因の効果を、その効果に影響しそうな誤差要素を統計的に除去できるように実験が「計画」されなくてはならないということである。これが適切に（つまり統計処理によって誤差を分離できるように）デザイン＝計画されていない場合は、取得したデータをどのように統計処理しようと注目している要因以外の要因の影響を除去できず、分析は失敗する。

この説明は、次のようなデータ区分が前提になっている。つまり、実験計画を実現できる、つまり環境をコントロールできる条件がある実験データ（experimental data）と他方、そうしたコントロールが不可能な、社会調査のような観察データ（observational data）という区分である（Hand, 2008=2014）。

こうした社会調査データのような観察データに対しては、日本での計量社会学は、主として重回帰分析を用いてきた。こうした傾向は、世界的にあり（ベネット 2009=2017 の解説論文で相澤が触れている）、そのことをブルデュエーは前提にして、多重対応分析を選択して理論形成をおこなってきた経緯があ

る。そこにあるのは（重回帰分析にみられる）「変数の社会学」への批判と、それに対して（多重対応分析による）関係性アプローチである。この部分については、近藤（2007）がわかりやすくまとめてくれている。

「そのアプローチ（引用者注：「変数の社会学」）の最大の特徴は、職業の階層的な分類や序列がデータ分析に先立って与えられるという点にある。そして既存の分離や尺度が捉えた現実のみが変数間の関係を通して解釈に委ねられるという点も、そのアプローチの重要な特徴として指摘することができる。これに対して、社会空間アプローチ（引用者注：ブルデューによるGDA）では、職業の階層的な分類や序列がデータ分析に先立って与えられることはなく、それらがデータ分析の中からデータの構造にもっとも適合的な関係として引き出されるという特徴をもっている。」（近藤2007:313）。

4.0 トニーベネット『文化・階級・卓越化』にみるGDAの実際

本書の第3章に用いられた統計的手法、多重対応分析についての説明がある。まず、強調しておきたいのは、「謝辞」「序論」でも触れられているが、トニーベネットら執筆者を統計処理の側面から支えたのは、本稿でもなんども触れているBrigitte Le RouxとHenry Rouanetだということである。このふたりは、ブルデューとも一緒に仕事をしており、対応分析の始祖であるベンズクリの直系の弟子である。

第3章に、多重対応分析についての要約がある。（Bennett.T, et al. 2009:44=2017:91-92）

第1に、文化マップは、嗜好の社会的規定要因に関する想定を暗に持ち込むことなく、文化的生活自体の様々な側面の組織化や違いの関係性だけにもとずいて構築される。

このあと、伝統的な多変量解析（重回帰分析）との関係にふれて、

我々は多重対応分析を伝統的な「変数中心」の分析手法とともに用いることができるという見方をもっている。

と続ける。この点は、Le Roux, Rouanet 2004, 2010でも繰り返し主張されている。

二つ目の点は、いったん文化的マップを構成すれば、その上に社会的なカテゴリを重ね合わせ、これらの「補足の変数」が文化的景観に関連しているかどうかを判断できることにある。このアプローチには、文化マップの座標はそのままに、ライフスタイル空間の様々な次元に社会的特性を「重ね合わせる」ことが含まれる。

この「補足の変数」とは本稿で触れてきた supplementary Variable である。

三つ目、それぞれの文化的変数の平均を示す点を配置することができるだけでなく、同じ座標軸上で調査対象となった個々人の点がどこにあるのかも特定できることにある。「個々人のクラウド」を精査することを通して、どのような個人と個人が近接しているのかを検討することで、文化実践の組織化について、さらに洞察を得ることができる。この「個人クラウド」の分析についての特筆すべき点は、質的インタビューと文化マップの位置を関連つけられることにある。それぞれのインタビュー対象者がどこに位置しているのかを突き止め、量的調査の回答と口頭でしゃべったこととを結びつけることができる。第4章で示すようにこの方法をとることで量的データと質的データの間に豊かな対話が生まれることになる。

少し長い引用となったが、対応分析が個人×変数の表を分析するにあたり、分析の過程で個人は、平均や誤差に解消されることなく、個人ポイントとして保持されるということであり、それゆえ、インタビュー調査とマップに表現された（必要に応じて伝統的多変量解析を用いた解釈をともなった）量的な関係が結びつけられる。

Le Roux, Rouanet (2010) の第1章のタイトルに下には、「量と質の間には幾何学がある」というフレーズが掲げられているが、その意味するところは、実践的展開を本書によって検討することが可能になる³⁾。

5.0 日本における GDA の紹介

MCA についての解説書もさることながら、GDA についての紹介を日本語で目にするのではないように思えるかもしれない。しかし、GDA のアイデアは、すでに紹介されている。それは、英語圏に対しても GDA のアイデアを紹介した先駆的な文献である、Lebert et al. 1984 の翻訳書である、大隅、他 1994 である。この翻訳書の IV.6 (MCA) 調査分析では、この GDA のアイデアが展開されており、「ある社会調査における統計的な要素を記述する変数は次の 2 つのグループに大別されている」として (a) デモグラフィック変数と (b) 調査内容に関連した変数にわけている。この (b) は更に 3 つの分類がされている (大隅 1994:163)。詳細は、本書にあたっていただきたいが、ここで指摘しておきたいのは、(a) のデモグラフィック (つまり人口統計学的変数) に対して MCA を行い、それを予測マップとし、そこに (b) の変数を射影する、という解説がされていることである。

ここでの例では、デモグラフィック変数によって構造を構築し、そこに調査内容を「追加変数」として射影する、という構成になっているが、どちらで構造を構築するのか、は調査のデザインの問題であり、デモグラフィック変数は必ず「追加変数」として処理するというのではない。

なお、p164 には、「追加変数の選択として最も重要な処理」という記述があるが、これが分析の枠組みの、データの構造把握の出発的になるからに他ならない。

6.0 結語

ブルデュー研究においては、彼が理論構築の核にすえた MCA、そして GDA に対する理解が普及していない中、分析手法に対する評価がとりのこされたまま様々な議論がおこなわれてきたように思う。近藤 2011 のように、MCA の手法的な特徴を押えて、ブルデューの手法を日本に適用した事例は稀である。

ブルデューの代表作であり、対応分析を用いた分析の最大の事例である『ディスタンクシオン』はフランス文学者である石井洋次郎による美しい日本語に訳されているとはいえ、統計的分析の核である対応分析については、まったく理解されていない (原著にも英語訳にもない奇妙な訳注が「慣性」につけられている。対応分析において「慣性」とは、分散のことである。こ

れを構造主義的な読み込みからなのか、構造の特性を表現するものとして「剛体の慣性」のような注になっている⁴⁾。

また、原文がフランス語であるということも理論的な検証を行う際のハードルを高いものにしていただと思われる。

しかし、2016年には、北田によるブルデューの多重対応分析の使用に対する率直な疑問が提示されることによって、従来正面から触れられてこられなかったブルデュー評価の新しい局面が開かれようとしているといえる。

加えて、2017年には、ブルデューの伝統を批判的にイギリス社会に展開したトニー・ベネットの『文化・階級・卓越化』の翻訳が、磯直樹たちによって翻訳刊行された。翻訳者は、対応分析の基本的なところを押さえているようである。さらに、原著は英語でありフランス語に比べて原著をレビューするハードルは低い。

GDA に関して体系的に書かれたものとしては、Le Roux, Rouane 2004 があるが、なにぶん、大著であるのと英語である。しかし、この同じ著者による SAGE 出版による「Multiple Correspondence Analysis」の翻訳刊行が近々行われるという動きも聞いている。(なお、Le Roux たちが本書で触れている Specific version の MCA がある。これはトニーベネットの著書のなかでも参照されている、Specific MCA⁵⁾ と Class Specific MCA、であるが、すでに R のパッケージとして公開されている。GDATools という名称である。<https://cran.r-project.org/web/packages/GDATools/index.html>)

このように、ブルデュー理論の批判的検討条件は明らかに拡大してきているとあってよいだろうし、多重対応分析、幾何学的データ解析の可能性を理論面、実践面から検討する条件も拡大している。

注

- 1) 村上, 2012: 48 「数量化Ⅲ類は多重対応分析と等価である」というしばしば唱えられる説が、少なくとも部分的には正しくないことを示す。数量化Ⅲ類には、多重対応分析とは異なった扱いを許す部分がある。あるいは、データのコーディングの方法によって、多重対応分析とはまったく異なった解を生み出すことがある。すなわち、多重対応分析は数量化Ⅲ類に完全にはとってかわることができない。」
- 2) Le Roux, Rouanet 2010 の第 1 章に、ブルデューの統計手法が年代区分とともにまとめられている。
- 3) なお、このような量的調査と質的調査の結合を「混合調査法」(mixed methods research) という名

称で参照することも行われている（『社会と調査』第11号特集）。ただし、このGDAが実現する結合のしかたを広義の「混合」調査法と呼べないかもしれないが、インタビュー調査の結果にテキスト分析を行い、それを計量分析の対象として調査票調査の分析に統合する、というアプローチではないことを記しておく。

- 4) ブルデュー（1990:10）日本語版のみの訳注：「※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づける時、それらと各階級集団との対応関係はほぼ一致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・惰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。」
- 5) Le Roux 達が、Specific MCA と呼ぶ MCA の version は、Greenacre が Subset MCA（Greenacre 2017）と呼ぶものである。

参考文献

- Clausen, Sten-Erik（1998）Applied Correspondence Analysis, Sage publication（訳：藤本一男, 2015, 『対応分析入門』オーム社）
- Greenacre M, 2017, Correspondence Analysis in Practice Third Edition, CRC Press.（訳：藤本一男, 2019 『対応分析の理論と実践——基礎・応用・発展——』（仮題）オーム社, 2019 年刊行予定）
- Hand David J, 2008, Statistics: A Very Short Introduction, Oxford University Press,（訳：上田 修功, 2014, 『統計学』丸善, サイエンス・パレット）
- Lebaron, Frederic, 2009, How Bourdieu “Quantified” Bourdieu: The Geometric Modeling of Data, ed: Karen Robson, Chris Sanders “Quantifying Theory: Pierre Bourdieu”, Springer
- Lebart, L., Morineau, A. & Warwick, K. (1984) Multivariate Descriptive Statistical Analysis. Chichester, UK: Wiley.（大隅昇・L. ルバル他, 1994, 『記述的多変量解析法』日科技連）
- Le Roux, B. & Rouanet, H. (2004), “Geometric Data Analysis: From Correspondence Analysis to Structured Data.”, Dordrecht: Kluwer.
- Le Roux, B. & Rouanet, H. (2010), Multiple Correspondence Analysis, SAGE publish,
- Rouanet, H. Ackermann W. & Le Roux, B. (1998, 2000, 2004) “THE GEOMETRIC ANALYSIS OF QUESTIONNAIRES: The Lesson of Bourdieu’s La Distinction” <http://www.math-info.univ-paris5.fr/~lerb/publications/LessonDistinction.html>
- Bennet. T et al, 2009, “Culture, Class, Distinction”, Rontledge（訳：磯直樹他）, 2017 『文化・階級・卓越化』青弓社
- 北田暁大、解体研, 2017, 『社会にとって趣味とはなにか——文化社会学的方法的基準——』河出書房
- ブルデュー（訳：石井洋次郎 1990）『ディスタンクシオン』藤原書店

三中信宏, 2018, 「統計思考の世界 曼荼羅で読み解くデータ解析の基礎」技術評論社

近藤博之, 2007, 「階層研究における社会空間アプローチの実践」日本教育社会学会大会発表要旨,
(<https://ci.nii.ac.jp/naid/110006479761>)

近藤博之, 2009, 「階層研究における社会空間アプローチの実践 (2)」日本教育社会学会大会発表要旨,
(<https://ci.nii.ac.jp/naid/110007635000>)

近藤博之, 2011, 「社会空間の構造と相同性仮説——日本のデータによるブルデュ——理論の検証——」
『理論と方法』26 (1), 161-177, 数理社会学会

村上隆, 2012, 「数量化 III 類と多重対応分析」『社会と調査』(9) 48-62, 社会調査協会 (http://jasr.or.jp/asr/asrpdf/asr09_020.pdf)

特集「量と質を架橋する：混合研究法 (mixed methods research) の可能性」『社会と調査』第 11 号
2013 (<http://jasr.or.jp/asr/asr11.html>)