『津田塾大学紀要』第 52 号（2020 年）169-184 頁

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

藤　本　一　男

概要

本稿では、対応分析は〈関係〉をどのように表現するのかという問いに対して、以下の構成で説明を行う。最初に、この手法に言及されるときに必ずといっていいほど参照される P. ブルデユーの、また最近でいえば、T. ベネットの論考での（多重）対応分析の扱いを整理する。そこには、対応分析へのアプローチの二つの領域が含まれていること、そのために本稿前半で、対応分析の原理、特徴を解説する。そこでは 2 元クロス表に対する（シンプル）対応分析（CA）と個体×変数の表を対象とし、3 変数以上を扱える多重対応分析（MCA）の関係も整理する。そして後半で、この対応分析を中心要素して構成される幾何学的データ解析（GDA）という分析フレームワークについて説明する。このフレームワークは、実験パラダイム（実験計画法が前提とするコントロールが可能な条件）がなりたたない社会調査データで、どのようにして、分散分析（ANOVA）と同じことを実現することを目指している。

なお、本稿は、日本社会学会第 92 回大会（於：東京女子大学）で開催されたテーマセッション「〈関係〉と〈プロセス〉の社会学の可能性」での発表草稿を元に加筆したものである。当日使用したスライド、および配布レジュメは、Web において公開している [1)]。

はじめに ― 対応分析と二つのアプローチ CA/MCA と GDA

社会学において対応分析を語るとき、P ブルデューの『デイスタンクシオン』が、最近では T ベネット他の『文化・階級・卓越性』が参照される。

そこで、この対応分析という手法を理解する際に、二つのアプローチがあることを最初に述べておく。第一は、この手法自体に関するもの、第二が、この手法を重要な要素として遂行されるデータ分析のフレームワーク、というアプローチである。もちろん、この両者は密接に関係しているが、別のものである。後者は、幾何学的データ解析（GDA）と呼ばれるもので、P ブルデュー、T ベネットの社会学的分析の計量的基礎になっているものである。

そこで用いられている計量的手法が他ならぬ対応分析であるが、それは、伝統的な計量社会学の手法である（重）回帰分析を「変数の社会学」として批判し、それに関係性の社会学を対置する理論的前提として位置付けられている（ブルデュー［1990］、ベネット［2009=2017］）。

しかし、この変数の社会学への批判は、対応分析という手法から自動的にでてくるものではない。問題になっているのは分析のフレームワークだからである。また、回帰分析に代表される「変数の社会学」を批判するからといって、回帰分析という手法の適用を否定し、すべてを対応分析でおこなう、という主張でもないのである。

多重対応分析（MCA）を基本ツールとする GDA は、まずデータをみて、そこに内在する構造に注目し、その上で適切な分析手法を採用せよ、と主張する。

以上のことを踏まえて、本稿では、前半に対応分析自体の特性について整理し、後半で、GDA という手法が、実験パラダイム（実験計画法＋分散分析）がなりたたない観察データにおいて、どのように変数 / カテゴリー、そして個体の関係を分析するのかを述べる。

1　対応分析（CA）と多重対応分析（MCA）

ブルデュー、ベネットともに用いている技法は、多重対応分析（MCA）である。対応分析（CA）が、2 変数の分割表を分析対象とするのに対して、多重対応分析（MCA）は、個体×多変数の表を分析処理の対象とする。数理的な基礎およびこの手法の特徴的な部分は CA の機能を理解することで得られるので、まず、CA を説明し、それを MCA に拡張していく。

（注：Le Roux 2010 のタイトルは “Multiple Correspondence Analysis”（多重対応分析）であるが、そこではCA の知識を前提にはしていない。そういう説明のアプローチも存在する。）

1.1 CA は二つの変数の関係を要約し可視化する

分析対象が量的変数である場合は、その 2 変数の関係を可視化するには、散布図が用いられる。同じことを質的変数で行う場合に、この散布図に相当するのがクロス表である。クロス表は、表頭、表側に変数と応答カテゴリを配置し、その交差したセルに対応する度数が記入される。

ここでその変数の関係の独立性を確認する場合には、カイ二乗検定を行い、得られ p 値をもとに、帰無仮説（関係は独立）の扱いを決定する。

しかし、この場合、検討している変数のカテゴリが 2 × 2 程度のものであれば、関係性をセルの数値から読み取れないわけではないが、カテゴリが 3 以上になったときに、クロス表から連関を読み取るのは困難である。

そこで mosaic plot（注：全体を 100 %とする棒グラフ表現に、帯の高さを周辺度数に対応させたタイル表示のグラフ。R では、mosaiplot ()、カナダの M.Frinedly らによる mosaic などがある。藤本 2018）などを用いて、全体の連関を可視化することができるが、そこでも、行変数に注目した行分析と列変数に注目した列分析は、異なった様相となる（図 1-1 が行分析、図 1-2 が列分析）。対応分析は、この行分析と列分析を一枚の平面に可視化する（図 2）。

表 1　余暇と職業のクロス集計 ノルウェイ統計「生活レベル調査 1995」から。Clausen= 藤本 p4,113

余暇	現業	Low 非現業	High 非現業	農民	学生	退職者
スポーツイベント	301	497	208	50	254	187
映画	261	550	250	27	339	157
ダンス / ディスコ	361	534	204	59	324	216
カフェ / レストラン	463	766	334	72	350	601
劇場	89	350	195	12	143	167
クラッシック・コンサート	23	182	124	10	60	110
ポップ・コンサート	117	298	145	11	184	56
絵画展	104	379	219	21	152	213
図書館	130	352	153	17	272	264
教会サービス	168	370	187	51	162	424

図 1-1　表 1 の行分析

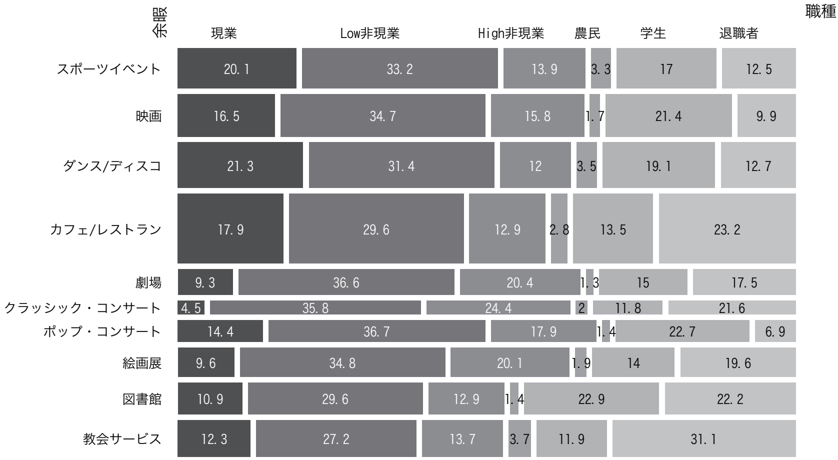


図 1-2　表 1 の列分析

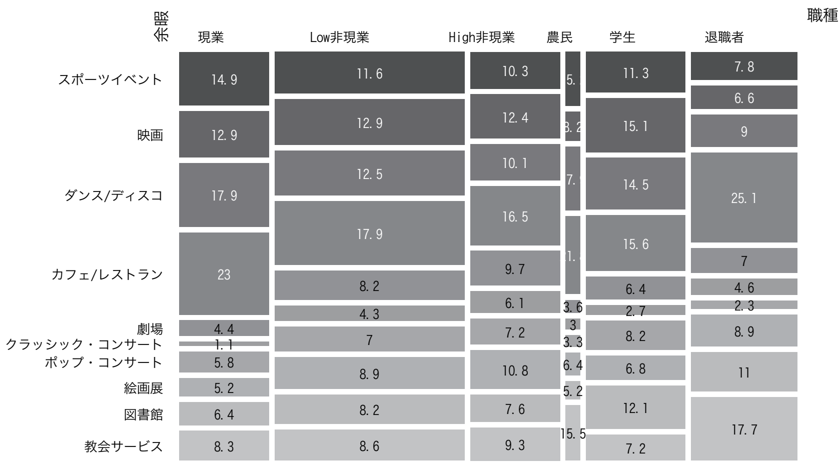
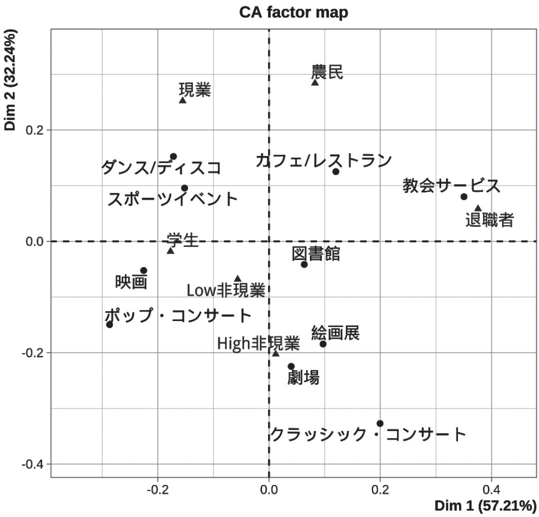


図２　表１の対応分析による 1-2 次元対象マップ



1.2　数理的原理

クロス表の i 行を、$A_i = a_{i1}, a_{i2}, ...a_{ij}, ...a_{im}$ というベクトルと考える。a_{ij} は、セル度数を総数で除した同時確率である。

表２　データ行列

		変数２						
		回答１	回答２	…	回答 j	…	回答 n	合計
変数１	回答１	a_{11}	a_{12}	…	a_{1j}	…	a_{1n}	$a_{1.}$
	回答２	a_{21}	a_{22}	…	a_{2j}	…	a_{2n}	$a_{2.}$
	回答３	a_{31}	a_{32}	…	a_{3j}	…	a_{3n}	$a_{3.}$
	回答	:	:	…		…	:	:
	回答	:	:	…		…	:	:
	回答	:	:	…		…	:	:
	回答 m	a_{m1}	a_{m2}	…	a_{mj}	…	a_{mn}	$a_{m.}$
	合計	$a_{.1}$	$a_{,2}$	…	$a_{.j}$	…	$a_{.n}$	N

行（同じように列）ポイントをこのようにベクトルとして考え、その距離（i 行、i' 行）を考える。おなじみの距離は、対応する次元の値を除したものを

2乗した総和の平方根というユークリッド距離であるが、この手法では、ここで、加重距離を考える。ウェイトは、その行（または列）の周辺度数（mass質量、と呼ばれる）である。こうして、得られる距離はカイ2乗距離となる。対応分析が扱う〈関係〉の基本は、このカイ2乗距離である。

実際の計算は、文献（Clausen：1998=藤本：2015など）を参照していただきたいが、特異値分解（SVD）を用いることで非常にシンプルに次元縮減の対象となるあらたな空間が生成される。概要はAppendix Aに掲載。

1.3 CA処理によって得られる統計量

分析処理の結果、各ポイントは、座標とあわせていくつかの統計量をもつ。一つは、軸の生成にどれだけ寄与したかという絶対的寄与（contribution）、また、そのポイントが各軸によってどれだけ表現されているのかの指標となる相対的寄与（COS2）、である。

前者は、生成された軸を解釈する際に必要となり、後者は、注目している平面上にどれだけの「質」でポイントが表現されているのかを判断する基準となる。あるポイントが注目している平面ではあまり表現されていなくて、別の軸上で表現されていることもある。

1.4 行ポイントと列ポイントの同時布置の問題

対応分析の利点としてあげれる行ポイントと列ポイントを同一平面で表示することであるが、これが可能になるのは、行分析によって得られる座標軸（およびそれが体現する分散）と列分析によって得られる座標軸（およびそれが体現する分散）が等しいからである。しかし、注意しなくてはならないのは、ともに主座標として同時布置することは可能であるが、ともに標準座標として図示することはできないということである。さらに、この前者、主座標として重ね合わせている場合でも、それぞれの空間は別のものであるために、重ね合わせは、あくまでも「概略」であるということを心しておかなくてはならない（これについては、藤本2017参照）。

1.5 サプリメンタリ・ポイント

対応分析/多重対応分析を要素手法とする分析手法が、幾何学的データ解析（GDA）と呼ばれることがある（藤本2019）。以下に概要を説明しておく。

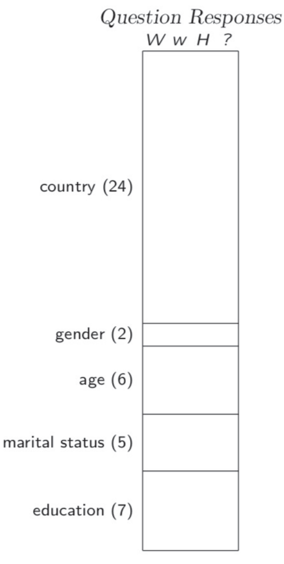
先に、行（もしくは列）ポイントは、同時確率のベクトル＋質量（mass）と書いた。このmassがあることによって、新たな座標軸が形成される。サ

プリメンタリ・ポイントは、この質量、mass をもたない行（そして列）ポイントのことであって、これらは、座標形成つまり構造構築には関与しないポイントなのである。しかし、これらのポイントは、転位公式（transition formula）によって、構造の中に投影される、つまり座標はもつことができるのである。

この関係を用いて、構造を構築し、そこに注目している変数を投影する、という使い方が可能になる。（例：例えば、ある年次のデータで基本的構造を生成して、その翌年、それ以降のポイントをサプリメンタリ・ポイントとして投影することで相対的な変化トレースすることが可能になる。）これが、幾何学的データ解析（GDA）の第一歩でもある。この点に入る前に、CA を多重対応分析（MCA）に拡張しておこう。

1.6 CA が対象とする「表」の拡張と基本的 MCA

CA が処理の対象とするのは、二元クロス表であるが、これ以外にも、％表記の成分表、積み重ね表、などさまざまな拡張が検討さてきた（Greenacre 2017、他）。MCA との関係で重要なのが、この積み重ね表である。



表頭の W w H ? は、それぞれ、フルタイム、パートタイム、在宅、わからない、という回答カテゴリに対応している。
それにたいして、5 つのクロス表が積み重ねられている。国、性別、年齢、婚姻状態、教育歴、である。() 内の数字は、行数（各変数のカテゴリ数）を示している。

図 3　積み重ね表の例 Greenacre2017, p130 Exhibit17.1 より引用

MCAでは、分析対象とする元表は、行に個体、列に変数という構成であるが、これを、二つのデータ表現方法で処理する。一つは、変数を回答カテゴリに分解し、応答を1/0とするindicator行列に変換し（MCAのソフトウェアが自動的に行う）、それにCAをおこなう、というアプローチである。

Questions				*Qu. 1*				*Qu. 2*				*Qu. 3*				*Qu. 4*			
1	*2*	*3*	*4*	*W*	*w*	*H*	*?*	*W*	*w*	*H*	*?*	*W*	*w*	*H*	*?*	*W*	*w*	*H*	*?*
1	3	2	2	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
2	3	3	2	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0
4	3	3	2	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
4	4	4	4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
1	3	2	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0
⋮	⋮	⋮	⋮	⋮				⋮				⋮				⋮			
. . . and so on for 3418 rows																			

表3　indicator行列：左の設問（Questions）に各問い（Qu1～Qu4）への回答番号が記述されている。右側に、その回答に対応した列が1になるように展開されている。(Greenacre 2017, p138 Exhibit18.1)

もう一つの対象データの変形は、それぞれの変数の総当たりクロス表で構成されるBurt表で、それに対してCA処理を行うのが、MCAのもう一つの基本形である。

	1W	*1w*	*1H*	*1?*	*2W*	*2w*	*2H*	*2?*	*3W*	*3w*	*3H*	*3?*	*4W*	*4w*	*4H*	*4?*
1W	2501	0	0	0	172	1107	1131	91	355	1710	345	91	1766	538	40	157
1w	0	476	0	0	7	129	335	5	16	261	181	18	128	293	17	38
1H	0	0	79	0	1	6	72	0	1	17	61	0	14	21	38	6
1?	0	0	0	362	1	57	108	196	7	96	55	204	51	45	2	264
2W	172	7	1	1	181	0	0	0	127	48	4	2	165	15	0	1
2w	1107	129	6	57	0	1299	0	0	219	997	61	22	972	239	13	75
2H	1131	335	72	108	0	0	1646	0	24	989	573	60	760	616	84	186
2?	91	5	0	196	0	0	0	292	9	50	4	229	62	27	0	203
3W	355	16	1	7	127	219	24	9	379	0	0	0	360	14	1	4
3w	1710	261	17	96	48	997	989	50	0	2084	0	0	1348	567	23	146
3H	345	181	61	55	4	61	573	4	0	0	642	0	202	286	73	81
3?	91	18	0	204	2	22	60	229	0	0	0	313	49	30	0	234
4W	1766	128	14	51	165	972	760	62	360	1348	202	49	1959	0	0	0
4w	538	293	21	45	15	239	616	27	14	567	286	30	0	897	0	0
4H	40	17	38	2	0	13	84	0	1	23	73	0	0	0	97	0
4?	157	38	6	264	1	75	186	203	4	146	81	234	0	0	0	465

表4　Burt表の例：設問、回答カテゴリは、表3と同じ。(Greenacre 2017 p141 Exhibit 18.4)

こうして2変数対応CAの多変数化は実現するのだが、カテゴリが増えることによって、生成される座標軸が体現する慣性（分散）が小さくなってしまうという問題に直面することになる。これはカテゴリ数が増えることによって避けられない現象である。そこで、多重対応分析MCAは、標準形をindicator行列版とBurt行列版としながら、いくつもの改良版へと発展していくことになる[2)]。そのため、分析においては、用いているのがどのMCAなのかが明示されなくてはならない。

ちなみに、CA/MCAのファミリのリストが、Beh, Lombardo 2014,32-33に掲載されているが、ここにあるだけで33ものCA/MCAが開発されている。

2　GDAという分析フレームワーク

Brigitte Le RouxはLe Roux, Rouanet 2010などで、MCAを使えばブルデューのような分析ができるわけではない、と書いている。これだけみれば、計量分析手法であるMCAを使ったとしても、それから自動的に社会学的分析になるわけではないのは当たり前ではないかと思うかもしれない。それはそうなのだが、Le Rouxの指摘は、なによりもMCAの使い方にある。

分析における社会学的視点の必要性は当然として、MCAの使い方も単なる数理的技法としてではなく、分析の手法、考え方として理解する必要があるということである。対応分析の始祖ベンゼクリの伝統で、対応分析が「データ解析」（Analyse des Donnees）と同義で用いられている、ということが言われるが（Clausen1989=2015:3）、それには、探索的データ解析というアプローチをふくめたデータ解析のフレームワークという意味がこめられている。この点は、日本で対応分析に相当する手法を数量化手法（III類）としてベンゼクリに10年先行して開発、実用化した林知己夫のスタンスと共通するものがあると言っていいだろう（数量化xx類という呼称は林がつけたものではなく、飽戸弘が命名したものである。林にとっては、まずデータがあり、それをいかに分析するのか、ということはデータに即して考えるという立場であった。林2001他）

こうした意味では、このGDAという分析フレームワークがベンゼクリの提唱した「データ分析」の正統な発展系列ということになる。

冒頭、回帰分析を「変数の社会学」として批判し、「関係性」を体現する手法としての多重対応分析が選ばれた、と書いたが、Le Rouxたちの立場は、回帰分析vs CA/MCAではない。

このGDAは、まず手に入れたデータの構造を確認する、ということをポイント・クラウドの取得として強調する。そこでの個体ポイントは、平均と分散に抽象化されずに個体として維持されるため、グラフ上のポイントからデータ表の個体番号を参照することが可能である。そこで、MCAによって全体構造を把握したうえで、インタビューイを選択することが可能になる。ベネット et al ではこの手法が意識的に用いられている（ベネット 2010=2017:92「三つ目」の有益な特徴）。また、GDAの実習テキストでもある Le Roux, Rouanet 2010, 第1章のタイトルの下には「量と質の間には幾何学がある」と書かれている。これはGDAが混合研究法であることの宣言でもある。

2.1 実験パラダイムと観察データ

さて、実験計画によって適切にデザインされた実験であれば「要因」による「効果」は、統計的処理によって「誤差項」と分離されて評価可能になる（Fisher 1925 他）。分散分析によってそれは行われる。しかし、社会調査データのような観察データは、そのようなデザインが入り込む余地がない。そこで、考案されたアプローチがGDAであり、その中心的技法の部分は、サプリメンタリ・ポイントとアクティブ変数の分離（＝デザイン）による構造の構築（構造化モデリング）と、その構造を前提にした、サブ・クラウドに注目した構造化データ解析（SDA）が続き、さらにそこを踏まえた統計的推定を行う、帰納的データ解析（IDA）が用意されている[3]（図4）。つまり、GDAとはPCA、CA、MCAを幾何学的にデータを扱う基本ツールとして位置付けながら、実験パラダイムとは異なる観察データに対して、変数サブ・クウラド（要因）の全体の中での位置を評価していくための分析フレームワークなのである（図4）。

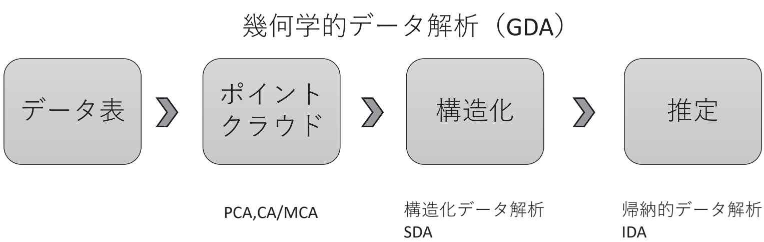


図4　GDAの構成要素

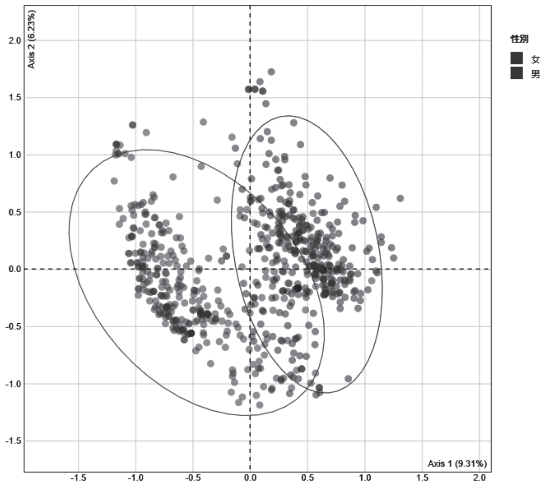


図5が示しているのは、プロットされた個体が有している回答カテゴリ（ここでは性別：男/女）の分布状況である。楕円に重なる部分がなければ、独立であるし、ほぼ重なっていれば、違いはない、ということになる。

図5　ある調査結果の個体クラウドと性別カテゴリの集中楕円

2.2 サプリメンタリ・ポイントからサブ・クラウド分析へ

サプリメントポイントをもちいた構造上への注目変数の投影、prediction は、CA や MCA の活用領域を拡げはしたが、その限界はすぐに明らかとなった。 Le Roux は、次のように述べる。

「サプリメンタリ変数の技法は、構造化データ解析にむけた一歩ではあった。しかし、この技法の限界は、サプリメンタリ・カテゴリがサブクラウドの平均ポイント（これはMCA の基本的特性である）に注意を留め、サブクラウドの分散を無視するという事態を考えるならば、明らかとなる。それゆえ『ディスタンクシオン』において、ブルデューは、最も強力な説明要因としての個体空間の等高線として形式化されたclass（区分）へと向ったが、それが構造化データ分析の精神であった。」Le Roux, Rouanet 2010:

2.3 集中楕円によるサブ・クラウドの要約

GDA のステップとしての構造化データ解析で、サブクラウドの集中楕円が導入される。これは先の説明にあるように、個体ポイントにおいて、ある設問のあるカテゴリに回答の「平均」を中心として等高線として描かれた楕円である。つまり、ある回答の位置がこの平均ポイント（中心）と広がり（分散）をもった楕円に表現されている。

ベネット ,et al 2009=2017 p103 にある、「このライフスタイル空間に社会人

口学的特徴についての「補足的変数」（引用者注：これまでサプリメンタリ変数と呼んできたもの）を重ね合わせ、各集団の平均の点がどこに位置づいているのかを見ることができる（11）。」この注11には「この背後にある統計的手続きが、それぞれの軸の社会的カテゴリーについての級内および級間分散の計算である」とあるが、ここでの「級」とはサブ・クラウドのことである[4)]。また、p105から参照されている図3-6は、まさにその集中楕円の例である[5)]。

なお、この集中楕円のところででてくるη2乗は、総分散に対するカテゴリ間分散の割合であり（Le Roux, Rouanet 2010:22）、そのカテゴリの軸への寄与が表されている。

ベネット et al は、「方法的補遺」を含めて、この幾何学的データ解析の非常に有益な実例なのである。

結語

以上の検討をまとめると以下のようになる。

1　対応分析（CA）は、多次元カテゴリカルデータを分析対象とし、応答の似たものを近くに似てないものを遠くに配置する。原点は、分析対象としている系の重心＝平均であり。ここに近いものは平均的な性格をもっている、ということである。

2　2変数の表（クロス表）に対する対応分析が基本であるが、この表を拡張し、多変量（個体×変数）データを扱えるようにしたものが多重対応分析（MCA）である。

3　CAにおいてもMCAにおいても、空間形成に関与しないサプリメンタリポイント（変数）を用いて、形成された空間に投影するという技法を用いることができる。この技法の一つの応用が、幾何学的データ解析（GDA）である。

4　サプリメンタリ変数を用いる技法は、LeRoux等によって、特定カテゴリをサプリメンタリ処理するMCAとして定式化され、Specific MCAと呼ばれている。また、サブ・クラウドに対するMCAは、Class specific MCAとして定式化されている（LeRoux & Rouanet 2010）。後者は、Greenacreがサブセット MCAと呼ぶものである（Greenacre 2017）。

5　GDAは、実験計画法的な構造の構築が不可能な社会調査データのような観察データに対して、「要因＝カテゴリ」の影響を評価可能にする試みである。それはGDAという枠組みのなかで、三つのパートに区分される。

1) まずもっとも基本的な部分としてのデータの構造化。これは、Active 変数とサプリメンタリ変数を区分することで始まる。これによって、構造上に注目している変数がどのような関係にあるかをみることができる。2) 続いて、この構造の解釈としてのサブクラドに注目する構造化データ解析である。ここでは、サブクラウドの慣性（分散）をあらわす集中楕円によってサブクラウドが要約される。また、このサブクラウドに関する、級間慣性と全体慣性の比、η 二乗によって、そのサブクラウドの影響が数量化されている。3) さらに、帰納的データ解析（IDA）によって、統計的推論がおこなれる。そこでは、まず「記述」が優先され、推定はその後とされる。本稿では扱えないが、典型性検定、等質性検定を行い、サブクラウドの信頼楕円を評価する。Rouanet、Le Roux たちは、そこからベイジアン推定へと展開する。

以上のまとめを、分析条件の実験パラダイム－観察データ、また、量的変数－質的変数という区分に当てはめてみると図 6 のように整理することができる。

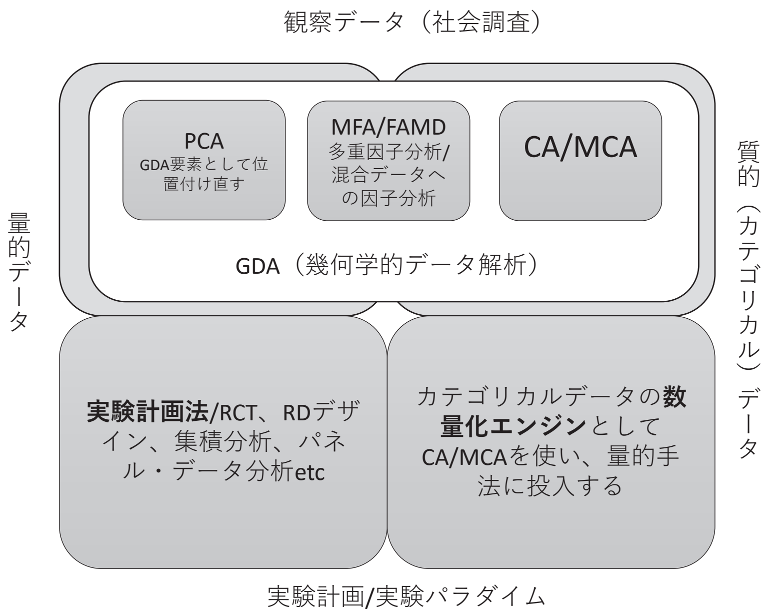


図 6　GDA と実験パラダイム、観察データ

注

1）レジュメ：https://speakerdeck.com/419kfj/how-correspondence-analysis-express-the-relations
スライド：https://www.slideshare.net/kazuofujimoto/jss20191006-ts13-gdaopen?qid=11cd4b91-c724-4aeb-9eea-9229fd1bcc5a&v=&b=&from_search=1

2）GDAにとって重要な拡張された多重対応分析（MCA）。csMCA class specific MCA 級に注目したMCA。全体の個体を参照しながら、特定の個体（級）に対するMCA。ex）学生生活の満足度が最低の学生個体を分析対象にする。Le Roux 2010、GDATools解説
speMCA specific MCA 不必要なカテゴリ（juck カテゴリ）をpassiveカテゴリとして行うMCA。1）度数が5以下のもの、2）解釈できない回答 - どちらでもない / 無回答、などをpassiveとする。
speMCAもcsMCA（ともにLe Roux & Rouanet）は、Greenacreが提案する、subsetMCAである。

3）帰納的データ解析 Inductive data analysis：IDA　ここで使われている帰納的は、Fisherが、1970「実験計画法」で述べた文脈での帰納的である（Fisher1935=1971、序章）。そこでは、有名な逆確率批判（ベイズ批判）が述べらており、「Bayesは帰納的推論、すなわち観測上の事実からそれを説明できる理論への議論に関する、正確で定量的な理論を展開することの重要性を認めた人として、ヨーロッパでは最初の人であったと見られるが、そのことは確かに科学史における地位を主張するのに十分である」という。このあとその理論（ベイズの理論＝逆確率）がいかに誤りであるのかが続く。Le Rouxたちはこれを踏まえて、GDAにおける推論パートをIDA（帰納的データ解析）と名付けている（Rouanet et al,1998）。

4）クラウド、クラス、クラスター　「初期のころは観測値のグループはクラス（級）という言葉が使われていたが、サーベイサンプリングや実験計画でクラスターサンプル（集落抽出法）が広く使われていることから「クラスター」という言葉がより用いられている。しかしながら「級内」という言葉は、クラスター内の相関を言及するものとして多く使われている。」級内相関係数、『統計科学百科事典』2-394。　このグループが、GDAにおいてはクラウドと呼ばれている。　参考：「VII. 級内相関と分散分析」フィッシャー 1925=1969、p67

5）磯他の翻訳では、「集中」が省略されており、ただ楕円となっているが、原文は、concentration ellipsesであって、これはGDA/SDAでの重要な分析概念である。それ以降の図3-7、図3-9も「集中楕円」である。

参考文献

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

ブルデュー .P（訳：石井洋次郎 1990）『ディスタンクシオン』藤原書店

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）

Fisher, R. A, 1969（1925）『研究者のための統計的方法』森北出版株式会社

Fisher, R. A, 1971（1935）『実験計画法』森北出版株式会社

藤本一男 , 2017,「対応分析のグラフを適切に解釈する条件」『津田塾大学紀要』第 49 号 , 141-153

藤本一男 , 2018,「プログラミング言語 R における 2 つの mosaic plot と日本語、多言語表示」『津田塾大学紀要』第 50 号 , 129-146

藤本一男 , 2019,「『Supplementary』変数から多重対応分析（MCA）を考える」『津田塾大学紀要』第 51 号 , 155-167

Friendly. M, 2002, “A Brief History of the Mosaic Display” , Journal of Computational and Graphical Statistics, Vol.11, No.1（Mar., 2002）, pp.89-107, American Statistical Association

Frendly. M, Meyer. D, 2016, “Discrete Data Analysis with R”, CRC Press

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）

Greenacre, 1984, “Theory and Application of Correspondence Analysis” , Academic Press.（http://www.carme-n.org/?sec=books5 PDF でダウンロート可能。）

林知己夫 , 2001,『データの科学』シリーズ〈データの科学〉1, 朝倉書店

Husson. F, 2017, “Exploratory Multivariate Analysis by Example Using R 2nd Edition”, CRC Press

Hjellbrekke. Johs,2019,MULTIPLE CORRESPONDENCE ANAYSIS for the Social Sciences, ROUTLEDGE

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. ルバール他 , 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,

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

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（仮訳 :「調査票の幾何学的分析 ブルデューの『ディスタンクシオン』から学ぶ」http://bit.ly/2zWLeCH）

Rouanet, H., Bernard, J.-M., Bert,M.C., Lecoutre, B., Lecputre,M.P.,& Le Roux, B, 1998, New way in statistical methodology : From significance test to Bayesian methods, Bern, Switzerland: Peter Lang

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）

R Core Team（2019）. R: A language and environment for statistical computing. R Foundation for Statistical

Computing, Vienna, Austria. URL https://www.R-project.org/.

Appendix A 対応分析の数理

対応分析の数理的要約

分析対象表（プロファイル行列 P）の標準化残差 S を求める。ここで、r は行和ベクトル。c は列和ベクトルである。いわゆる周辺度数。

1. ステップ 1

$S = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$

この S の特異値分解（SVD）を行う。D の要素が特異値。

$D_r^{-1/2}$ は ${}_r^{-1/2}$ を要素とした対角行列。T は転置行列。

$S = \mathrm{U}D_\alpha V^T$ ここで、$U^TU = V^TV = I$

2. 座標の計算

ここで得られた結果から標準座標（standard coordinate）（Φ、Γ）、主軸座標（principal coordinate）（*F*、*G*）が計算される。

$\Phi = D_r^{-1/2}\mathrm{U}$

$\Gamma = D_c^{-1/2}\mathrm{V}$

$F = D_r^{-1/2}UD_\alpha = \Phi D_\alpha$

$G = D_c^{-1/2}VD_\alpha = \Gamma D_\alpha$

3. 固有値 / 特異値

特異値分解の結果得られた *D*α は固有値を成分とした対角行列である。この操作の結果、n 次元空間であった行列 P は、固有値（として表現される P の分散＝慣性）ごとの軸に分解されていく。

Appendix B 対応分析に関するソフトウェア

CA/MCA 処理

FactoMineR http://factominer.free.fr/

ca http://www.carme-n.org/?sec=ca

GDATools https://cran.r-project.org/web/packages/GDAtools/index.html

Viewer

factoextra http://www.sthda.com/english/wiki/factoextra-r-package-easy-multivariate-data-analyses-and-elegant-visualization

factoshiny http://factominer.free.fr/graphs/factoshiny.html

explor https://cran.r-project.org/web/packages/explor/vignettes/introduction_en.html