

2×2 分割表に対するモデルの 図式表現と潜在構造分析

統計数理研究所 鈴木 義一郎

(1984 年 7 月 受付)

要 旨

この論文では、2×2 分割表の場合に限って、種々のタイプの特殊モデルと、それらの混合モデルを考えることによって、分割表の形で与えられるデータに対する潜在構造分析の諸相を概観する。さらに、2×2 分割表のモデルに対する幾何学的表現を導入して、個々の特殊モデルや、それらの混合モデルとして扱われる所与のデータに呼応するモデルの"位置づけ"などを、視覚的に把握することも可能であることを示す。また実際の分割表タイプのデータに対して、潜在構造分析の事例もいくつか紹介する。さらに、2組の分割表を併合すると association の尺度の符号が必ずしも保存されないという、いわゆる Simpson's paradox に対する考察が、分割表モデルに対応する幾何学的表現を用いて為される。

- § 1. はじめに (Goodman-Kruskal (1959) の抄訳による潜在構造分析の展開の概説)
- § 2. 2×2 分割表とその特殊モデル
- § 3. 2×2 分割表に対する図式表現
 - 3.1 周辺確率座標空間
 - 3.2 条件付確率座標空間
- § 4. 混合モデルのいろいろなパターン
 - 4.1 混合モデルと Simpson's paradox
 - 4.2 少なくとも 1 方の周辺確率が等しいモデルどうしの混合
 - 4.3 2組の周辺確率がいずれも異なるモデルどうしの混合
- § 5. 潜在構造分析
 - 5.1 潜在構造分析の諸相の概観
 - (i) 対角モデルと独立モデルの混合
 - (ii) 三角モデルと独立モデルの混合
 - (iii) その他の混合モデル
 - 5.2 潜在構造分析の実際
 - (i) 初婚の夫は再婚の妻との結婚を嫌う
 - (ii) 生れ変わったときの希望の性

§ 1. はじめに (Goodman-Kruskal (1959) の抄訳による潜在構造分析の展開の概説)

今から丁度 100 年前の 1884 年に、大竜巻の予想結果的中率の評価に関連して、3つの文献

が発表された。そのうちの2つは、*The American Meteorological Journal* の創刊号に掲載された J.P. Finley と G.K. Gilbert によるもので、もう1つは *Science* の4巻のエディターへの手紙の欄に記載された C.S. Peirce によるものである。

1884年4月、アメリカ合衆国の18の地区について発表された大竜巻の予想と実際の結果の一例を掲げてみると次のようになっている。

結果 予想	実際に大竜巻が 起きた(B)	大竜巻が 起こらなかった(B̄)	計
大竜巻が起きる (A)	11 (0)	14 (0)	25
大竜巻が起きない(Ā)	3 (14)	906 (920)	909
計	14	920	934

この予想に対する的中率として Finley が当初考えたものは、対角要素(網目の部分)の和を合計で割算した

$$\frac{11+906}{934}=98.18\%$$

という値であった。この種の考え方を適用する限り、例えば表のカッコ内の数値のように、“大竜巻は起こらない”とする(愚にもつかない)予想の的中率が

$$\frac{0+920}{934}=98.50\%$$

のように、先の予想の的中率を上まわってしまう結果となる。

このような2×2分割表は通例

(1)

	(B)	(B̄)	計
(A)	p_{11}	p_{12}	p
(Ā)	p_{21}	p_{22}	$1-p$
計	q	$1-q$	1

のような比率を用いて表現される。そこで

$$d = p_{11}p_{22} - p_{12}p_{21}$$

と置いてみると、この値が0になるのは

$$p_{11} = pq$$

のときに限ることが分かる。Gilbert は

$$\frac{d}{p_{12} + p_{21} + d}$$

によつて的中率が評価できると考えた。このような定義によれば、 d の値が間違つて予想を行うケースの確率 $p_{12} + p_{21}$ よりかなり大きくない限り、100%に近くなる。実際、先の表で与えられた予想の場合は

$$d = 0.0112, p_{12} + p_{21} = 0.0182$$

であるから、的中率は38.1%とかなり低い値にしかならない。

Finley にしろ、Gilbert にしろ、どうにもアドホックに過ぎるとして、Peirce は次のような考

え方を提唱した。ここに、完璧無比の予言者 (infallible predictor) と、全くデタラメな予言者 (completely ignorant predictor) とがいるとする。前者の予測結果は常に

q	0
0	$1-q$

といったパターンであるのに対し、後者は

$\bar{p}q$	$\bar{p}(1-q)$
$(1-\bar{p})q$	$(1-\bar{p})(1-q)$

のようなパターンで表わされる。もし完璧無比な予言者と全くデタラメな予言者とが $\alpha : (1-\alpha)$ という比率で混じり合って予想を行ったとすれば、全体としては

$\alpha q + (1-\alpha)\bar{p}q$	$(1-\alpha)\bar{p}(1-q)$
$(1-\alpha)(1-\bar{p})q$	$\alpha(1-q) + (1-\alpha)(1-\bar{p})(1-q)$

のような結果になる筈である。これと、実際に観測された分割表から得られる比率の表とをつき合わせて考えれば、

$$\alpha = \frac{d}{q(1-q)}$$

という関係が成立しなければならない。この α という値を、混合された表の "的中率" と解釈するのが妥当である、こう Peirce は考えたのである。またこの α は

$$\alpha = \frac{p_{11}}{q} - \frac{p_{12}}{1-q}$$

といった表現もできるから、大竜巻が実際に起こったケースでの的中率から、起こらなかったケースでの不的中率を差し引いた値になっており、起こらないのに起きると予想して周囲を騒がせた場合のマイナス面をも考慮に入れた尺度となっている。Finley の与えた当初のデータについてこの α の値を求めてみると、77.05% のように算出され、Gilbert の与えた数値の 2 倍強になっている。

その後 1901 年に、イタリアの人口統計学者 R. Bénini が、2×2 分割表に対する attraction and repulsion の尺度を提示した。1898 年の 1 年間に、イタリアで結婚した夫婦の結婚前の状態に関して、次のような 4 分表 (%表示) を得た。

夫 \ 妻	初 婚	再 婚	計
初婚	86.68	2.75	89.43
再婚	7.42	3.15	10.57
計	94.10	5.90	100.00

一般に (1) のような表に対して、彼の与えた尺度は

$$\frac{d}{\min(p, q) - pq}$$

のように表現される。分子は2×2分割表に対する通常の association の尺度、分母は周辺確率が固定されている場合の (d の) 最大可能値を表わしている。

初婚の夫が再婚の妻との結婚を忌避 (repulsion) しようとする行動には

p	0
$q-p$	$1-q$

というパターンが対応し、他方そのようなことに拘わらない行動には

pq	$p(1-q)$
$(1-p)q$	$(1-p)(1-q)$

という独立モデルを対応させる。実際のデータには、これら2つのパターンが $\alpha:(1-\alpha)$ という比率で混じり合っているものとして

$\alpha p + (1-\alpha)pq$	$(1-\alpha)p(1-q)$
$\alpha(q-p) + (1-\alpha)(1-p)q$	$\alpha(1-q) + (1-\alpha)(1-p)(1-q)$

という“混合”モデルを対応させれば

$$\alpha = \frac{d}{p(1-q)}$$

という関係の成立することが確かめられる。これは、Bénini が与えた repulsion の尺度そのものである。Bénini の与えたデータについてこの α の値を算出してみると、0.479 となる。初婚の夫が再婚の妻との結婚を忌避しようとする傾向が半分近くみられる。

また、Bénini の与えたデータを、次のような2つのパターンの混合として把える考え方も成立する：

86.68	0
10.17	3.15

86.68	5.085
5.085	3.15

前者はやはり、初婚の夫が再婚の妻との結婚を忌避しようとするパターンで、後者は初婚と再婚との組合せが男女同じ比率とする対称なケースである。この場合もデータとつき合わせて、 $\alpha=0.458$ という結果が得られ、先に求めた場合と類似の結論が得られる。

その後、半世紀近い歳月が流れて、P.F. Lazarsfeld, P. Kerdall, E. Katz 等の統計学者によって、この種の混合モデルを用いた2×2分割表の分析法の展開が再開されることになる。実際、“潜在構造 (latent structure)” という呼称は Lazarsfeld によるものである。

§ 2. 2×2 分割表とその特殊モデル

2次元モデルの、最も原始的なものといえば、ある属性をもつ (A) かもたない (\bar{A}) か、さらに他の属性をもつ (B) かもたない (\bar{B}) かによって、4分割して考えていく方法である。いわゆる、2×2 分割表とか、4分表とか呼ばれている。

一般に、観測データは

	(B)	(\bar{B})	計
(A)	n_{11}	n_{12}	$n_{1.}$
(\bar{A})	n_{21}	n_{22}	$n_{2.}$
計	$n_{.1}$	$n_{.2}$	n

といった形で与えられる。これに対応する "確率モデル" としては

	(B)	(\bar{B})	計
(A)	p_{11}	p_{12}	$p_{1.}$
(\bar{A})	p_{21}	p_{22}	$p_{2.}$
計	$p_{.1}$	$p_{.2}$	1

を考えてやる。

いま

$$p_{1.} = p, p_{.1} = q, p_{11}p_{22} - p_{12}p_{21} = d$$

のように置いてみると、上記のモデルは

$pq + d$	$p(1-q) - d$
$(1-p)q - d$	$(1-p)(1-q) + d$

のように表わされることが分かる。以下このモデルのことを、 $M\{p, q; d\}$ という記号で表わすことにする。

明らかに、 $M\{p, q; 0\}$ は独立なモデルに対応している。従って、 d というパラメータは、独立性からの偏位を表わしている。また d の表現式から、cross-product difference と呼ばれることもある。

更に、 d の変動し得る範囲は

$$-b \leq d \leq a$$

である。ここで

$$a = a(p, q) = \min \{p(1-q), (1-p)q\},$$

$$b = b(p, q) = \min \{pq, (1-p)(1-q)\},$$

そこで、 $M\{p, q; a\}$, $M\{p, q; -b\}$ といったモデルのことを、the most positively dependent model, the most negatively dependent model と呼ぶことにしよう。

後の目的のために、いくつかの特殊モデルを定義しておくことにしよう。まずパラメータが1つだけのモデルとしては

$$\text{独立 (independent) モデル: } I_0(p) = M\{p, p; 0\},$$

$$I_0^*(p) = M\{p, 1-p; 0\}$$

対角 (diagonal)モデル: $D^+(p) = M\{p, p; p(1-p)\}$,
 上三角 (upper triangle)モデル: $U_0^+(p) = M\{p, 1-p; p^2\}, \left(0 \leq p \leq \frac{1}{2}\right)$
 下三角 (under triangle)モデル: $L_0^+(p) = M\{p, 1-p; (1-p)^2\}, \left(\frac{1}{2} \leq p \leq 1\right)$

のようなものが考えられる。なおここでは、主として正の association をもつモデルだけを考
 えることにするが、負の対角モデル $D^-(p)$ や負の三角モデル $U_0^-(p), L_0^-(p)$ のようなものも
 同様に定義できる。

次に、パラメータが2つのモデルとしては

独立モデル : $I(p, q) = M\{p, q; 0\}$
 上三角モデル: $U^+(p, q) = M\{p, q; p(1-q)\}, (p \leq q)$
 下三角モデル: $L^+(p, q) = M\{p, q; (1-p)q\}, (p \geq q)$
 正対称 (Symmetric) モデル: $S(p, d) = M\{p, p; d\}, (0 \leq d \leq p(1-p))$
 歪対称 (skew-symmetric) モデル: $S^*(p, d) = M\{p, 1-p; d\},$
 $(0 \leq d \leq \min\{p^2, (1-p)^2\})$

といったようなものが考えられる。

さらに、これらの特殊モデルの間には

$$\begin{aligned} I(p, p) &= S(p, 0) = I_0(p) \\ I(p, 1-p) &= S^*(p, 0) = I_0^*(p) \\ U^+(p, p) &= L^+(p, p) = S(p, p(1-p)) = D^+(p) \\ U^+(p, 1-p) &= S^*(p, p^2) = U_0^+(p) \\ L^+(p, 1-p) &= S^*(p, (1-p)^2) = L_0^+(p) \end{aligned}$$

といった関係が成り立っていることも確かめられる。また、これらの対角モデルや三角モデル
 は、いずれも the most (positively) dependent model になっていることも分かる。

§3. 2×2 分割表に対する図式表示

データ解析を行う際に、種々の図式表現法を用いることにより、視覚を援用して分析を進め
 ることの有効性は、よく知られているところである。ここでは、2×2 分割表モデル $M\{p, q; d\}$

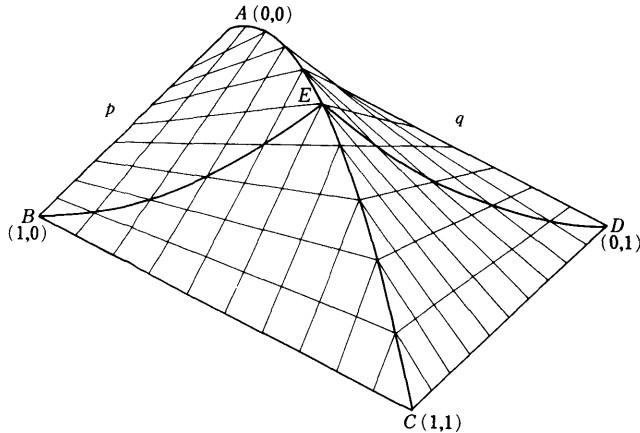


図1 周辺確率座標空間

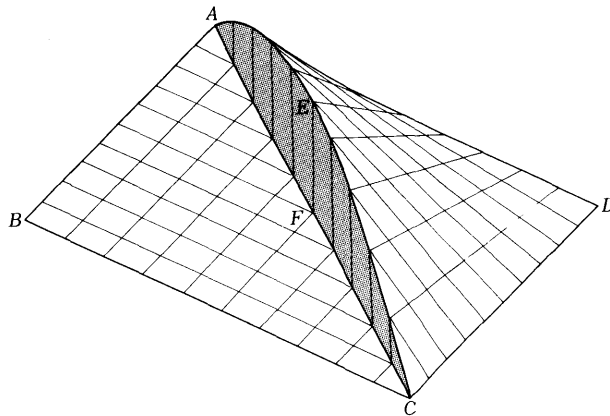


図2 正対称のモデル

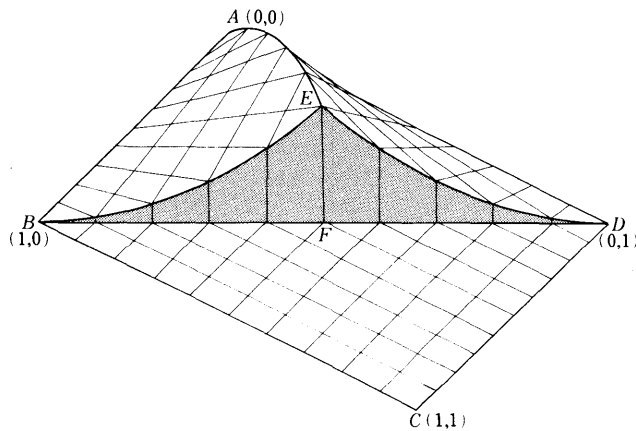


図3 歪対称のモデル

に対して、2種類の図式表示法を提案してみる。独立なパラメータが一般には3つあるので、3次元空間を用いて表現されることになる。

3.1 周辺確率座標空間 (marginal probability coordinate space)

これは、2組の属性を持つ確率 (周辺確率)

$$\Pr\{A\}=p, \Pr\{B\}=q$$

の座標系の点を足とし、独立性からの偏位を示すパラメータ d の値だけ鉛直に上下する ($d > 0$ なら上、 $d < 0$ なら下) 点として表示する方法である。図1に示したのは、正の association をもつモデル ($d \geq 0$) の全体である。この図の上表面で示される点の全体には、the most positively dependent model が対応していて、特に曲線 AEC は対角モデルの全体 $\{D^+(p) | 0 \leq p \leq 1\}$ を表わす。同様に曲線 BE には下三角モデル $L_0^+(p)$ 、曲線 ED には上三角モデル $U_0^+(p)$ が対応する。さらに下辺の平面 $ABCD$ は、いわゆる独立モデルの全体が対応し、一般の正の association をもったモデルは、この図形の内側の点で表示されることになる。

この図形を、 AC という対角線上で垂直に切りとった切り口を示したのが、図2である。網目を施した部分が正対称モデル $S(p, d)$ の全体に対応し、直線 AC はパラメータが1つの独立

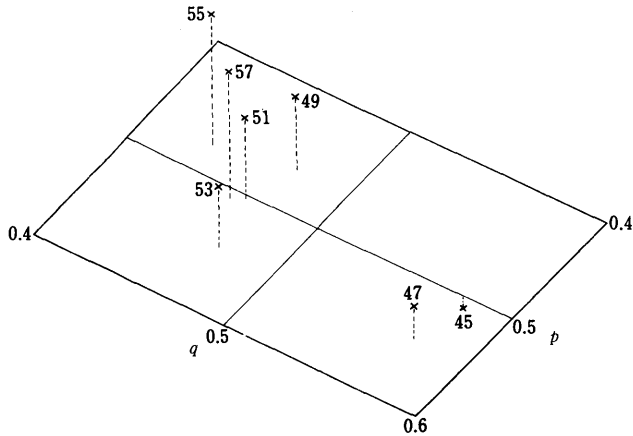


図4

表1. 先進国・途上国別輸出入の年次推移 (日本統計年鑑)

年次	輸 出		輸 入	
	先 進 国	途 上 国	先 進 国	途 上 国
45	3758	2818	3755	2723
46	4554	3435	3604	2975
47	4915	3445	3798	3053
48	5176	4323	5372	4412
49	7709	7349	7505	9652
50	6958	8198	7088	9189
51	9377	9167	7731	10652
52	10224	10098	7446	10790
53	9628	9537	6953	8962
54	10655	10270	9394	13670
55	13848	13471	11171	19310
56	16224	15163	11278	18480
57	16855	15490	11987	18832

モデル $I_0(p)$ に対応している。特に点 F に対応するモデルは、各セルの確率の値が $1/4$ である等確率モデル (equi-probability model) である。

図1の BD という対角線で切りとれば、図3のような切り口が得られ、歪対称のモデル $S^*(p, d)$ の全体に対応する。直線 BD に対応するのは、 $I_0^*(p)$ という独立モデルの全体である。

図4に、表1のデータの図式表現を与えてみた (全部を表示するとやや繁雑になるので奇数の年次のみを示してある)。輸出と輸入の金額が不均衡となるにつれ、分割表として眺めれば p の値が $1/2$ からずれるはずで、総じてみればバランスしていることが分かる。また q の値は、輸出入の合計 (貿易額) の、先進国と途上国のトータルに対する先進地域が占める比率を表わしていて、最近ではむしろ途上国との取り引きのウェイトが高くなっている様子が読みとれる。また独立モデルからのずれを示す垂線の足の長さは、先進国だけ (または途上国だけ) に対する輸出入の不均衡の程度を示しており、この数年間での先進国からでている不満のパターンを

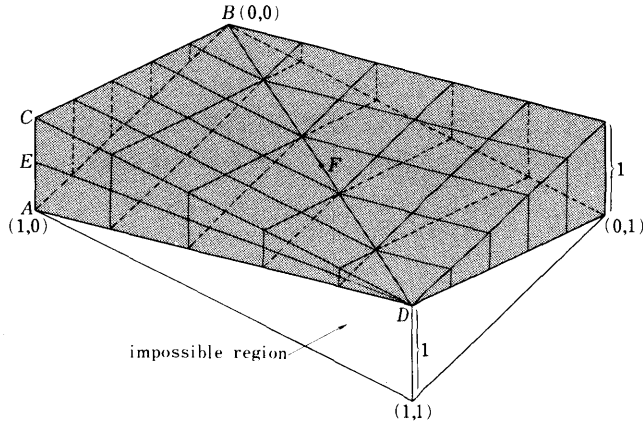


図5 条件付確率座標空間

観察することができる。

3.2. 条件付確率座標空間 (conditional probability coordinate space)

これは、2組の条件付確率

$$\Pr\{B | A\} = C_1, \Pr\{B | \bar{A}\} = C_2$$

の座標系の点を足とし、

$$\Pr\{B\} = q$$

という値を高さにもつ点として表示する方法である。図5に、このような表示法によるモデルの全体が示されている。

一般に $M\{p, q; d\}$ というモデルを、このような空間で表示すると、

$$C_1 = \Pr\{B | A\} = \frac{pq + d}{p} = q + \frac{d}{p}$$

$$C_2 = \Pr\{B | \bar{A}\} = \frac{(1-p)q - d}{1-p} = q - \frac{d}{1-p}$$

$$d = (C_1 - C_2)p(1-p)$$

$$p = \frac{-C_2}{C_1 - C_2}, \min\{C_1, C_2\} \leq q \leq \max\{C_1, C_2\}$$

といったような関係のあることから、 $C_1 = C_2$ という関係を満足する直線 BD が独立モデルの集合を表わしており、この直線を境にして左半分、右半分の領域がそれぞれ正、負の association をもつモデルに対応している。さらに

$$\text{線分}(AC) \iff D^+(p)$$

$$\text{三角形}(ACB) \iff L^+(p, q)$$

$$\text{三角形}(ACD) \iff U^+(p, q)$$

$$\text{線分}(ED) \iff U_0^+(p)$$

$$\text{三角形}(ABD) \iff p=0 \text{ のモデル}$$

$$\text{三角形}(CBD) \iff p=1 \text{ のモデル}$$

のような対応関係のあることも容易に確かめることができる。なお、図5での記号のつけ方は、図1の表示法と類似の記号の部分に対応するようにしてある。

なおこの条件付確率座標空間による図式表現法は、Shapiro (1982) の表示法を改良したもの

である。つまり彼は、 C_1, C_2 という2組の座標系のみ用いた平面内での表現だけを考えている。これでは、各点に単一のモデルではなく、モデルの部分集合が対応していることになる。Simpson (1951) は、2組の分割表を結合すると(つまり混合すると)、associationの強さが保証されないという例を提示した。Shapiro は彼の図式表現法を用いて、この Simpson's paradox の説明を試みている。そこで彼は、証明なしで次のような叙述を行っている。つまり、2つのモデル(の集合)

$$R=(C_1, C_2), \bar{R}=(\bar{C}_1, \bar{C}_2)$$

が独立モデルの集合を表わす $C_1=C_2$ という直線の同じ側にあっても、線分 $R\bar{R}$ を対角線とする C_1 軸、 C_2 軸に平行な矩形が対角線と交わるならば、 R, \bar{R} というモデル(族)と反対の符号の association をもつ "混合モデル" を作ることができると結論している。この結論は部分的には正しいが、一般には正しくない。このことを、条件付確率座標空間で考えるとかなり厄介になるが、先の周辺確率座標空間で考えれば簡単に証明できることを次節の初めで示すことにしよう。

§4. 混合モデルのいろいろなパターン

4.1 混合モデルと Simpson's paradox

いま

$$M\{p, q; d\}, M\{\tilde{p}, \tilde{q}; \tilde{d}\}$$

という2つのモデルが与えられたとして、これらが $1-\alpha, \alpha(0 \leq \alpha \leq 1)$ という比率で混合されたモデルについて考えてみる。

属性 A, B をもつ比率の、当初の2つのモデルでの値を p_{11}, \tilde{p}_{11} 、さらに混合モデルでの値を \bar{p}_{11} とすると

$$\begin{aligned} \bar{p}_{11} &= \alpha \tilde{p}_{11} + (1-\alpha)p_{11} \\ &= \alpha(\tilde{p}\tilde{q} + \tilde{d}) + (1-\alpha)(pq + d) \\ &= \bar{p}\bar{q} + \tilde{d} + e \end{aligned}$$

のように表現できることが分かる。ここで

$$\begin{aligned} \bar{p} &= \alpha \tilde{p} + (1-\alpha)p, \\ \bar{q} &= \alpha \tilde{q} + (1-\alpha)q, \\ \bar{d} &= \alpha \tilde{d} + (1-\alpha)d, \\ e &= \alpha(1-\alpha)(p-\tilde{p})(q-\tilde{q}). \end{aligned}$$

これより混合モデルは

$$M\{\bar{p}, \bar{q}; \bar{d} + e\}$$

のように表現できる。

特に当初の2つのモデルにおいて、少くとも一方の周辺確率が共通の場合には、 e の値が0となるので、混合モデルは $M\{\bar{p}, \bar{q}; \bar{d}\}$ のように単純な形となる。しかし、この単純な場合でも、混合モデルに対応する分割表から、混合比率 α を推定できるかとなると、一般には不可能である。

まずは、Simpson's paradox についての説明を済ませておくことにしよう。いま $M=M\{p, q; d\}, \bar{M}=M\{\tilde{p}, \tilde{q}; \tilde{d}\}$ という2つのモデルが与えられ、いずれも正の association をもつ ($d > 0, \tilde{d} > 0$) と仮定しよう。もし

$$(p-\tilde{p})(q-\tilde{q}) > 0$$

であれば、混合モデル

$$\bar{M} = (1-\alpha)M + \alpha\tilde{M}$$

に対する association は

$$\alpha\tilde{d} + (1-\alpha)d + \alpha(1-\alpha)(p-\tilde{p})(q-\tilde{q}) \geq \alpha\tilde{d} + (1-\alpha)d > 0$$

のように常に正となり, Simpson's paradox は成立しない. それでは

$$(p-\tilde{p})(q-\tilde{q}) = -\gamma < 0$$

の場合はどうか.

$$(2) \quad \alpha\tilde{d} + (1-\alpha)d < \gamma\alpha(1-\alpha)$$

という不等式が成り立つときに限り, 混合モデル \bar{M} の association は負になることが分かる. そこで

$$f(x) = \tilde{d}x + d(1-x) - \gamma x(1-x)$$

と置くと

$$f(0) = d, f(1) = \tilde{d}$$

で, $f(x)$ は

$$x^* = \frac{1}{2} + \frac{d-\tilde{d}}{2\gamma}$$

のときに最小値

$$-\frac{\{\gamma - (\sqrt{d} - \sqrt{\tilde{d}})^2\} \{\gamma - (\sqrt{d} + \sqrt{\tilde{d}})^2\}}{4\gamma}$$

をとることが分かる. 従って, (2) という不等式を満たすような α が (0, 1) の間に存在するためには

$$\gamma > |d - \tilde{d}|$$

でしかも

$$\gamma < (\sqrt{d} - \sqrt{\tilde{d}})^2 \text{ 又は } \gamma > (\sqrt{d} + \sqrt{\tilde{d}})^2$$

という 2 組の不等式が同時に満足する場合に限られる. ところで d, \tilde{d} が正の場合には

$$(\sqrt{d} - \sqrt{\tilde{d}})^2 < |d - \tilde{d}| < (\sqrt{d} + \sqrt{\tilde{d}})^2$$

という関係が成立しているから, 結局

$$(3) \quad \gamma > (\sqrt{d} + \sqrt{\tilde{d}})^2$$

という不等式が成立する場合, Simpson's paradox が成立することになる. 図 6 のような周辺

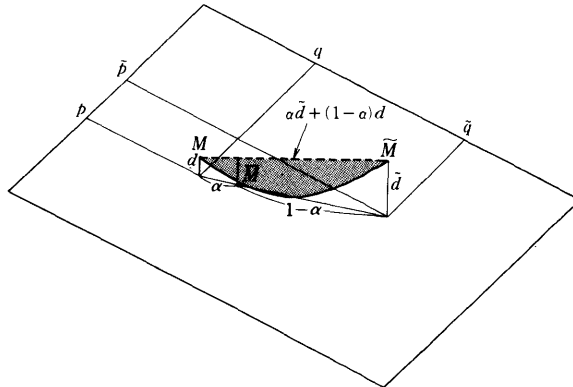


図 6

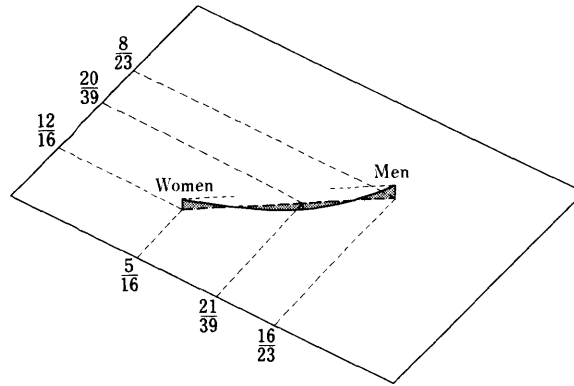


図7

確率座標空間で考えれば、モデル M, \bar{M} の混合モデルの全体は図の網目を施した部分の下側の曲線（点 M, \bar{M} を結ぶ線分 $\alpha\bar{d} + (1-\alpha)d$ から $\gamma\alpha(1-\alpha)$ だけ下がった曲線）で示されるから、この曲線が独立モデルの全体を表わす平面の下側に出る部分が存在すれば、負の association の混合モデルが対応することになる。このような曲線が独立モデルの平面と交わるための条件が (3) に他ならない。

Simpson (1951) の与えたデータの場合

$$d=0.0189, \bar{d}=0.0156, \gamma=0.154$$

であるから (3) という不等式を満しており、実際 $\alpha=16/39$ と置いた混合モデルの association は負になり、Simpson's paradox が成り立っている（図7参照）。

以下では、主として独立モデルとの混合モデルで、 α が推定可能となるケースの例をみていくことにしよう。

4.2 少なくとも一方の周辺確率が等しいモデルどうしの混合

まず、周辺確率が2つとも共通の場合でも

$$\alpha M\{p, q; d\} + (1-\alpha)M\{p, q; 0\} = M\{p, q; \alpha d\}$$

であるから、 d を任意の値としたのでは、混合モデルから α と d の値とを分離することはできない。そこで the most dependent model との混合モデルを考えてみる。

$$p \leq q \text{ ならば } d = p(1-q)$$

$$p \geq q \text{ ならば } d = (1-p)q$$

とすると、 $M\{p, q; d\}$ はそれぞれ $U^+(p, q), L^+(p, q)$ という三角モデルになり、

$$\bar{d} = \alpha p(1-q), \bar{d} = \alpha(1-p)q$$

といった関係より α を推定することができる（表2の(i), (ii)）。特に $p=q$ の場合には、 U^+ や L^+ は対角モデル D^+ となり、独立モデルのほうも $I_0(p)$ というパラメータが1個のモデルになる（表2のモデル(i)）。

次に、一組の周辺確率だけが等しい、 $p=\bar{p}$ という場合について考えてみる。

$$\alpha M\{p, \bar{q}; \bar{d}\} + (1-\alpha)M\{p, q; 0\} = M\{p, \bar{q}; \alpha\bar{d}\}$$

であるから、 $\bar{d}=d(p)$ のように、 \bar{d} が p の関数として与えられる場合には

$$\bar{d} = \alpha d(p)$$

の関係より α を推定することができる。また q も p の関数で与えられていれば

表 2. 少くとも一組の周辺確率が共通の独立モデルとの混合 ($p = \bar{p}, d = 0$)

	q	\bar{q}	\bar{d}	$M\{p, q; 0\}$	$M\{\bar{p}, \bar{q}; \bar{d}\}$	\bar{d}	α				
(イ)	q	q	$p(1-q)$ ($p \leq q$)	$I(p, q)$	$U^+(p, q)$	$\alpha p(1-q)$	$\bar{d}/\bar{p}(1-\bar{q})$				
(ロ)			$(1-p)q$ ($p \geq q$)					$L^+(p, q)$	$\alpha(1-p)q$	$\bar{d}/(1-\bar{p})\bar{q}$	
(ハ)	$1-p$	p	$p(1-p)$	$I_0(p)$	$D^+(p)$	$\alpha p(1-p)$	$\bar{d}/\bar{p}(1-\bar{p})$				
(ニ)				$I_0^*(p)$							
(ホ)	q	$1-p$	p^2 ($p \leq \frac{1}{2}$)	$I(p, q)$	$U_0^+(p)$	αp^2	\bar{d}/\bar{p}^2				
(ヘ)	p										
(ト)	$1-p$										
(チ)	q										
(リ)	p			$(1-p)^2$ ($p \geq \frac{1}{2}$)				$I_0(p)$	$L_0^+(p)$	$\alpha(1-p)^2$	$\bar{d}/(1-\bar{p})^2$
(ル)	$1-p$										
(レ)	q										
(ヲ)	$1-p$	q	$p(1-q)$ ($p \leq q$)	$I_0(p)$	$U^+(p, q)$	$\alpha p(1-q)$	$1-\bar{p}_{12}/\bar{p}^2$				
(ワ)			$I_0^*(p)$	$1-\bar{p}_{12}/\bar{p}(1-\bar{p})$							
(カ)	$1-p$	q	$(1-p)q$ ($p \geq q$)	$I_0(p)$	$L^+(p, q)$	$\alpha(1-p)q$	$1-\bar{p}_{21}/\bar{p}(1-\bar{p})$				
(コ)			$I_0^*(p)$	$1-\bar{p}_{21}/(1-\bar{p})^2$							
(ク)	p	$1-p$	0	$I_0(p)$	$I_0^*(p)$	0	$\frac{\bar{q}-\bar{p}}{1-2\bar{p}}$				
(ケ)			d	$S^*(p, d)$				αd			
(コ)	$1-p$	p	0	$I_0^*(p)$	$I_0(p)$	0	$\frac{1-\bar{p}-\bar{q}}{1-2\bar{p}}$				
(セ)			d	$S(p, d)$				αd			

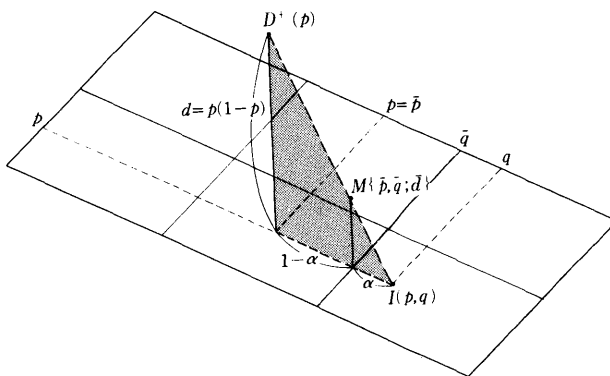


図 8 対角モデルと独立モデルの混合

$$\bar{q} = \alpha \bar{q} + (1-\alpha)q$$

の関係より \bar{q} の値も推定可能である。特に

$$q = p, \bar{d} = p(1-p) \Rightarrow D^+(p) \quad (\text{表 2 の (イ), (ニ), (ホ)})$$

$$q = 1-p, \bar{d} = p^2 \left(p \leq \frac{1}{2} \right) \Rightarrow U_0^+(p) \quad (\text{表 2 の (ヘ), (ト), (チ)})$$

$$q = 1-p, \bar{d} = (1-p)^2 \left(p \geq \frac{1}{2} \right) \Rightarrow L_0^+(p) \quad (\text{表 2 の (リ), (ヌ), (ル)})$$

図8には、(㊸)のタイプの混合パターンを示してみた。

さらに、 $M\{p, \bar{q}; \bar{d}\}$ が the most dependent model ならば、少なくとも1個のセルは0になるから、 $M\{p, q; 0\}$ の q が p の関数になっていれば、 α は推定可能となる。例えば

$$aL^+(p, q) + (1-\alpha)I_0(p)$$

の場合で考えてみると

$$\bar{p}_{12} = (1-\alpha)p(1-p) = (1-\alpha)\bar{p}(1-\bar{p})$$

となるから、これより α を推定できる。このようなケースは、表2の(㉗)、(㉘)、(㉙)、(㉚)に示される。一般に、任意の \bar{d} に対して α が推定可能となる場合は、 p, q, \bar{p}, \bar{q} が唯一つのパラメータで表わされる場合で、例えば

$$p = \bar{p} = q = 1 - \bar{q} \implies S(p, \bar{d}) \quad (\text{表2の(㉘), } \bar{d}=0 \text{ なら(㉙)})$$

$$p = \bar{p} = 1 - q = \bar{q} \implies S^*(p, \bar{d}) \quad (\text{表2の(㉚), } \bar{d}=0 \text{ なら(㉛)})$$

のようなパターンである。この場合には

$$\bar{q} = \alpha p + (1-\alpha)(1-p) \quad (\text{又は } \bar{q} = \alpha(1-p) + (1-\alpha)p)$$

といった関係が成立するから、混合モデルの周辺確率の値だけを用いて α が推定できる。次いで

$$\bar{d} = \alpha \bar{d}$$

の関係より \bar{d} の値も推定できる。(㉘)や(㉚)のパターンで特に $\bar{d}=0$ の場合は、それぞれ(㉙)、(㉛)のパターンの混合モデルになる。

4.3 2組の周辺確率がいずれも異なるモデルどうしの混合

まず2つのモデルが、共に独立であるケースについて考える。この場合には、 $\bar{d}=0$ となるから、混合モデルの cross-product difference を用いて

$$(4) \quad e = \alpha(1-\alpha)(p-\bar{p})(q-\bar{q})$$

という値が推定できる。次に周辺確率に関する関係

$$(5) \quad \bar{p} = \alpha p + (1-\alpha)\bar{p} = \bar{p} + \alpha(p-\bar{p})$$

$$(6) \quad \bar{q} = \alpha q + (1-\alpha)\bar{q} = \bar{q} - (1-\alpha)(q-\bar{q})$$

表3. 2組の周辺確率が異なる独立モデルどうしの混合

	q	\bar{p}	\bar{q}	$M\{p, q; 0\}$	$M\{\bar{p}, \bar{q}; 0\}$	$p^* = 2p - 1$	$q^* = 2q - 1$	$2(2\alpha - 1)$
(ネ)	p	$1-p$	$1-p$	$I_0(p)$	$I_0(1-p)$	$B \pm A^*$	—	$\pm(1-2\bar{p})/A^*$
(ナ)	$1-p$	$1-p$	p	$I_0^*(p)$	$I_0^*(1-p)$	$A \pm B^*$	—	$\pm(1-2\bar{p})/B^*$
(ラ)	q	$1-p$	$1-p$	$I(p, q)$	$I_0(1-p)$	$B \pm A^*$	*	$(1-2\bar{p})/(B \pm A^*)$
(ル)		$1-q$	$1-q$		$I_0(1-q)$	*	$-B \pm A^*$	$(1-2\bar{q})/(-B \pm A^*)$
(ロ)		$1-p$	p		$I_0^*(1-p)$	$A \pm B^*$	*	$(1-2\bar{p})/(A \pm B^*)$
(リ)		q	$1-q$		$I_0^*(q)$	*	$A \pm B^*$	$(1-2\bar{q})/(A \pm B^*)$
(レ)		q	p		$I(p, q)$	$A \pm B^*$	$A \mp B^*$	$\pm 2B/B^*$
(ク)		q	$1-p$		$I(q, 1-p)$	$-B \pm A^*$	$A \pm B^*$	$(A \pm B^*)/(-B \pm A^*)$
(ケ)		$1-q$	p		$I(1-q, p)$	$A \pm B^*$	$-B \pm A^*$	$(A \pm B^*)/(-B \pm A^*)$
(コ)		$1-p$	$1-q$		$I(1-p, 1-q)$	*	*	$\pm 1/\sqrt{1+C}$

$$A = \bar{p} + \bar{q} - 1, \quad A^* = \sqrt{A^2 + 4e}$$

$$B = \bar{p} - \bar{q}, \quad B^* = \sqrt{B^2 - 4e}$$

$$C = 4e/(1-2\bar{p})(1-2\bar{q})$$

から得られる

$$\alpha(p - \tilde{p}) = \bar{p} - \tilde{p}, (1 - \alpha)(q - \tilde{q}) = q - \tilde{q}$$

という式を(4)に代入して

$$(7) \quad (\bar{p} - \tilde{p})(q - \tilde{q}) = e$$

という関係式を得る。同様の考えによって

$$(8) \quad (p - \tilde{p})(\bar{q} - \tilde{q}) = e$$

という関係も得られる。

まず q や \tilde{p} が p の関数になっている場合には、(7)の関係より p が推定できるから、これを(5)に代入することにより α が推定できる(表3の(ネ)、(ハ)のケース)。

また \tilde{p} , \tilde{q} が p の関数として与えられている場合には、(8)の関係よりまず p を推定し、これを(5)に代入して α の値を推定することができる(ウ)、(ウ)のケース)。

さらに、 \tilde{p} , \tilde{q} が q の関数として与えられていれば(表3の(ム)、(ノ)のパターン)、(7)によって q の値を推定し、これを(6)に代入して α が推定できる。

次いで \tilde{p} が q の関数、 \tilde{q} が p の関数といった場合はどうか。これも(8)、(5)の関係式よりそれぞれ p , q を推定し、これらを(6)の関係に代入して α を推定することができる(表3の(カ)、(ク)、(ケ)のケース)。

最後に \tilde{p} は p の関数、 \tilde{q} は q の関数といった場合には、一般には α が推定できない。しかし、 $\tilde{p} = f(p)$ という関数が

$$(9) \quad \frac{f(\tilde{p}) - \tilde{p}}{f(p) - p} = c \quad (\text{一定})$$

という関係を満す1次関数、 $\tilde{q} = g(q)$ についても

$$\frac{g(\tilde{q}) - \tilde{q}}{g(q) - q} = d$$

を満す1次関数になっていればよい。(4)の関係より

$$f(\tilde{p}) = \alpha f(p) + (1 - \alpha)f(\tilde{p})$$

従って

$$\begin{aligned} \bar{p} - f(\tilde{p}) &= \alpha(p - f(p)) + (1 - \alpha)(\tilde{p} - f(\tilde{p})) \\ &= (c + (1 - c)\alpha)(p - \tilde{p}) \end{aligned}$$

同様に

$$\bar{q} - g(\tilde{q}) = (d + (1 - d)\alpha)(q - \tilde{q})$$

といった関係も得られるから、これらを(4)の関係に代入して

$$(10) \quad \frac{\alpha(1 - \alpha)}{[c + (1 - c)\alpha][d + (1 - d)\alpha]} = \frac{e}{[\bar{p} - f(\tilde{p})][\bar{q} - g(\tilde{q})]}$$

という α に関する2次式が得られ、これから α の値が推定できる。

ところで(9)のような関係を満すためには

$$f(p) = b + cp$$

のようになっていればよいことが確かめられる。特に $b = 1$, $c = -1$ とし、 g も同じ形の関数とすると(10)の関係は

$$\frac{4\alpha(1 - \alpha)}{(1 - 2\alpha)^2} = C \quad \left(C = \frac{4e}{(1 - 2\tilde{p})(1 - 2\tilde{q})} \right)$$

となり、

$$\alpha = \frac{1}{2} \pm \frac{1}{2\sqrt{C+1}}$$

これは表3の(マ)のケースになっている。

2組の周辺確率が異なり、しかも独立でないモデルとの混合で、混合比率 α が推定できる場合となると、対称モデルと三角モデルとの組合せのケースに限られる。(歪対称モデルと三角モデルの混合モデルでも α は推定可能であるが、この場合は負の association をもつケースになる。)

対称モデルのほうを $M\{p, p; p(1-p)-\delta\}$ 、三角モデルのほうを $M\{\bar{p}, \bar{q}; \bar{d}\}$ と表わす。ここで \bar{p} や \bar{q} は p と δ の関数、また \bar{d} は

$$\bar{d} = \min\{\bar{p}(1-\bar{q}), (1-\bar{p})\bar{q}\}$$

でなければならない。このような混合モデルから α が推定できる場合の一例は

$$\frac{\bar{q}-\bar{p}}{\delta} = c \quad (\text{一定})$$

といったケースである。 c が正なら上三角モデル、負の値なら下三角モデルになる。 $c > 0$ のときには

$$\begin{aligned} \bar{p}_{12} &= (1-\alpha)\delta \\ \bar{p}_{21} &= \alpha(\bar{q}-\bar{p}) + (1-\alpha)\delta = (1+(c-1)\alpha)\delta \end{aligned}$$

となるから

$$\frac{\bar{p}_{12}}{\bar{p}_{21}} = \frac{1-\alpha}{1+(c-1)\alpha}$$

これを α について解いて

$$\alpha = \frac{\bar{p}_{21} - \bar{p}_{12}}{(c-1)\bar{p}_{12} + \bar{p}_{21}}$$

という関係が得られる。また $c < 0$ の場合も同様に

$$\alpha = \frac{\bar{p}_{12} - \bar{p}_{21}}{\bar{p}_{12} - (c+1)\bar{p}_{21}}$$

といった関係が示される。表4の(ウ)、(フ)に示したモデルは、それぞれ $c=2$ 、 $c=-2$ のケースに相当する。

§5. 潜在構造分析

前節では、特殊モデルどうしの混合モデルのいろいろなパターンについて概観した。そこでこの節では、これらを利用して、実際の分割表に対する潜在構造分析を行ってみる。

5.1 潜在構造分析の諸相の概観

(i) 対角モデルと独立モデルとの混合

常に視聴率が一定の番組があるとする。前回その番組をみた(A)かみなかった(\bar{A})か、また今回もみる(B)かみない(\bar{B})かにより、結果は4分割される。まず、その番組を毎回みるか、絶対みないという極端な視聴パターンには、 $D^+(p)$ というモデルを対応させる。もう一方の極端は、前後関係を無視してみたりみなかったりというパターンで、 $I_0(p)$ という独立モデルを対応させることができる。

従って全体的な視聴パターンには

$$\alpha D^+(p) + (1-\alpha)I_0(p)$$

という混合モデルが対応できる。これは表2(ウ)のパターンである。 α は、番組を継続して視聴させたり、させなかったりする度合いを表すことになる。

また、同じ相手に同じ質問を2度続けて行うというパネル調査の結果に対して、1回目の調

査である意見に賛成 (A) か反対 (\bar{A}) かを問い、続いて 2 回目の調査でもその意見に賛成 (B) か反対 (\bar{B}) かにより、結果は 4 分割される。このような調査で、2 回とも必ず同じ答えをするパターンには、 $D^+(p)$ というモデルが対応する。反対に、1 回目と 2 回目の調査に対し前後の脈絡なしで答えるタイプには $I(p, q)$ という独立モデルを対応させることができる。

パネル調査の全体的な反応には

$$\alpha D^+(p) + (1-\alpha)I(p, q)$$

という混合モデルを対応させることができる。これは、表 2(†) のパターンになる。この α で、パネル調査の結果の安定性をみることができる。

さらに前にみたように、大竜巻のような災害の生起に関する予想と結果にも、同じような混合モデルを対応させることができた。その際の完璧な予測には、 $D^+(p)$ ではなく $D^+(q)$ であったから、混合モデルは

$$\alpha D^+(q) + (1-\alpha)I(p, q)$$

となり、(†) のパターンとは若干異なった形になる。

(ii) 三角モデルと独立モデルの混合

ある特定銘柄の商品を、前回購入した (A) かしなかった (\bar{A}) か、また今回も同じ銘柄を購入する (B) かしない (\bar{B}) かにより結果は 4 分割される。一つの銘柄を一度購入したら以後は必ず同じものを求めるというパターンは、 $U^+(p, q)$ ($p \leq q$) というモデルを対応させることができる。また銘柄には一切無傾着な層には、全体としての購入率は一定として、 $I_0(p)$ という独立モデルを対応させればよい。

結局、全体的な購入パターンには

$$\alpha U^+(p, q) + (1-\alpha)I_0(p)$$

というモデルを対応させればよい。これは、表 2(ア) のパターンになっている。 α はその銘柄の固定客が増加する傾向の度合いを表わしている。

パネル調査の 1 回目で革新を支持する (A) か保守を支持する (\bar{A}) か、また 2 回目の調査でも革新を支持すると答える (B) か、保守と答える (\bar{B}) かによって、結果は 4 分割される。保守から革新へは決して変更しないといったグループには $L^+(p, q)$ というモデルが対応し、保守か革新かはその時で変わるが全体としては比率が一定とするグループには $I_0(p)$ が対応する。

かくて全体的な反応パターンには

$$\alpha L^+(p, q) + (1-\alpha)I_0(p)$$

という (カ) のモデルが対応する。なお、この種の混合モデルが対応できるためには、 $p > q$ という関係が成りたたなければいけないから、結局、保守のほうがのびているといった傾向になければならない。 α はそのような傾向の度合いを表わす尺度と考えられる。

ある年に結婚した夫婦の前歴で、夫が初婚 (A) か再婚 (\bar{A}) か、妻が初婚 (B) か再婚 (\bar{B}) かにより、結果は 4 分割される。初めて結婚する男は必ず初婚の女性と結婚するという行動パターンは、 $U^+(p, q)$ ($p \leq q$)、というモデルで表わされる。もう一方の極端としては $I(p, q)$ という独立モデルを対応させれば、観測データには

$$\alpha U^+(p, q) + (1-\alpha)I(p, q)$$

という (イ) の混合モデルが対応できる。この α は、Bénini の与えた初婚の夫が再婚の妻との結婚を忌避する度合いの尺度と一致する。

(iii) その他の混合モデル

先にパネル調査の結果を、対角モデル $D^+(p)$ と独立モデル $I(p, q)$ の混合モデルとして捉える方法について考えたが、これを次のように考えることもできる。パネル調査の意見に潜在的には賛成であるが、調査の都度、意味をとり違えるなどして反対の答えをしてしまう可能性が

1-p だけあると考えると、パネル調査の結果には $I_0(p)$ という独立モデルを対応させる。また潜在的には反対だが、やはり調査の都度賛成と答えてしまう可能性が q だけあるといった反応パターンには、 $I_0(q)$ というモデルを対応させる。

結局パネル調査の全体的な結果には

$$\alpha I_0(p) + (1-\alpha)I_0(q)$$

という独立モデルどうしの混合モデルを対応させることができる。ところがこのモデルから α を推定することは不可能なことが分かる。特に

$$p=1-\varepsilon, q=\varepsilon$$

といったケースでは表3の(※)のパターンになり、 ε や α の値が推定できる。Lazarsfeld-Kendall はこの ε に対して、index of turnover という呼称を与えている。

このモデルはもう少し一般化できる。つまり、1回目の調査での回答での、潜在的に賛成グループと反対グループの誤回答率を $1-p$ とし、2回目の回答では $1-q$ とすれば、パネル調査全体の結果には

$$\alpha I(p, q) + (1-\alpha)I(1-p, 1-q)$$

という混合モデル(表3の(㉞)のパターン)を対応させることができる。特に $p=q$ の場合が先に考えたケースになる。

最後に従属モデルどうしの混合モデルのケースについて考えてみよう。前に、ある年に結婚した夫婦の組合せを初婚と再婚別の4分表に対して、三角モデル $U^+(p, q)$ と独立モデル $I(p, q)$ との混合モデルを対応させて、 α は初婚の夫が再婚の妻との結婚を忌避する度合いを表わすものと考えた。この例に対して、次のような混合モデルを対応させることもできる。つまり三角モデルと独立モデルとの混合の代わりに、三角モデルと対称モデルとの混合

$$\alpha U^+(p-\delta, p+\delta) + (1-\alpha)S^+(p, d)$$

を考えてやる。ここで

$$d=p(1-p)-\delta$$

という関係が成立しているものとする。これは表4の(㉟)のパターンである。次の節の初めで、これら2つのパターンの混合モデルの比較を行ってみる。

表4. 2組の周辺確率が異なる従属モデルどうしの混合

	q	\bar{p}	\bar{q}	d	\bar{d}	$M\{p, q; d\}$	$M\{\bar{p}, \bar{q}; \bar{d}\}$
(ケ)	p	$p-\delta$	$p+\delta$	$p(1-p)-\delta$	$p(1-p)-\delta(1-\delta)$	$S^+(p, d)$	$U^+(p-\delta, p+\delta)$
(フ)		$p+\delta$	$p-\delta$				$L^+(p+\delta, p-\delta)$

5.2 潜在構造分析の実際

(i) 初婚の夫は再婚の妻との結婚を嫌う

表5のデータは、厚生省の人口動態統計による初婚・再婚別婚姻件数を5年間隔で示したもので、カッコ内の数値は各年ごとのパーセントである。(婚姻件数とはその年に届出のあったもので、実際に結婚したという事実婚とは多少の喰い違いがある。)

これら6組の2×2分割表タイプの時系列データを、三角モデルと独立モデルとの混合として表わしたのが表6である。また、三角モデルと対称モデルの混合として表わすと表7のようになる。三角モデルに対する係数が α で、初婚の夫が再婚の妻との結婚を忌避する度合いを表わすと考えられた。ところで表6、と表7のこれらの数値を比較してみると、最近のデータほど大きく喰い違っていることが分かる。これは、Bénini の提示したイタリアの1898年のデータで

表 5. 初婚・再婚別婚姻件数 (人口動態統計)

年次	夫初婚		夫再婚		計
	妻初婚	妻再婚	妻初婚	妻再婚	
1955	601924 (84.20)	24470 (3.42)	54667 (7.65)	33800 (4.73)	714861
1960	758429 (87.57)	23592 (2.72)	54168 (6.25)	29926 (3.46)	866115
1965	847564 (88.76)	25085 (2.63)	52740 (5.52)	29463 (3.09)	954852
1970	914870 (88.87)	28913 (2.82)	52846 (5.13)	32776 (3.18)	1029405
1975	822382 (87.34)	33443 (3.55)	49063 (5.21)	36740 (3.90)	941628
1980	657373 (84.86)	33512 (4.32)	44042 (5.69)	39775 (5.13)	774702

表 6. 表 5 に対する三角モデルと独立モデルの混合

1955													
	(.521) ×	<table border="1"> <tr><td>87.62</td><td>0</td></tr> <tr><td>4.23</td><td>8.15</td></tr> </table>	87.62	0	4.23	8.15	+	(.479) ×	<table border="1"> <tr><td>80.48</td><td>7.14</td></tr> <tr><td>11.37</td><td>1.01</td></tr> </table>	80.48	7.14	11.37	1.01
87.62	0												
4.23	8.15												
80.48	7.14												
11.37	1.01												
1960													
	(.512) ×	<table border="1"> <tr><td>90.29</td><td>0</td></tr> <tr><td>3.53</td><td>6.18</td></tr> </table>	90.29	0	3.53	6.18	+	(.488) ×	<table border="1"> <tr><td>84.71</td><td>5.58</td></tr> <tr><td>9.11</td><td>0.60</td></tr> </table>	84.71	5.58	9.11	0.60
90.29	0												
3.53	6.18												
84.71	5.58												
9.11	0.60												
1965													
	(.497) ×	<table border="1"> <tr><td>94.39</td><td>0</td></tr> <tr><td>2.90</td><td>5.71</td></tr> </table>	94.39	0	2.90	5.71	+	(.503) ×	<table border="1"> <tr><td>86.17</td><td>5.22</td></tr> <tr><td>8.12</td><td>0.49</td></tr> </table>	86.17	5.22	8.12	0.49
94.39	0												
2.90	5.71												
86.17	5.22												
8.12	0.49												
1970													
	(.489) ×	<table border="1"> <tr><td>91.68</td><td>0</td></tr> <tr><td>2.33</td><td>5.99</td></tr> </table>	91.68	0	2.33	5.99	+	(.511) ×	<table border="1"> <tr><td>86.19</td><td>5.49</td></tr> <tr><td>7.82</td><td>0.50</td></tr> </table>	86.19	5.49	7.82	0.50
91.68	0												
2.33	5.99												
86.19	5.49												
7.82	0.50												
1975													
	(.476) ×	<table border="1"> <tr><td>90.89</td><td>0</td></tr> <tr><td>1.65</td><td>7.46</td></tr> </table>	90.89	0	1.65	7.46	+	(.524) ×	<table border="1"> <tr><td>84.11</td><td>6.78</td></tr> <tr><td>8.43</td><td>0.60</td></tr> </table>	84.11	6.78	8.43	0.60
90.89	0												
1.65	7.46												
84.11	6.78												
8.43	0.60												
1980													
	(.487) ×	<table border="1"> <tr><td>89.18</td><td>0</td></tr> <tr><td>1.36</td><td>9.46</td></tr> </table>	89.18	0	1.36	9.46	+	(.513) ×	<table border="1"> <tr><td>80.74</td><td>8.44</td></tr> <tr><td>9.80</td><td>1.02</td></tr> </table>	80.74	8.44	9.80	1.02
89.18	0												
1.36	9.46												
80.74	8.44												
9.80	1.02												

は、2つの混合モデルによる α の推定値がそうは違わなかったことと、かなり様変わりしている。

表 6 で与えられる独立モデルの項を眺めてみると、各年でかなりばらついているが、表 7 の対称モデルのほうはかなり安定していることが分かる。従って表 6 のデータに関する限りでは、対称モデルとの混合モデルを当てはめるほうが妥当な感じである。表 7 の α の値が最近になって減少していることから、初婚の夫が再婚の妻との結婚を嫌う傾向が少なくなっていることが分かる。

(ii) 生れ変わったときの希望の性

表 8 のデータは、統計数理研究所が 5 年ごとに行っている「日本人の国民性の研究」の標本調査の 1 つの質問で、もしもう一度生れ変るとして男と女のどちらを好むかという問に関する

表7. 表5に対する三角モデルと対称モデルの混合

1955	(.382) ×	<table border="1"><tr><td>84.20</td><td>0</td></tr><tr><td>11.07</td><td>4.73</td></tr></table>	84.20	0	11.07	4.73	+	(.618) ×	<table border="1"><tr><td>84.20</td><td>5.53</td></tr><tr><td>5.54</td><td>4.73</td></tr></table>	84.20	5.53	5.54	4.73
84.20	0												
11.07	4.73												
84.20	5.53												
5.54	4.73												
1960	(.393) ×	<table border="1"><tr><td>87.57</td><td>0</td></tr><tr><td>8.97</td><td>3.46</td></tr></table>	87.57	0	8.97	3.46	+	(.607) ×	<table border="1"><tr><td>87.57</td><td>4.48</td></tr><tr><td>4.49</td><td>3.46</td></tr></table>	87.57	4.48	4.49	3.46
87.57	0												
8.97	3.46												
87.57	4.48												
4.49	3.46												
1965	(.355) ×	<table border="1"><tr><td>88.76</td><td>0</td></tr><tr><td>8.15</td><td>3.09</td></tr></table>	88.76	0	8.15	3.09	+	(.645) ×	<table border="1"><tr><td>88.76</td><td>4.08</td></tr><tr><td>4.07</td><td>3.09</td></tr></table>	88.76	4.08	4.07	3.09
88.76	0												
8.15	3.09												
88.76	4.08												
4.07	3.09												
1970	(.293) ×	<table border="1"><tr><td>88.87</td><td>0</td></tr><tr><td>7.95</td><td>3.18</td></tr></table>	88.87	0	7.95	3.18	+	(.707) ×	<table border="1"><tr><td>88.87</td><td>3.98</td></tr><tr><td>3.97</td><td>3.18</td></tr></table>	88.87	3.98	3.97	3.18
88.87	0												
7.95	3.18												
88.87	3.98												
3.97	3.18												
1975	(.189) ×	<table border="1"><tr><td>87.34</td><td>0</td></tr><tr><td>8.76</td><td>3.90</td></tr></table>	87.34	0	8.76	3.90	(.811) ×	<table border="1"><tr><td>87.34</td><td>4.38</td></tr><tr><td>4.38</td><td>3.90</td></tr></table>	87.34	4.38	4.38	3.90	
87.34	0												
8.76	3.90												
87.34	4.38												
4.38	3.90												
1980	(.136) ×	<table border="1"><tr><td>84.86</td><td>0</td></tr><tr><td>10.01</td><td>5.13</td></tr></table>	84.86	0	10.01	5.13	(.864) ×	<table border="1"><tr><td>84.86</td><td>5.00</td></tr><tr><td>5.01</td><td>5.13</td></tr></table>	84.86	5.00	5.01	5.13	
84.86	0												
10.01	5.13												
84.86	5.00												
5.01	5.13												

表8. 生れ変りの希望の性 ([6])

年次	生れ変り		男 性		女 性		計
	男 に	女 に	男 に	女 に	男 に	女 に	
1958	614 (45.42)	36 (2.66)	492 (36.39)	210 (15.53)			1352
1963	1099 (43.91)	89 (3.56)	796 (31.80)	519 (20.73)			2503
1968	1264 (45.14)	73 (2.61)	695 (24.82)	768 (27.43)			2800
1973	1833 (42.71)	111 (2.59)	1067 (24.86)	1281 (29.84)			4292
1978	749 (41.84)	37 (2.07)	445 (24.86)	559 (31.23)			1790

集計結果を示したものである。

これもやはり、2×2分割表タイプの5組の時系列データと考えられる。これに対して、三角モデルと対称モデルの混合として表9、また対角モデルと独立モデルの混合として表10のような結果が得られる。表9の三角モデルや対称モデルの最後のセルの値が大きく変わっていて、今度のデータにはこの混合モデルがうまく適合しているとはいえない。これは表8のデータで、女性の意見だけが大きく変化してきているためである。

一方、表10の対角モデルと独立モデルの混合モデルでみると、こちらはいずれの年についてもかなり安定していて、こちらのモデルを用いるほうがよいと判断できる。次に生れ変わるなら、男性は次も男、女性は次も女というように、同一の性を希望するパターンは対角モデルで表わされるから、そのモデルに対する混合比率に注目すると、25%弱から50%強に増加している。従って最近では、同一の性への生れ変りを希望する傾向が強まってきていることが分かる。男性優位の社会が崩壊しつつあるといった現象への意識の変化を読みとることができる。

表9. 表8に対する三角モデルと対称モデルの混合

<u>1958</u>													
	(.864) ×	<table border="1"><tr><td>45.42</td><td>0</td></tr><tr><td>39.05</td><td>15.53</td></tr></table>	45.42	0	39.05	15.53	+	(.136) ×	<table border="1"><tr><td>45.42</td><td>19.53</td></tr><tr><td>19.52</td><td>15.53</td></tr></table>	45.42	19.53	19.52	15.53
45.42	0												
39.05	15.53												
45.42	19.53												
19.52	15.53												
<u>1963</u>													
	(.799) ×	<table border="1"><tr><td>43.91</td><td>0</td></tr><tr><td>35.36</td><td>20.73</td></tr></table>	43.91	0	35.36	20.73	+	(.201) ×	<table border="1"><tr><td>43.91</td><td>17.68</td></tr><tr><td>17.68</td><td>20.73</td></tr></table>	43.91	17.68	17.68	20.73
43.91	0												
35.36	20.73												
43.91	17.68												
17.68	20.73												
<u>1968</u>													
	(.810) ×	<table border="1"><tr><td>45.14</td><td>0</td></tr><tr><td>27.43</td><td>27.43</td></tr></table>	45.14	0	27.43	27.43	+	(.190) ×	<table border="1"><tr><td>45.14</td><td>13.72</td></tr><tr><td>13.71</td><td>27.43</td></tr></table>	45.14	13.72	13.71	27.43
45.14	0												
27.43	27.43												
45.14	13.72												
13.71	27.43												
<u>1973</u>													
	(.812) ×	<table border="1"><tr><td>42.71</td><td>0</td></tr><tr><td>27.45</td><td>29.84</td></tr></table>	42.71	0	27.45	29.84	+	(.188) ×	<table border="1"><tr><td>42.71</td><td>13.73</td></tr><tr><td>13.72</td><td>29.84</td></tr></table>	42.71	13.73	13.72	29.84
42.71	0												
27.45	29.84												
42.71	13.73												
13.72	29.84												
<u>1978</u>													
	(.846) ×	<table border="1"><tr><td>41.84</td><td>0</td></tr><tr><td>26.93</td><td>31.23</td></tr></table>	41.84	0	26.93	31.23	+	(.154) ×	<table border="1"><tr><td>41.84</td><td>13.46</td></tr><tr><td>13.47</td><td>31.23</td></tr></table>	41.84	13.46	13.47	31.23
41.84	0												
26.93	31.23												
41.84	13.46												
13.47	31.23												

表10. 表8に対する対角モデルと独立モデルの混合

<u>1958</u>													
	(.244) ×	<table border="1"><tr><td>48.08</td><td>0</td></tr><tr><td>0</td><td>51.92</td></tr></table>	48.08	0	0	51.92	+	(.756) ×	<table border="1"><tr><td>44.57</td><td>3.51</td></tr><tr><td>48.13</td><td>3.79</td></tr></table>	44.57	3.51	48.13	3.79
48.08	0												
0	51.92												
44.57	3.51												
48.13	3.79												
<u>1963</u>													
	(.320) ×	<table border="1"><tr><td>47.47</td><td>0</td></tr><tr><td>0</td><td>52.53</td></tr></table>	47.47	0	0	52.53	+	(.680) ×	<table border="1"><tr><td>42.14</td><td>5.33</td></tr><tr><td>46.63</td><td>5.90</td></tr></table>	42.14	5.33	46.63	5.90
47.47	0												
0	52.53												
42.14	5.33												
46.63	5.90												
<u>1968</u>													
	(.470) ×	<table border="1"><tr><td>47.75</td><td>0</td></tr><tr><td>0</td><td>52.25</td></tr></table>	47.75	0	0	52.25	+	(.530) ×	<table border="1"><tr><td>42.82</td><td>4.93</td></tr><tr><td>46.85</td><td>5.40</td></tr></table>	42.82	4.93	46.85	5.40
47.75	0												
0	52.25												
42.82	4.93												
46.85	5.40												
<u>1973</u>													
	(.488) ×	<table border="1"><tr><td>45.30</td><td>0</td></tr><tr><td>0</td><td>54.70</td></tr></table>	45.30	0	0	54.70	+	(.512) ×	<table border="1"><tr><td>40.23</td><td>5.07</td></tr><tr><td>48.57</td><td>6.13</td></tr></table>	40.23	5.07	48.57	6.13
45.30	0												
0	54.70												
40.23	5.07												
48.57	6.13												
<u>1978</u>													
	(.510) ×	<table border="1"><tr><td>43.91</td><td>0</td></tr><tr><td>0</td><td>56.09</td></tr></table>	43.91	0	0	56.09	+	(.490) ×	<table border="1"><tr><td>39.69</td><td>4.22</td></tr><tr><td>50.70</td><td>5.39</td></tr></table>	39.69	4.22	50.70	5.39
43.91	0												
0	56.09												
39.69	4.22												
50.70	5.39												

以上の分析事例でみたように、同じタイプの2×2分割表のデータが与えられた場合には、分割された特殊モデルの構造が比較的安定しているような混合モデルを対峙させて、その混合比率の推移を眺めることによって、データに含まれる潜在構造の変異を探ることができる。

謝 辞

本稿の作成にあたり、適切な助言をいただいた査読者および編集委員会に謝意を表します。

参 考 文 献

Goodman, L.A. and Kruskal, W.H. (1959). Measures of association for cross classifications II.

Further discussion and references, *J. Amer. Statist. Assoc.*, **58**, 310-364.

厚生省 (1955~1980). 人口動態統計.

Shapiro, S.H. (1982). Collapsing contingency tables—a geometric approach, *Amer. Statistician*, **36-1**, 43-46.

Simpson, E.H. (1951). The interpretation of interaction in contingency tables, *J. Roy. Statist. Soc. B-13*, 238-241.

総理府統計局 (1960~1982). 日本統計年鑑.

統計数理研究所国民性調査委員会 (1982). 第4日本人の国民性, 出光書店.

Graphical Representation for 2x2 Contingency Table Models and Latent Structure Analysis

Giitiro Suzuki

(The Institute of Statistical Mathematics)

In general, 2x2 contingency table can be expressed using three parameters as follows :

$pq + d$	$p(1-q) - d$	p
$(1-p)q - d$	$(1-p)(1-q) + d$	$1-p$
q	$1-q$	1

This model may be denoted by $M(p, q; d)$. The parameter d denotes an association of the model and can change

$$-\min\{pq, (1-p)(1-q)\} \leq d \leq \min\{p(1-q), (1-p)q\}.$$

Some kinds of specification generates one- and two-parameter specific models as follows (only positive association models) :

	Independent Model	Dependent Model	The most Dependent Model
	$I(p, q) = M(p, q; 0)$	$M(p, q; d)$ (three parameter model)	$U^+(p, q) = M(p, q; p(1-q))$ $L^+(p, q) = M(p, q; (1-p)q)$
Symmetric ($q = p$)	$I_0(p) = I(p, p)$	$S^+(p, d) = M(p, p; d)$ ($0 \leq d \leq p(1-p)$)	$D^+(p) = M(p, p; p(1-p))$ (pos. diagonal model)
Skew-symmetric ($q = 1-p$)	$I_0^*(p) = I(p, 1-p)$	$T^+(p, d) = M(p, 1-p; d)$ ($0 \leq d \leq p^2 \wedge (1-p)^2$)	$U^+(p, 1-p)$ ($p \leq \frac{1}{2}$) $L^+(p, 1-p)$ ($p \geq \frac{1}{2}$)

The methods of graphical representation for $M(p, q; d)$ are introduced, which enable us to grasp the "position" of some specific models and their mixtures. Especially, this geometric approach is appropriate to explain the so called Simpson's paradox (a phenomenon of the reverse sign of association by collapsing two contingency tables).

Next, the mixture of the two Models $M(p, q; d)$ and $M(\tilde{p}, \tilde{q}; \tilde{d})$ with weights α and $1-\alpha$ respectively is expressed as $M(\bar{p}, \bar{q}; \bar{d} + e)$, where

$$\begin{aligned} \bar{p} &= \alpha p + (1-\alpha)\tilde{p}, \\ \bar{q} &= \alpha q + (1-\alpha)\tilde{q}, \\ \bar{d} &= \alpha d + (1-\alpha)\tilde{d}, \\ e &= \alpha(1-\alpha)(p-\tilde{p})(q-\tilde{q}), \end{aligned}$$

Various cases where α is estimable from the contingency table of the mixed model are illustrated.

Finally, using these mixed models, it can be revealed some "latent structure" hidden in contingency table. Some kinds of latent structure analysis are also illustrated for a series of real data.