北海道大学工学部研究報告
第88号（昭和53年）

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 88 (1978)

潜在クラス分析におけるある手法の開発について

須川和明* 佐藤義治* 河口至商* 金田 嶢**

（昭和52年12月28日受理）

On the Development of a Method for Latent Class Analysis

Kazuaki SUGAWA, Yoshiharu SATO, Michiaki KAWAGUCHI
and Takashi KANETA

(Received December 28, 1977)

Abstract

In latent class analysis, the main problem is to estimate the latent parameters. Its difficulties are due to the fact that the number of estimated parameters is extremely large as compared with the observations, in addition the structure equations are nonlinear. The traditional methods are reduced to the eigenvalue problems. But, in some of these methods, the estimates depend on the selection of items. In this paper, the solution is obtained by directly solving the structure equations using M. J. D. Powell's algorithm for the nonlinear algebraic equations. Because of the direct solution, this has an advantage over other solutions reported hitherto

1. 序　　論

潜在構造分析は，P. F. Lazarsfeld によって提案された手法であり，いくつかの質問項目によって，ある集団の態度を測定したときに，その構造を潜在変数を仮定することにより，その確率構造として表現しょうとするものである。

ここで述べる潜在クラス分析は潜在構造分析の特殊な場合であり，潜在変数の確率構造が離散的である場合に相当する。これらの分析におけるデータは，何個かの質問項目に対する各個体の反応が2者択一的なものからなる。すなわち，反応は yes, no, 賛成，反対等であり，これを1又は0で表現する。潜在クラス分析では，潜在変数は離散的な値をとるものとし，それを x^1, x^2, $\ldots, x^m$ とし，それらの確率をν^1, $\nu^2, \ldots, \nu^m$ とする。また各潜在変数の値において，各個体が，1又は0と反応する確率は各質問項目について互いに独立である（局所独立性）と仮定する。各潜在変数の値をとる各個体の集合を潜在クラスと呼ぶ。潜在クラスの数を m と仮定し，各クラス α 内において，各質問項目 i に1と反応する確率を λ_i^α とする。以上の仮定のもとに潜在クラス分析のモデルはつぎのように表わされる。

$$1=\sum_{\alpha=i}^{m}\nu^\alpha$$

$$\pi_i=\sum_{\alpha=1}^{m}\nu^\alpha\lambda_i^\alpha$$

---

*　工学部　情報数理工学第一講座
**　苫小牧工業高等専門学校

$$\pi_{ij}=\sum_{\alpha=1}^{m}\nu^{\alpha}\lambda_i^{\alpha}\lambda_j^{\alpha}$$

$$\pi_{ijk}=\sum_{\alpha=1}^{m}\nu^{\alpha}\lambda_i^{\alpha}\lambda_j^{\alpha}\lambda_k^{\alpha}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

ここで π_i, π_{ijk}, はデータから推定可能な量である。すなわち，π_i はデータ全体において質問項目 i に1と反応する確率，π_{ij} は質問 i と j に同時に1と反応する確率であり，π_{ijk} 等も同様な確率である。モデルにある ν^{α}, λ_i^{α} を潜在パラメータという。この分析の主要な目的は，この潜在パラメータを推定することである。

潜在パラメータを推定する手法は，現在までにいくつか考え出されており，実用化されている。Lazarsfeld は潜在クラスの数と質問の個数が特別な場合の解法を与えたが，潜在パラメータの一般的な推定法は B. F. Green[1] によって始めて与えられた。この方法は，モデル式を行列表現し，対称行列の固有値問題に帰着させることによって，比較的簡単な計算手続きで潜在パラメータを推定することができるが，潜在パラメータの推定の前に観測不可能な値 (π_{ii}, π_{iii}, ... 等) の推定を必要とするため，質問項目の少ない場合には，あまり良い推定値を得ることはできない。T. W. Anderson[2] はそのような値を必要としない推定法を与えたが，非対称行列の固有値問題となるため計算は Green の場合に較べて複雑である。又，Green の方法とは違い，一度に推定できるパラメータの数に制限があるため，推定値を一意的に決めることができない。これは，推定に際してすべてのデータを使用できないためである。W. A. Gibson[3] は Anderson の手法を一般化して，一度に全部の潜在パラメータを推定する手法を提案しているが，これは，パラメータの推定時に質問項目をいくつかに分割することが必要となり，この分割のし方に解が従属する。さらに，A. Madansky[4] により高次のモデル式を用いる方法が提案されているが，計算が複雑である。

以上の解法はすべて，モデル式を行列表現することにより固有値問題を解く方法であったが，本報告ではモデル式の非線型連立方程式を M. J. D. Powell[5] による非線型連立方程式の数値解法アルゴリズムを使って，潜在パラメータを直接推定する方法を試みた。Powell のアルゴリズムの概要と計算に必要な初期パラメータの設定を以下に述べる。さらに，本手法の応用計算例と推定結果に対する考察をする。

2. 潜在パラメータの推定法

潜在クラス分析におけるモデル式は，序論で述べたように潜在パラメータ $\nu^{\alpha}, \lambda_i^{\alpha}$ $(i=1,2,\ldots,k,\alpha=1,2,\ldots m)$ を未知数とする非線型の連立方程式であることから，これを，Powell による非線型連立方程式の解法アルゴリズムを使って解くことを考えてみた。そのためにモデルの方程式を以下のように書き換えることにする。

$$f_1=\sum_{\alpha=1}^{m}\nu^{\alpha}-1=0$$

$$f_l=\sum_{\alpha=1}^{m}\nu^{\alpha}\lambda_i^{\alpha}-\pi_i=0 \qquad (l=2,\ldots,k+1)$$

$$f_l=\sum_{\alpha=1}^{m}\nu^{\alpha}\lambda_i^{\alpha}\lambda_j^{\alpha}-\pi_{ij}=0 \qquad (l=k+2,\ldots,k(k+1)/2+1)$$

$$f_l=\sum_{\alpha=1}^{m}\nu^{\alpha}\lambda_i^{\alpha}\lambda_j^{\alpha}\lambda_k^{\alpha}-\pi_{ijk}=0 \qquad (l=k(k+1)/2+2,\ldots,(k+1)(k^2-k+6)/6)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

Powell のアルゴリズムの反復計算は次の 4 つの部分から成っている。

(1) 解の修正ベクトル $\boldsymbol{\delta}$ の計算。

(2) 各ステップの長さ Δ の修正。

(3) ヤコビアン J の修正。

(4) $\boldsymbol{\delta}$ の独立性の維持。

また，反復計算に必要な値は，方程式の解の推定値，ヤコビアン J とその逆行列 J^{-1}，n ステップ前までの $\boldsymbol{\delta}$ 方向を表わす行列 Ω，ステップの長さ Δ，である。方程式の解ベクトル $\boldsymbol{x}$ に対する修正ベクトル $\boldsymbol{\delta}$ はニュートン法と $F(\boldsymbol{x})=\sum_{i=1}^{n}\{f_i(\boldsymbol{x})\}^2$ に関する最急降下法を組み合せた方法で，$|\boldsymbol{\delta}|\leqq\Delta$ となるものが計算される。2 つの方法のバランスは Δ に依存して，Δ が大きいときにはニュートン法に重みが置かれ，Δ が小さいときには最急降下法に重みが置かれる。計算された $\boldsymbol{\delta}$ はその方向が n ステップ前までの $\boldsymbol{\delta}$ の方向と独立（ここでの独立の概念は一般の線型代数での独立性とは異なる。）であるかどうかが行列 Ω によって判定され，さらに $F(\boldsymbol{x}+\delta)<F(\boldsymbol{x})$ を満たすかどうかが判定される。Δ は反復回数を少なくするためには，大きい方が望ましいが，$F(\boldsymbol{x}+\delta)\geqq F(\boldsymbol{x})$ となる場合には小さくするように各ステップごとに修正される。ヤコビアン J は，適当に与えられた DSTEP に対して，

$$J_{ij}=\frac{f_i(\boldsymbol{x}+\mathrm{DSTEP}\ \boldsymbol{e}_j)-f_i(\boldsymbol{x})}{\mathrm{DSTEP}} \qquad (i, j-1, 2, \ldots, n)$$

として近似的に計算される。

計算を始めるためには，方程式の解の初期値と 4 つの初期パラメータ DSTEP, DMAX, ACC, MAXFUN が必要である。DSTEP はヤコビアンの近似計算のためと，Δ の下限を与えるために使われる。計算機の丸め誤差の影響を受けない限りは小さい程望ましい。DMAX は Δ の上限

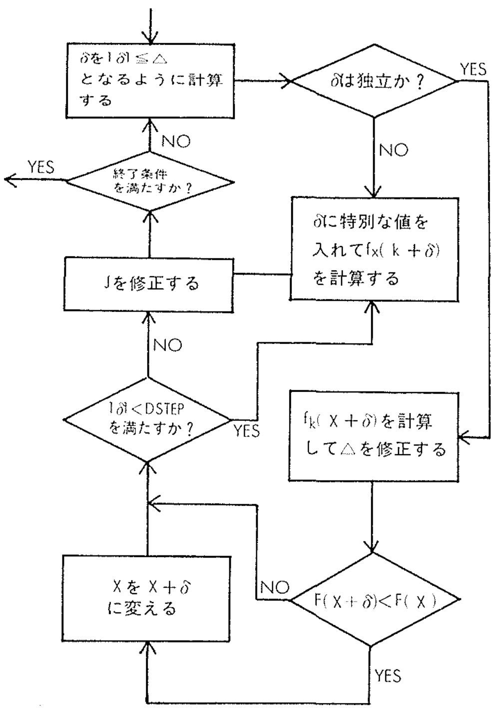

図1 反復計算のフローチャート

を与える。このためには，初期解と真の解との距離の大よその推定値を DMAX とすることができる。ACC は解の精度を与える。すなわち，$F(\boldsymbol{x}) \leqq \mathrm{ACC}$ となったとき計算は終了する。MAXFUN は反復計算を確実に終了させるために使われる。$f_i(\boldsymbol{x})$ の計算の回数が MAXFUN を越えた時に計算は終了する。

以上が Powell のアルゴリズムの概要であるが，参考のためにフローチャート*) を図1に示す。尚，このアルゴリズムによるサブルーチンプログラムが Powell[5]の Appendix に示されている。

3. 本手法の応用とその考察

前章で述べた Powell による非線型連立方程式の解を求めるサブルーチンプログラムを使って，2つのデータについて計算を行なってみた。まず最初は，表1のように潜在パラメータを仮定して，架空のデータを作ることにする。

この値をもとにして，1,000人の人に質問したと仮定したときの各応答パターンの人数を計算すると表2のようになる。これを与えられたデータとすると，質問 i に1と応答する比率 π_i, 質問 i と j に同時に1と応答する比率 π_{ij}, 質問 i と j と k に同時に1と応答する比率 π_{ijk} を計算することができる。それを表3に示す。

これをもとに，Powell のサブルーチンを使って潜在パラメータの計算を行なった。計算に必要な初期値はそれぞれ次のように与えた。解の初期値はすべて 0.3 とする。サブルーチンの初期パラメータはそれぞれ $\mathrm{DSTEP}=10^{-4}$, $\mathrm{ACC}=10^{-15}$, MAXFUN=300, DMAX=1.0 とする。その結果，計算は79回反復した後，残差の2乗和 $F(\boldsymbol{x})$ を減少させることができなくなって終了した。最終的な解と残差の2乗和が表4であるが，この結果と最初に仮定した潜在パラメータを比較すると，完全に一致していることがわかる。

表1

		質問		
		1	2	3
潜在クラス1	$\nu^1=0.6$	$\lambda_1^1=0.35$	$\lambda_2^1=0.40$	$\lambda_3^1=0.85$
潜在クラス2	$\nu^2=0.4$	$\lambda_1^2=0.75$	$\lambda_2^2=0.95$	$\lambda_3^2=0.25$

表2

反応パターン			
1	2	3	人数
1	1	1	143
1	1	0	111
1	0	1	156
1	0	0	200
0	1	1	226
0	1	0	30
0	0	1	95
0	0	0	39

表3

	反応率
π_1	0.51
π_2	0.62
π_3	0.61
π_{12}	0.369
π_{13}	0.2535
π_{23}	0.299
π_{123}	0.14265

*) M. J. D. Powell, A Fortran Subroutine For Solving Systems of Nonlinear Algebraic Equations から転載。

表 4

		質問		
		1	2	3
潜在クラス 1	$\hat{\nu}^1=0.60$	$\hat{\lambda}_1^1=0.35$	$\hat{\lambda}_2^1=0.40$	$\hat{\lambda}_3^1=0.85$
潜在クラス 2	$\hat{\nu}^2=0.40$	$\hat{\lambda}_1^2=0.75$	$\hat{\lambda}_2^2=0.95$	$\hat{\lambda}_3^2=0.25$
残差 2 乗和		0.3553×10^{-13}		

表 5-1

反応パターン			
1	2	3	人数
1	1	1	61
1	1	1	44
1	0	1	89
0	1	1	20
1	0	0	71
0	1	0	49
0	0	1	39
0	0	0	131
合計			504

表 5-2

	反応率
π_1	0.525793
π_2	0.345238
π_3	0.414683
π_{12}	0.208333
π_{13}	0.297619
π_{23}	0.160714
π_{123}	0.121032

表 6

		質問		
		1	2	3
潜在クラス 1	$\hat{\nu}^1=0.5831$	$\hat{\lambda}_1^1=0.8206$	$\hat{\lambda}_2^1=0.4103$	$\hat{\lambda}_3^1=0.6077$
潜在クラス 2	$\hat{\nu}^2=0.4169$	$\hat{\lambda}_1^2=0.1135$	$\hat{\lambda}_2^2=0.2543$	$\hat{\lambda}_3^2=0.1448$
残差 2 乗和		0.1776×10^{-13}		

表 7

		質問		
		1	2	3
潜在クラス 1	$\hat{\nu}^1=0.5863$	$\hat{\lambda}_1^1=0.8174$	$\hat{\lambda}_2^1=0.4101$	$\hat{\lambda}_3^1=0.6074$
潜在クラス 2	$\hat{\nu}^2=0.4137$	$\hat{\lambda}_1^2=0.1134$	$\hat{\lambda}_2^2=0.2544$	$\hat{\lambda}_3^2=0.1449$

次に，金田[7] によって実際に行なわれた質問調査のデーに対して潜在パラメータの推定を行なってみた。この調査は，数学への親近感について，質問を 504 名に行なったもので，その反応パターンの分布と各質問の反応比率が表5 (1)，表 5(2) である。

解の初期値とサブルーチンの初期パラメータを DSTEP $=10^{-5}$，その他は前の例と同じにして計算してみると，127 回反復した後，残差の 2 乗和を減少できなくなって計算が終了した。最終的な結果が表 6 である。参考までに金田が Anderson の方法で推定した結果を表 7 に示す。

2 つの値を比較してみると，ほぼ同じ結果が得られているわけであるが，潜在パラメータの推

定値を $\hat{\nu}^{\alpha}$, $\hat{\lambda}_i^{\alpha}$ として，この推定値を用いて応答パターン z ($z=101$, $z=110$ など) を持つ応答者数の理論度数 $\hat{n}_z$（例えば, $z=101$については, $\hat{n}_z=N\sum_{\alpha=1}^{m}\hat{\nu}_\alpha\hat{\lambda}_1^{\alpha}(1-\hat{\lambda}_2^{\alpha})\hat{\lambda}_3^{\alpha}$ となる。）を計算し，観測度数 n_z に対するあてはまりの良さを示す指標として，$\chi^2=\sum_z(n_z-\hat{n}_z)^2/\hat{n}_z$ を計算してみると，Anderson の方法による場合は $\chi^2=0.0089$ であり，本手法を用いた場合は $\chi^2=0.000000302$ となることから，本手法の結果の方があてはまりの良い推定値であると考えることができる。

ここであげた2つの計算例は，共に質問項目が3で，仮定しているクラスの数が2の場合であった。この場合には，方程式の数と未知数の数が共に8となり等しいために，一意的に解が求められる。ところが一般には，クラスの数を m，質問項目の数を k とすれば，未知数は $m(k+1)$ であり，方程式の数は 2^k となる。方程式の解は $m(k+1)\leqq 2^k$ となる限り求めることができるが，方程式の数が未知数の数より多い場合には一意的に解が決まるとは限らない。しかし，解が存在する限りは，その解はすべての方程式を満たすはずであると考えることができるから，ある方程式の組から計算された解が残りの方程式を満たすのであればその値を求めている解とすることができる。

4. 結　　論

序論で述べたように，Green の推定法には，観測不可能な値を推定する際に用いなければならないという欠点があり，潜在パラメータの推定を困難なものにしていた。また，Anderson の方法は Green の欠点を解決しているが，そのために非対称行列の固有値を求めることが要求され，計算を複雑なものにしている。さらに，Anderson の方法は，クラスの数を m とするとき，一度に $2m^2$ 個の潜在パラメータしか推定できないという欠点を持っていた。Gibson の推定法はこの点を解決しているがデータの分割の方法によって推定値が異なることがある。

この報告で述べた手法は，Anderson の推定法の欠点を解決して，一度に全部の潜在パラメータを推定することができ，さらに，前章の終わりに述べたように方程式の数が未知数の数より多い場合には，方程式の組の選び方によって解が異なることがあるが，この場合にも，残りの方程式がこの値を解に持つかどうかを調べることによって，真の解であるかどうかを判定することができる。

参 考 文 献

1) Green, B. F.: Psychometrika, 16 (1951), pp. 151-166.
2) Anderson, T. W.: Psychometrika, 19 (1954), pp. 1-10.
3) Gibson, W. A.: Psychometrika, 20 (1955), pp. 69-73.
4) Madansky, A.: Psychometrika, 25 (1960), pp. 183-198.
5) Powell, M. J. D.: Numerical Methods for Nonlinear Algebraic Equations (1970), pp. 87-114, Gorden and Breach Science Publishers.
6) Powell, M. J. D.: 同上, pp. 115-161.
7) 金田 嶢：苫小牧工業高等専門学校紀要, 6 (昭46), pp. 67-76.