順序カテゴリー・データ解析 (第 2 報)

辻 谷 将 明*

A REVIEW OF THE ANALYSIS OF ORDERED CATEGORICAL DATA (Part II)

Masaaki TSUJITANI

In the earlier paper, we surveyed the strategies for modeling the two-way contingency tables having ordered categories. Especially, Goodman's association models (1979) were discussed by reanalyzing the well-known sets of data.

This manuscript presents the comparisons of the results by use of association models with those obtained using the canonical correlation analysis to locate the association models in the analysis of ordered categorical data. The examination of the agreement of the association models and Plackett's model (1965) with the bivariate normal may clarify the utility of association models. Then, the inference in contingency tables with ordered categories is mentioned based on Plackett's coefficient of association.

Furthermore, the association models may be generalized to multiway contingency tables. Some examples are illustrated in the context of potential application of the association models to analyze the various types of ordered categorical data.

1. は じ め に

先の報告 (辻谷 (1984)) では, カテゴリーに順序を有する 2 次元分割表について, Goodman の連関モデルを中心に, その解析法の展開過程を整理した。

本稿は, 前報の続きとして, 2 次元分割表における連関モデルの位置づけとその特徴を明瞭に把握するため, 正準相関分析と比較する (Goodman (1981 a)). また, 2 変量正規性という観点に注視し, 連関モデルが Plackett (1965) モデルより有効であることを述べ (Goodman (1981 b)), 分割係数に基づく推測法も取り上げる。そして, 連関モデルの適用範囲を広げるため, 多次元表 (特に, 3 次元表) への拡張について論述する。

これらの解析法が容易に実地適用されるように, 具体例を通じてそれらの有効性を立証する。

2. 連関モデルと正準相関

2.1 分割表の正準相関分析

2 次元分割表において, (i, j) セルの観測度数を $n_{ij} (i=1, 2, \dots, I; j=1, 2, \dots, J)$, (第 1) 正準スコア

における第 i 行カテゴリーのスコアを x_i , 第 j 列カテゴリーのスコアを y_j とする。そして

$$\bar{x} = \frac{1}{n} \sum_i n_{i\cdot} x_i, \quad \bar{y} = \frac{1}{n} \sum_j n_{\cdot j} y_j,$$

$$\text{Var}(x) = \frac{1}{n} \sum_i n_{i\cdot} (x_i - \bar{x})^2,$$

$$\text{Var}(y) = \frac{1}{n} \sum_j n_{\cdot j} (y_j - \bar{y})^2$$

$$\text{Cov}(x, y) = \frac{1}{n} \sum_i \sum_j n_{ij} (x_i - \bar{x})(y_j - \bar{y})$$

を定義する。ただし

$$n_{i\cdot} = \sum_j n_{ij}, \quad n_{\cdot j} = \sum_i n_{ij}, \quad n = \sum_i \sum_j n_{ij}$$

である。このとき, 分割表の正準相関分析 (これは, 双対尺度法と呼ばれることがある (西里 (1982))) は, 次のような最適化問題と等価になる (Kendall & Stuart (1979: 33 章), 柳井ら (1977)):

$$\left\{ \begin{array}{l} \text{制約条件} \\ \bar{x} = \bar{y} = 0, \text{Var}(x) = \text{Var}(y) = 1 \\ \text{のもとで} \\ \text{Cov}(x, y) \longrightarrow \text{最大化} \end{array} \right.$$

* 神戸女子大学 (Kobe Women's University)

具体的には、ラグランジェ未定乗数 λ, κ を導入し、関数

$$F = \sum_i \sum_j n_{ij} x_i y_j - \frac{\lambda}{2} \left(\sum_i n_i x_i^2 - n \right) - \frac{\kappa}{2} \left(\sum_j n_{\cdot j} y_j^2 - n \right)$$

について

$$\begin{cases} \frac{\partial F}{\partial x_i} = \sum_j n_{ij} y_j - \lambda n_i x_i = 0; & i=1, 2, \dots, I \\ \frac{\partial F}{\partial y_j} = \sum_i n_{ij} x_i - \kappa n_{\cdot j} y_j = 0; & j=1, 2, \dots, J \end{cases} \quad (2.1)$$

$$(2.2)$$

を解けばよい。(2.1)式の両辺に x_i を掛け、 i について和をとると

$$\lambda R =$$

となる。ここに、 R は正準相関係数で

$$R = \frac{1}{n} \sum_i \sum_j n_{ij} x_i y_j$$

とする。

$$N = \left(\frac{n_{ij}}{(n_i n_{\cdot j})^{1/2}} \right)_{I \times J}$$

$$u = (u_i)_{I \times 1} = (x_i n_i^{1/2})_{I \times 1}$$

とおけば、一般固有値問題

$$NN^T u = R^2 u$$

に帰着できる。ただし、 T は転置行列を表す。ゆえに、

$$\begin{cases} x_i = \frac{u_i}{n_i^{1/2}} \\ y_j = \frac{1}{R n_{\cdot j}} \sum_i n_{ij} x_i \end{cases}$$

から、正準スコア x_i, y_j を推定する。

2.2 RC 連関モデルと正準相関分析との比較

あるモデルのもとでの n_{ij} の期待度数を m_{ij} とする。Goodman (1979, 1981 a) は、RC モデル

$$m_{ij} = \alpha_i \beta_j e^{\phi \mu_i \nu_j} \quad (2.3)$$

を考察した。ここに、パラメータ μ_i を第 i 行カテゴリーのスコア、 ν_j を第 j 列カテゴリーのスコア、 $|\mu_i - \mu_{i+1}|$ と $|\nu_j - \nu_{j+1}|$ を連続したカテゴリー間の距離 (スコア間隔) と見なす。

本節では、これらのスコア μ_i, ν_j を、正準相関分析から得られる正準スコア x_i, y_j と比較したい。 x_i, y_j は制約

$$\begin{cases} \sum_i x_i p_{i\cdot} = 0; & \sum_i x_i^2 p_{i\cdot} = 1 \\ \sum_j \nu_j p_{\cdot j} = 0; & \sum_j \nu_j^2 p_{\cdot j} = 1 \end{cases}$$

を満たす。ここに

$$p_{i\cdot} = n_i/n, \quad p_{\cdot j} = n_{\cdot j}/n$$

とする。一方、RC モデルは、 μ_i, ν_j をそれらの1次式で置換えても不変であるから、一般性を失うことなく、制約条件

$$\begin{cases} \sum_i \mu_i P_{i\cdot} = 0; & \sum_i \mu_i^2 P_{i\cdot} = 1 \\ \sum_j \nu_j P_{\cdot j} = 0; & \sum_j \nu_j^2 P_{\cdot j} = 1 \end{cases} \quad (2.4)$$

を課すことができる。ここに

$$P_{i\cdot} = m_i/n, \quad P_{\cdot j} = m_{\cdot j}/n$$

$$m_i = \sum_j m_{ij}, \quad m_{\cdot j} = \sum_i m_{ij},$$

$$n = \sum_i \sum_j m_{ij}$$

と書ける。

さて、正準スコア x_i, y_j は、

$$\begin{cases} \sum_j y_j \frac{n_{ij}}{n_i} = \lambda x_i \\ \sum_i x_i \frac{n_{ij}}{n_{\cdot j}} = \lambda y_j \end{cases} \quad (2.5)$$

を満たす。 x_i, y_j と最尤推定量 $\hat{\mu}_i, \hat{\nu}_j$ との関係についても論じよう。

RC モデルのもとでの m_{ij} の最尤推定量 $\hat{m}_{ij}$ は、尤度方程式

$$\hat{m}_{i\cdot} = n_i \quad (2.6)$$

$$\hat{m}_{\cdot j} = n_{\cdot j} \quad (2.7)$$

$$\sum_j \hat{\nu}_j \hat{m}_{ij} = \sum_j \hat{\nu}_j n_{ij} \quad (2.8)$$

$$\sum_i \hat{\mu}_i \hat{m}_{ij} = \sum_i \hat{\mu}_i n_{ij} \quad (2.9)$$

を満たす。ここに、 $\hat{m}_{i\cdot} = \sum_j \hat{m}_{ij}$, $\hat{m}_{\cdot j} = \sum_i \hat{m}_{ij}$ とする。

(2.8)式は、(2.6)式を用いると

$$\sum_j \hat{\nu}_j \hat{m}_{ij} / \hat{m}_{i\cdot} = \sum_j \hat{\nu}_j n_{ij} / n_i \quad (2.10)$$

と書き換えられる。同様に、(2.9)式は(2.7)式から

$$\sum_i \hat{\mu}_i \hat{m}_{ij} / \hat{m}_{.j} = \sum_i \hat{\mu}_i n_{ij} / n_{.j} \quad (2.11)$$

となる。パラメータ μ_i, ν_j に、制約条件(2.4)式を課したとき、対応する最尤推定量は

$$\left\{ \begin{aligned} \sum_i \hat{\mu}_i \hat{P}_{i.} &= 0, \quad \sum_i \hat{\mu}_i^2 \hat{P}_{i.} = 1 \\ \sum_j \hat{\nu}_j \hat{P}_{.j} &= 0, \quad \sum_j \hat{\nu}_j^2 \hat{P}_{.j} = 1 \end{aligned} \right\} \quad (2.12)$$

を満たす。ここに

$$\hat{P}_{i.} = \hat{m}_{i.} / n, \quad \hat{P}_{.j} = \hat{m}_{.j} / n$$

とする。(2.6)式と(2.7)式から、(2.12)式は

$$\left\{ \begin{aligned} \sum_i \hat{\mu}_i p_{i.} &= 0, \quad \sum_i \hat{\mu}_i^2 p_{i.} = 1 \\ \sum_j \hat{\nu}_j p_{.j} &= 0, \quad \sum_j \hat{\nu}_j^2 p_{.j} = 1 \end{aligned} \right\} \quad (2.13)$$

となる。ゆえに、(2.3)式より、最尤推定量 $\hat{\mu}_i, \hat{\nu}_j$ は、対応する正準スコア x_i, y_j と同じ制約条件を満たす。

次に、 $\hat{\mu}_i, \hat{\nu}_j$ が対応する正準スコア x_i, y_j と一致するための条件を調べてみよう。いま、

$$\sum_j \hat{\nu}_j \hat{m}_{ij} / \hat{m}_{i.} = \lambda' \hat{\mu}_i \quad (2.14)$$

$$\sum_i \hat{\mu}_i \hat{m}_{ij} / \hat{m}_{.j} = \lambda' \hat{\nu}_j \quad (2.15)$$

を満たす場面を考える。ただし、 λ' は指定されていない。

なお、(2.6)、(2.8)式を(2.14)式へ代入すると

$$\sum_j \hat{\nu}_j n_{ij} = \lambda' \hat{\mu}_i n_{i.}$$

を得、上式の両辺に $\hat{\mu}_i$ を掛け、 i に関する和をとると

$$\lambda' = \sum_i \sum_j \hat{\mu}_i \hat{\nu}_j p_{ij}$$

となる。ゆえに、 λ' は $\hat{\mu}_i$ と $\hat{\nu}_j$ の積率相関である(これを、**連関係数**と呼ぶ)。

(2.10)、(2.11)、(2.14)、(2.15)式より

$$\sum_j \hat{\nu}_j n_{ij} / n_{i.} = \lambda' \hat{\mu}_i$$

$$\sum_i \hat{\mu}_i n_{ij} / n_{.j} = \lambda' \hat{\nu}_j$$

が成り立つ。(2.5)式より、(2.14)式(あるいは、(2.13)式)を満たすとき、 $\hat{\mu}_i, \hat{\nu}_j$ は x_i, y_j に一致する。(2.14)

式は

$$\sum_j \hat{\nu}_j m_{ij} = \lambda' \hat{\mu}_i \hat{m}_{i.} \quad (2.16)$$

と書け、

$$\hat{m}_{ij} = \hat{a}_i \hat{\beta}_j e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}}$$

より、

$$\hat{m}_{i.} = \hat{a}_i \sum_j \hat{\beta}_j e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}}$$

であるから、(2.16)式を用いると

$$\sum_j \hat{\nu}_j \hat{a}_i \hat{\beta}_j e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}} = \lambda' \hat{\mu}_i \hat{a}_i \sum_j \hat{\beta}_j e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}}$$

すなわち

$$\sum_j \hat{\nu}_j \hat{\beta}_j e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}} = \lambda' \hat{\mu}_i \sum_j \hat{\beta}_j e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}} \quad (2.17)$$

を得る。同様に

$$\sum_i \hat{\mu}_i \hat{a}_i e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}} = \lambda' \hat{\nu}_j \sum_i \hat{a}_i e^{\hat{\phi}_{\hat{\mu}_i \hat{\nu}_j}} \quad (2.18)$$

が成り立つ。

以上より、 $\hat{\beta}_j, \hat{a}_i$ が(2.17)式と(2.18)式を満たしたとき、 $\hat{\mu}_i$ と $\hat{\nu}_j$ は正準スコアになる。

2.3 数値例

表1のデータは、Clogg (1982)が取り上げた性別にみた幸福度("あまり幸福でない", "やや幸福", "非常に幸福"), および受けた学校教育の年数("12未満", "12", "13-16", "17以上")間の関係についてのクロス分類表である。これらのデータに対し、正準相関分析による正準スコア、およびRCモデルのもとでのスコアを求めると表2のようになる。(ちなみに、RCモデルに対する尤度比カイ二乗値は、それぞれ5.2942 ($df=2$), および0.6127 ($df=2$)である。)これらのスコアを用いれば、カテゴリーの順序を一意に決められ

表1a 男性に対する幸福度、および受けた学校教育の年数によるクロス分類表

幸福度	受けた学校教育の年数			
	<12	12	13-16	17+
あまり幸福ではない	40	21	14	3
やや幸福	131	116	112	27
非常に幸福	82	61	55	27

表1b 女性に対する幸福度、および受けた学校の年数によるクロス分類表 (カッコ内は ϕ 値)

幸 福 度	受けた学校教育の年数			
	<12	12	13-16	17+
あまり幸福ではない	62 (3.024)	26 (2.246)	12 (1.437)	3
やや幸福	155 (1.771)	156 (1.357)	95 (1.454)	15
非常に幸福	87	127	76	15

表2. 表1のデータに対する正準スコア、およびRCモデルのもとでのスコア推定値

	表 1 a		表 1 b	
	正準スコア	RC モデル 推 定 値	正準スコア	RC モデル 推 定 値
連関係数	0.118	0.115	0.201	0.201
行				
1	2.798	-2.704	2.442	-2.409
2	-0.344	0.136	0.011	-0.041
3	-0.380	0.704	-0.839	0.870
列				
1	1.197	-0.962	1.309	-1.304
2	-0.185	-0.061	-0.667	0.630
3	-0.949	0.586	-0.909	0.960
4	-1.659	2.622	-0.769	0.789

る。また、順序が事前に決まっているなら、その順序と推定されスコアに基づく順序との比較ができる。

さて、表2から明らかなように、表1aに対する i 行の正準スコアは、 i と共に減少し、 j 列のそれも($j=4$ の傾向の変化を除けば) j と共に減少している。表1bについては、 i 行は i と共に減少、 j 列も j と共に減少する。ゆえに、これらのデータに関し、行変数と列変数は、正の連関がある。特に、表1bに対する正準スコアとRCモデルのスコアは大きさが良く似た例である。参考のため、表1a, 1bのデータについて、順序を考慮した連関係数として、Goodman & Kruskal (1954, p. 749) の γ , Kendall (1975, p. 45, p. 49) の τ_a , τ_b , τ_c , および Somers (1962) の d_{AB} , d_{BA} を算出すると表3のようになり、いずれも正の連関を示している。

3. 連関モデルと2変量正規

3.1 2変量分布族との比較

Goodman (1981 b) によれば、行と列の分類が潜在

表3. 種々の連関係数の値

連関係数	表 1 a	表 1 b
γ	0.106	0.233
τ_a	0.043	0.093
τ_b	0.067	0.148
τ_c	0.064	0.140
b_{AB}	0.060	0.138
b_{BA}	0.075	0.158

的に1変量正規分布から生じたとき、連関モデルは Plackett (1965) モデル (Pearson (1913) は、これを一定の Yule 連関をもつモデルと呼んだ) より2変量正規に良く一致している^{付録}。

いま、(2.3)式のスコア μ_i, ν_j の分布を、2変量正規分布と比較しよう。(2.4)式の制約を受けることにより、相関係数 ρ の2変量正規分布

表4. 2変量正規分布 ($\rho=0.5$), Plackett モデル, 一様連関モデル, RC モデル, および独立モデル, 確率=成分値/ 10^4

y	x	0.0	0.5	1.0	1.5	2.0	2.5
2.5	6	11	14	13	9	7	
	11	12	14	7	5	1	
	6.5	11.4	14.5	13.3	8.8	5.5	
	6.0	10.8	14.0	13.3	9.2	6.8	
	23.0	18.0	11.0	5.3	2.0	0.8	
2.0	23	35	37	29	16	9	
	27	36	33	23	8	5	
	23.1	34.5	37.2	29.1	16.4	8.7	
	23.6	34.6	36.8	28.7	16.2	9.2	
	57.0	44.7	27.3	13.1	4.9	1.9	
1.5	79	99	90	60	29	13	
	76	97	88	52	24	6	
	77.7	98.9	90.8	60.4	29.1	13.1	
	79.1	99.1	90.1	59.8	28.7	13.3	
	141.6	111.0	67.9	32.6	12.3	4.7	
1.0	191	205	160	90	37	14	
	180	210	165	88	32	15	
	189.3	205.2	160.2	90.7	37.2	14.3	
	191.1	205.1	159.5	90.3	37.0	14.0	
	266.7	209.0	127.9	61.5	23.1	8.8	
0.5	336	309	205	99	35	11	
	340	331	210	97	36	12	
	335.2	309.2	205.6	99.1	34.4	11.3	
	335.6	308.8	205.6	99.4	34.8	10.7	
	380.7	298.4	182.6	87.7	33.0	12.5	
0.0	426	336	191	79	23	6	
	458	340	180	77	26	11	
	428.8	336.7	190.6	78.2	23.2	6.5	
	428.6	336.7	191.0	78.5	23.4	5.8	
	407.1	319.1	195.3	93.8	35.3	13.4	
-0.5	396	266	129	45	12	2	
	397	239	115	49	17	6	
	397.5	265.7	128.1	44.7	11.3	2.7	
	396.0	266.1	128.9	45.2	11.4	2.3	
	325.3	254.9	156.0	74.9	28.2	10.7	
-1.0	266	152	63	19	4	1	
	238	131	64	26	10	3	
	267.2	152.1	62.4	18.5	4.0	0.8	
	265.2	152.8	63.4	19.0	4.1	0.7	
	193.2	151.4	92.7	44.5	16.8	6.4	
-1.5	129	63	22	6	1	0	
	116	63	30	14	4	2	
	129.6	62.8	21.9	5.5	1.0	0.2	
	129.0	63.1	22.2	5.6	1.0	0.1	
	84.6	66.3	40.6	19.5	7.3	2.8	
-2.0	45	19	6	1	0	0	
	49	27	13	5	2	1	
	45.3	18.7	5.5	1.2	0.2	0.0	
	45.5	18.7	5.5	1.2	0.2	0.0	
	27.2	21.3	13.0	6.3	2.4	0.9	
-2.5	1212	4	1	0	0	0	
	17	9	5	2	1	0	
	11.7	4.1	1.0	0.2	0.0	0.0	
	12.2	3.8	0.9	0.1	0.0	0.0	
	6.5	5.1	3.1	1.5	0.6	0.2	
	2	1	0	0	0	0	
	6	3	2	0	1	0	
	2.2	0.7	0.1	0.0	0.0	0.0	
	2.2	0.6	0.1	0.0	0.0	0.0	
	1.1	0.9	0.6	0.3	0.0	0.0	

表5. 2変量正規分布 ($\rho=0.5$) をセル度数とした場合の種々の連関モデルのもとでの Pearson 適合度カイニ乗値, 自由度および AIC 値

連関モデル	Pearson 適合度 カイニ乗値	自由度	AIC
一様連関モデル	2.0	54	359.6
RC モデル	1.3	40	386.7
独立モデル	1144.5	55	884.6

$$f(X=x, Y=y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \cdot (x^2 - 2\rho xy + y^2) \right\} \\ = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{x^2+y^2}{2(1-\rho^2)} \right\} \\ \times \exp \left(\frac{\rho}{1-\rho^2} xy \right) \quad (3.1)$$

を, $\phi = \tau/(1-\tau^2)$ とした

$$m_{ij} = \alpha_i \beta_j \exp \left(\frac{\tau}{1-\tau^2} \mu_i \nu_j \right)$$

に対応できる.

表4の第1成分は, (3.1)式で $\rho=0.5$ の2変量正規分布, 第2成分は Plackett モデルの値である. 第3成分の一様連関モデルに対する成分は, 第1成分の値をセル度数として Goodman の反復法により算出した期待度数である. 同様に, RC モデルおよび独立モデルのもとでの期待度数を, それぞれ第4および第5成分に与えておく. 一様連関モデル, RC モデル, 独立モデルに対する Pearson 適合度カイニ乗値, 自由度, および AIC ($= -2 \ln(\text{最大尤度}) + 2(\text{自由なパラメータ数})$) 値を表5に示す. 表4において, 連続した行と列カテゴリーは, 等間隔に取り扱われているので, モデルの当てはまりは (AIC に準拠すれば) RC モデルより一様連関モデルのほうが良い.

ちなみに, RC モデルのもとでのスコアの点推定値, およびスコア間隔は表6のようになる.

3.2 分割係数に基づく推測

Wahrendorf (1980) は, 2変量分布に対する Plackett の分割係数 ϕ (付録を参照) を用い, 順序カテゴリーを有する $I \times J$ 分割表に関する推測を考案している. 分割表の周辺分布 $F(x)$ と $G(y)$ が与えられたとき,

表6. 2変量正規分布 ($\rho=0.5$) をセル度数とした場合の RC モデルのもとでのスコア推定値, およびスコア間隔

	点推定値	スコア間隔
行		
1	2.644	
2	1.975	0.669
3	1.440	0.535
4	0.910	0.530
5	0.384	0.527
6	-0.158	0.541
7	-0.693	0.535
8	-1.220	0.527
9	-1.782	0.562
10	-2.377	0.595
11	-3.309	0.932
12	-3.491	0.182
列		
1	-0.940	0.825
2	-0.115	0.834
3	0.719	0.846
4	1.564	0.847
5	2.412	1.089
6	3.501	

$$\phi = \frac{H(1-F-G+H)}{(F-H)(G-H)}, \phi > 0$$

に基づいて, 2変量分布族 $H(x, y)$ を一意に定義できる. $\phi=1$ (すなわち, $\log \phi=0$) なら, 周辺確率変数が統計的に独立であることを意味する.

カテゴリーに順序を有する分割表について, 隣接したセルの合併度数 $N(\alpha, \beta; \gamma, \delta) = \sum_{i=\alpha}^{\beta} \sum_{j=\gamma}^{\delta} n_{ij}$ ($1 \leq \alpha \leq \beta \leq I; 1 \leq \gamma \leq \delta \leq J$) から構成される $(I-1)(J-1)$ 個の分割係数は

$$\hat{\phi}_{kl} = \frac{N(1, k; 1, l)N(k+1, I; l+1, J)}{N(1, k; l+1, J)N(k+1, I; 1, l)}, \\ k=1, \dots, I-1; l=1, \dots, J-1 \quad (3.2)$$

で与えられる. この対数値

$$\hat{d}_{kl} = \log N(1, k; 1, l) + \log N(k+1, I; l+1, J) \\ - \log N(1, k; l+1, J) \\ - \log N(k+1, I; 1, l) \quad (3.3)$$

の近似的分散は

$$\begin{aligned}\text{Var}(\hat{d}_{kl}) = & \frac{1}{N(1, k; 1, l)} \\ & + \frac{1}{N(k+1, I; l+1, J)} \\ & + \frac{1}{N(1, k; l+1, J)} \\ & + \frac{1}{N(k+1, I; 1, l)}\end{aligned}\quad (3.4)$$

となる。また近似的共分散は、 $1 \leq i, k \leq I-1$ および $1 \leq j, l \leq J-1$ について

$$\begin{aligned}\text{Cov}(\hat{d}_{ij}, \hat{d}_{kl}) = & \text{Cov}\{\log N(1, i; 1, j) \\ & + \log N(i+1, I; j+1, J) \\ & - \log N(1, i; j+1, J) \\ & - \log N(i+1, I; 1, j), \log N(1, k; 1, l) \\ & + \log N(k+1, I; l+1, J) - \log N(1, k; l \\ & + 1, J) - \log N(k+1, I; 1, J)\}\end{aligned}\quad (3.5)$$

と書ける。上式の右辺は、展開して各項に

$$\begin{aligned}\text{Cov}\{\log N(\alpha, \beta; \gamma, \delta), \log N(\alpha', \beta'; \gamma', \delta')\} \\ = \frac{N(\alpha \vee \alpha', \beta \wedge \beta'; \gamma \vee \gamma', \delta \wedge \delta')}{N(\alpha, \beta; \gamma, \delta)N(\alpha', \beta'; \gamma', \delta')} \\ - \frac{1}{N(1, I; 1, J)}\end{aligned}$$

を適用すれば求められる。ここに、 $x \wedge y = \max\{x, y\}$, $x \vee y = \max\{x, y\}$ とする。 $t = (I-1)(J-1)$ について、(3.4) と (3.5) 式から推定される分散共分散 ($t \times t$) 行列を C , (3.3) 式を要素とする $t \times 1$ ベクトルを $\hat{d}$ と書く。このとき、(3.2) 式（あるいは (3.3) 式）の分割係数が同等であるという帰無仮説のもとで

$$z = \hat{d}^T T^T (TCT^T)^{-1} T\hat{d} \quad (3.6)$$

は、漸近的に自由度 $(t-1)$ のカイ二乗分布に従う。ここに、 $(t-1) \times t$ 行列 T を

$$T = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

とする。

この検定の結果、有意になれば結合された ϕ の推定値（分散が最小になる重み付け平均） $\hat{d}_w$ は、重みを $w = C^{-1}(\mathbf{1}^T C^{-1})^{-1}$, 成分が 1 の $t \times 1$ ベクトルを $\mathbf{1}$ としたとき

$$\hat{d}_w = w^T \hat{d}$$

で、その分散は

$$\text{Var}(\hat{d}_w) = w^T C w$$

で与えられる。

一例として、表 1a および表 1b の 3×4 分割表について、(3.6) 式の z を算出すると、それぞれ 11.700, および 10.376 となり、表 1b の分割係数のみが有意水準 5% で同等であると見なされる。かくて、 $\hat{d}_w = 0.583$, $\text{Var}(\hat{d}_w) = 0.0153$ となる。また、 ϕ 値を表 1b の分割点（カッコ内の値）に与えておく。

4. 多次元表

多次元表における連関モデルについて論述する。本稿では、3次元表のみを詳しく考察するが、容易に m 次元表 ($m=4, 5, \dots$) へも拡張可能である (Clogg (1982), Agresti & Kezouh (1983), Agresti (1984))。

4.1 見込み比

3次元分割表において、セル (i, j, k) の観測度数を n_{ijk} , あるモデルのもとでの期待度数を m_{ijk} ($i=1, \dots, I; j=1, \dots, J; k=1, \dots, K$), そして行, 列, 層変数を A, B, C とする。見込み比

$$\theta_{ij(k)} = \frac{m_{ijk} m_{i+1, j+1, k}}{m_{i+1, j, k} m_{i, j+1, k}}; \quad (4.1)$$

$$i=1, \dots, I-1; j=1, \dots, J-1; k=1, \dots, K$$

は、 C の固定された水準内での A と B との局所偏連関を表す。同様に、 $\theta_{i(j)k}$ および $\theta_{(i)jk}$ も定義できる。見込み比の比

$$\theta_{ijk} = \frac{\theta_{ij(k+1)}}{\theta_{ij(k)}} = \frac{\theta_{i(j+1)k}}{\theta_{i(j)k}} = \frac{\theta_{(i+1)jk}}{\theta_{(i)jk}}$$

は、局所 3 因子交互作用である。 $(I-1)(J-1)(K-1)$ 個すべての θ_{ijk} が 1 に等しければ、3 因子交互作用が存在しない。

4.2 線形効果連関モデル

変数 A, B, C がすべて順序を有すると仮定する。相互独立モデル

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C \quad (4.2)$$

の一般化で、2次元表の線形 $\times$ 線形連関モデル (Haberman (1974)) に対応したモデルを考える。変数 A, B, C のそれぞれの水準に、カテゴリー・スコア $\{u_i\}, \{v_j\}, \{w_k\}$ を割付けたとき、等質線形効果連関モデル (homogeneous linear effects association model)

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \beta^{AB} u_i v_j + \beta^{AC} u_i w_k + \beta^{BC} v_j w_k \quad (4.3)$$

が得られる。ただし、 $\Sigma \lambda_i^A = \Sigma \lambda_j^B = \Sigma \lambda_k^C = 0$ とする。自由度は $df = IJK - I - J - K - 1$ で、

$$\left. \begin{aligned} \ln \theta_{ij(k)} &= (u_{i+1} - u_i)(v_{j+1} - v_j) \beta^{AB} \\ \ln \theta_{i(j)k} &= (u_{i+1} - u_i)(w_{k+1} - w_k) \beta^{AC} \\ \ln \theta_{(i)jk} &= (v_{j+1} - v_j)(w_{k+1} - w_k) \beta^{BC} \\ \ln \theta_{ijk} &= 0 \end{aligned} \right\} \quad (4.4)$$

となる。

特に、等間隔スコア (例えば、 $\{u_i\} = \{i\}$, $\{v_j\} = \{j\}$, $\{w_k\} = \{k\}$) を与えた場合、(4.4) の局所連関は一定である。このモデルを、等質一様連関モデル (homogeneous uniform association model) と名付ける。

また、3 因子交互作用を含めば

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \beta^{AB} u_i v_j + \beta^{AC} u_i w_k + \beta^{BC} v_j w_k + \beta^{ABC} u_i v_j w_k$$

となる。自由度は $df = IJK - I - J - K - 2$ である。

4.3 等質加法効果連関モデル

2 次元表におけるモデル I (Goodman (1979)) は

$$\ln m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \beta^{AB} u_i v_j + \tau_i^B u_i + \tau_i^A v_j$$

と一般化できる。ここに、 $\Sigma \lambda_i^A = \Sigma \lambda_j^B = 0$, $\tau_i^A = \tau_i^B = \tau_i^B = \tau_i^A = 0$ とする。このモデルを 3 次元表へ拡張すれば

$$\begin{aligned} \ln m_{ijk} &= \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \beta^{AB} u_i v_j + \beta^{AC} u_i w_k \\ &\quad + \beta^{BC} v_j w_k + u_i \tau_{1j}^B + u_i \tau_{1k}^C + v_j \tau_{1i}^A \\ &\quad + v_j \tau_{2k}^C + w_k \tau_{2i}^A + w_k \tau_{2j}^B \end{aligned} \quad (4.5)$$

となる。ただし、 $\Sigma \lambda_i^A = \Sigma \lambda_j^B = \Sigma \lambda_k^C = 0$ かつ $\tau_{11}^A = \tau_{11}^B = \tau_{11}^C = \tau_{21}^A = \tau_{21}^B = \tau_{21}^C = \tau_{21}^A = \tau_{21}^B = \tau_{21}^C = \tau_{21}^A = \tau_{21}^B = \tau_{21}^C = 0$ である。自由度は、 $df = IJK - 3(I + J + K) + 11$ となり、

$$\begin{aligned} \ln \theta_{ij(k)} &= (u_{i+1} - u_i)(v_{j+1} - v_j) \beta^{AB} + (v_{j+1} - v_j) \\ &\quad (\tau_{1,i+1}^A - \tau_{1,i}^A) + (u_{i+1} - u_i)(\tau_{1,j+1}^B - \tau_{1,j}^B) \end{aligned} \quad (4.6)$$

が成り立つ。 $\ln \theta_{i(j)k}$, $\ln \theta_{(i)jk}$ についても同様に定義できる。(4.6) 式から、 $\{\tau_{1i}^A\}$ と $\{\tau_{1j}^B\}$ がそれぞれ $A-B$ 偏連関上の加法行効果、および加法列効果を表し、それらの効果は C の水準に関し等質であることが分かる。 $\ln \theta_{i(j)k}$ や $\ln \theta_{(i)jk}$ についても類似のことがいえる。ゆえに、(4.5) 式を等質加法効果連関モデル (homogeneous additive effects association model) と呼ぶ。特に、(4.6) 式の τ パラメータがすべてゼロなら (4.6) 式は (4.3) 式になる。

変数の一つが名義的 (例えば、 A と B が順序変数、 C が名義変数) なら、

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \beta^{AB} u_i v_j + u_i \tau_{1k}^C + v_j \tau_{2k}^C$$

とすればよい。ただし、 $\Sigma \lambda_i^A = \Sigma \lambda_j^B = \Sigma \lambda_k^C = \Sigma \tau_{1k}^C = \Sigma \tau_{2k}^C = 0$ である。この自由度は、 $df = IJK - I - J - 3K + 3$ となる。

また、変数の二つが名義的 (例えば、 A と B が名義変数、 C が順序変数) なら

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + w_k \tau_{2i}^A + w_k \tau_{2j}^B \quad (4.7)$$

とする。ただし

$$\Sigma \lambda_i^A = \Sigma \lambda_j^B = \Sigma \lambda_k^C = \Sigma \tau_{2i}^A = \Sigma \tau_{2j}^B = \Sigma \lambda_{ij}^{AB} = \Sigma \lambda_{ij}^{AB} = 0$$

である。自由度は、 $df = IJK - IJ - I - J - K + 3$ となる。

4.4 乗法効果連関モデル

等質加法効果連関モデル (4.3) 式において、スコア u_i , v_j , w_k が割付けられたものでなく、未知パラメータなら

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \beta^{AB} \mu_i \nu_j + \beta^{AC} \mu_i \omega_k + \beta^{BC} \nu_j \omega_k \quad (4.8)$$

となる。ただし

$$\begin{aligned} \Sigma \lambda_i^A &= \Sigma \lambda_j^B = \Sigma \lambda_k^C = \Sigma \mu_i p_{i..} = \Sigma \nu_j p_{.j.} = \Sigma \omega_k p_{..k} = 0; \\ \Sigma \mu_i^2 p_{i..} &= \Sigma \nu_j^2 p_{.j.} = \Sigma \omega_k^2 p_{..k} = 1; \quad p_{i..} = \sum_j \sum_k n_{ijk} / n, \\ p_{.j.} &= \sum_i \sum_k n_{ijk} / n, \quad p_{..k} = \sum_i \sum_j n_{ijk} / n, \\ n &= \Sigma \Sigma \Sigma n_{ijk} \end{aligned}$$

とする。自由度は、 $df = IJK - 2(I + J + K) + 5$ で

$$\begin{aligned} \ln \theta_{ij(k)} &= (\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j) \beta^{AB} \\ \ln \theta_{i(j)k} &= (\mu_{i+1} - \mu_i)(\omega_{k+1} - \omega_k) \beta^{AC} \\ \ln \theta_{(i)jk} &= (\nu_{j+1} - \nu_j)(\omega_{k+1} - \omega_k) \beta^{BC} \\ \ln \theta_{ijk} &= 0 \end{aligned}$$

が成り立つ。これは、Goodman (1979) のモデル II (RC モデル) の 3 次元表への拡張である。

更に、各変数の効果が、連関ごとに異なっているなら

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \beta^{AB} \mu_{1i} \nu_{1j} + \beta^{AC} \mu_{2i} \omega_{1k} + \beta^{BC} \nu_{2j} \omega_{2k} \quad (4.9)$$

とすればよい。自由度は、 $df = IJK - 3(I + J + K) + 11$

である。

4.5 条件付き連関モデル

本節では、順序を有する2次元表が、各 K 個のグループごとに観測された場合を考える (Clogg (1982))。

Goodman のモデル I, および II の3次元表への拡張として、それぞれ

$$m_{ijk} = \tau_{13(i,k)} \tau_{23(j,k)} \alpha_i^j \beta_j^i$$

および

$$m_{ijk} = \tau_{13(i,k)} \tau_{23(j,k)} \exp \{ \mu_{i(k)} \nu_{j(k)} \}$$

を定義する。これらを (4.1) 式へ代入すると

$$\theta_{ij(k)} = (\alpha_{i+1,(k)} / \alpha_{i,(k)}) (\beta_{j+1,(k)} / \beta_{j,(k)})$$

および

$$\ln \theta_{ij(k)} = (\mu_{i+1,(k)} - \mu_{i,(k)}) (\nu_{j+1,(k)} - \nu_{j,(k)})$$

が得られる。すなわち

$$\theta_{ij(k)} = \theta_{i,(k)} \theta_{j,(k)} \quad (4.10)$$

および

$$\ln \theta_{ij(k)} = \phi_{i,(k)} \phi_{j,(k)} \quad (4.11)$$

となる。これらのモデルは、グループごとに異なった行効果および列効果をもち、異質行および列効果条件付き連関モデル (heterogeneous row and column effects conditional association model) と呼ばれる。Goodman (1979) の2次元表の場合と同様に、(4.10) 式の特殊な場合が考えられる。

まず、

$$\theta_{ij(k)} = 1$$

なら、条件付き無連関モデル (null conditional association model) で、これは条件付き独立モデル

$$\ln m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

と等価になる。自由度は、 $df = K(I-1)(J-1)$ である。

$$\theta_{ij(k)} = \theta$$

のとき、グループごとに等質な一様連関が存在する。自由度は、 $df = K(I-1)(J-1) - 1$ となる。これを、等質一様連関モデル (homogeneous uniform association model) と名付ける。また、この一様連関がグループごとに異なるなら

$$\theta_{ij(k)} = \theta_{i,(k)}$$

とする。自由度は、 $df = K(IJ - I - J)$ である。

行効果のみがあり、それらが K 個のグループにおいて同一なら

$$\theta_{ij(k)} = \theta_{i..}$$

とする。自由度は、 $df = (I-1)(KJ - K - 1)$ である。これを、等質行効果連関モデル (homogeneous row effect association model) と呼ぶ。一方、行効果がグループごとに異なれば

$$\theta_{ij(k)} = \theta_{i \cdot (k)}$$

とすればよい。自由度は、 $df = K(I-1)(J-1) - IK + K$ となる。これを、異質行効果連関モデル (heterogeneous row effect association model) と名付ける。

また、列効果が存在し、それがグループごとに同じなら

$$\theta_{ij(k)} = \theta_{j..}$$

とする。自由度は、 $df = (J-1)(IK - K - 1)$ である。これを、等質列効果連関モデル (homogeneous column effect association model) という。その列効果が、グループごとに異なれば

$$\theta_{ij(k)} = \theta_{j \cdot (k)}$$

とする。自由度は、 $df = K(I-1)(J-1) - JK + K$ である。これを、異質列効果連関モデル (heterogeneous column effect association model) と名付ける。

行および列効果は存在するが、それらがグループごとに同じなら

$$\theta_{ij(k)} = \theta_{i..} \theta_{j..}$$

とすればよい。自由度は、 $df = K(I-1)(J-1) - I - J + 3$ となる。これを、等質行および列効果連関モデル (homogeneous row and column effects association model) と呼ぶ。特に、グループごとに列効果は等しいが、行効果が異なるなら

$$\theta_{ij(k)} = \theta_{i \cdot (k)} \theta_{j..}$$

とする。自由度は、 $df = K(I-1)(J-1) - IK + K - J + 2$ である。これを、異質行効果および等質列効果連関モデル (heterogeneous row and homogeneous column effects association model) という。逆に、グループごとの行効果は等しいが、列効果が異なるなら

$$\theta_{ij(k)} = \theta_{i..} \theta_{j \cdot (k)}$$

表7. 幸福度, 受けた学校教育の年数, および兄弟(姉妹)の人数によるクロス分類表

受けた学校教育の年数	兄弟(姉妹)の人数				
	0-1	2-3	4-5	6-7	8+
幸福でない					
<12	15	34	36	22	61
12	31	60	46	25	26
13-16	35	45	30	13	8
17+	18	14	3	3	4
少し幸福					
<12	17	53	70	67	79
12	60	96	45	40	31
13-16	63	74	39	24	7
17+	15	15	9	2	1
非常に幸福					
<12	7	20	23	16	36
12	5	12	11	12	7
13-16	5	10	4	4	3
17+	2	1	2	0	1

表8. 表7のデータに対する乗法効果連関モデル(4.8)のもとでのスコア推定値

点推定値	
行	
1	1.218
2	-0.278
3	-1.081
4	-1.589
例	
1	-1.443
2	-0.571
3	0.228
4	0.597
5	1.640
層	
1	-0.563
2	-0.233
3	2.685

とする。自由度は, $df = K(I-1)(J-1) - JK + K - I + 2$ である。これを, 等質行効果および異質列効果連関モデル (homogeneous row effect and heterogeneous column effect association model) と名付ける。

4.6 数値例

表7の3次元分割表データは, 受けた学校教育の年数(A), 兄弟(姉妹)の人数(B), および幸福感(C)によってクロス分類したものである(Clogg (1982))。このデータへ乗法効果連関モデル(4.8)式を当てはめたところ, 尤度比カイ二乗値は $G^2 = 46.15 (df = 41)$ となり, 有意水準5%で, このモデルの適合度は良いと見なせる。また, カテゴリー・スコアは, 表8のように推定される。

次に, カテゴリー間の距離(これを, $|\mu_i - \mu_{i+1}|, |\nu_j - \nu_{j+1}|, |\omega_k - \omega_{k+1}|$ で定義する)について考察を加えよう。受けた学校教育の年数(A)に関し, 最初の二つのカテゴリー間のスコア間隔の推定値は1.496で, これは第3と第4カテゴリー間のそれ(0.508)の約3倍である。兄弟の人数(B)について, 第1と第2カテゴリー間のスコア間隔の推定値(0.872)は, 第2と第3カテゴリー間のそれ(0.799)にほぼ等しい。また, 幸福感(C)において, 第1と第2カテゴリー間のスコア間隔の推定値(-0.330)は, 第2と第3カテゴリー間の

のそれ(-2.918)の約1/10である。

付 録

Plackett モデル (2 変量分布族) について

Plackett (1965)は, 周辺計が与えられた1パラメータから成る2変量分布族の構成法を考えた。任意の点 (x, y) で2分法化された2変量分布 $H(x, y)$ が, 周辺分布 $F(x)$, $G(y)$ をもつとき, 四分表 ($p_1 = n_{11}/n$, $p_2 = n_{12}/n$, $p_3 = n_{21}/n$, $p_4 = n_{22}/n$) に関し

$$\begin{aligned} p_1 &= H(x, y), \quad p_2 = F(x) - H(x, y), \quad p_3 \\ &= G(y) - H(x, y), \\ p_4 &= 1 - F(x) - G(y) + H(x, y) \end{aligned}$$

とおけば, 分割係数(見込み比)

$$\phi = \frac{p_1 p_4}{p_2 p_3}$$

に対応して

$$\phi = \frac{H(1-F-G+H)}{(F-H)(G-H)}$$

となる(Mardia (1970 a, b))。ゆえに, Mardia (1967)によると, 1パラメータの2変量分布 $H(x, y)$ は

$$H = \frac{S - \{S^2 - 4\phi(\phi-1)FG\}^{1/2}}{2(\phi-1)}, \quad \phi > 0$$

で与えられる。ここに

$$S=1+(F+G)(\phi-1)$$

で、 H は不等式

$$\max\{F+G-1, 0\} \leq H \leq \min\{F, G\}$$

を満たす。これは、分割表型分布 (contingency-type distribution) と呼ばれる。表4の2段目の要素は、周辺分布 $F(x)$, $G(y)$ が標準正規で $\phi=4$ をもつ2変量分布 $H(x, y)$ について、 (x, y) 平面を覆っている領域の集合に対する確率 ($\times 10^4$) を示している。

謝 辞

本稿および第1報をまとめるにあたり、審査員の方々から有益なご教示をいただきました。ここに記して謝意を表します。

参 考 文 献

- Agresti, A. (1984) *Analysis of Ordinal Categorical Data*, New York: John Wiley & Sons.
- Agresti, A. and Kezouh, A. (1983) Association models for multidimensional cross-classifications of ordinal variables, *Commun. Statist.*, **12**, 1261-1276.
- Clogg, C.C. (1982) Some models for the analysis of association in multiway cross-classifications having ordered categories, *J. Amer. Statist. Assoc.*, **77**, 803-815.
- Goodman, L.A. (1979) Simple models for the analysis of association in cross-classifications having ordered categories, *J. Amer. Statist. Assoc.*, **74**, 537-552.
- Goodman, L.A. (1981 a) Association models and canonical correlation in the analysis of cross-classifications having ordered categories, *J. Amer. Statist. Assoc.*, **76**, 320-334.
- Goodman, L.A. (1981 b) Association models and the bivariate normal for contingency tables with ordered categories, *Biometrika*, **68**, 347-355.
- Goodman, L.A. and Kruskal, E.H. (1954) Measures of association for cross-classifications, I, *J. Amer. Statist. Assoc.*, **49**, 732-764.
- Haberman, S.J. (1974) Loglinear models for frequency tables with ordered classifications, *Biometrics*, **30**, 589-600.
- Kendall, M.G. (1975) *Rank Correlation Methods*, London: Griffin.
- Kendall, M.G. and Stuart, A. (1979) *The Advanced Theory of Statistics; Vol. 2. Inference and Relationship*, London: Griffin.
- Mardia, K.V. (1967) Some contributions to contingency-type bivariate distributions, *Biometrika*, **54**, 235-249.
- Mardia, K.V. (1970 a) *Families of Bivariate Distributions*, London: Griffin.
- Mardia, K.V. (1970 b) Some problems of fitting for contingency-type bivariate distributions, *J. Roy. Statist. Soc.*, **B 32**, 254-264.
- 西里静彦(1982) 質的データの数量化—双対尺度法とその応用—, 朝倉書店.
- Pearson, K. (1913) Note on the surface of constant association, *Biometrika*, **9**, 534-537.
- Plackett, R.L. (1965) A class of bivariate distributions, *J. Amer. Statist. Assoc.*, **60**, 516-522.
- Somers, R.H. (1962) A new asymmetric measure of association for ordinal variables, *Amer. Soc. Rev.*, **27**, 799-811.
- 辻谷将明(1984) 順序カテゴリー・データ解析, 行動計量学, **11**, 22-38.
- Wahrendorf, J. (1980) Inference in contingency tables with ordered categories using Plackett's coefficient of association for bivariate distributions, *Biometrika*, **67**, 15-21.
- 柳井晴夫, 高根芳雄(1977) 多変量解析法, 朝倉書店.

—1984年12月受付, 1985年8月再受付—