モデルの分解を用いての 分割表データの解析

東京理科大学理工学部情報科学科 富澤 貞男

Analysis of Data of Contingency Tables Using the Decompositions for Models

Sadao TOMIZAWA

要 旨

本論文の目的は、分割表解析においてモデルの分解を考えることが、如何に重要で、また有効的であるかを実際に3種類のデータを解析することにより示すことである。扱われるデータのうちの2種類は、行と列が同じ分類の正方分割表データであり、対称性に関するモデルの分解が与えられ適用される。残りのデータは2次元分割表データであって、独立モデルの分解を用いて解析される。

1. は じ め に

たとえば、2次元分割表データを解析する際に、通常の統計家はまず、独立モデル(独立性の仮説検定)が成り立つかどうかをPearsonのカイ2乗または尤度比カイ2乗統計量に基づいて適合度を検定するであろう。このとき、適当な有意水準に対して、もし採択されたならば、このデータに対して独立モデルが成り立つという判断を通常の統計家は下すであろう。このとき注意すべきことは、当然のことであるが、これは単に適合度が良いということであって独立モデルが適切であることにはならないということである。独立モデルの適切性を調べる一つの方法としては、このモデルよりも更に制約条件を緩めたいいくつかの(特に2つの)モデルへ独立モデルを分解することが考えられる。

一般に、独立モデルに限らず、あるモデルがあったならばそのモデルの分解を考えることは極めて重要である。それは特に、次のような2つの場合が起こり得ることから理解できるであろう。即ち、「モデル H_A が成り立つための必要十分条件は、モデル H_B と H_C の両方が成り立つことである。但し、適合度カイ2乗検定におけるモデル H_A に対する自由度は、モデル H_B と H_C に対す

Key words: Conditional symmetry, Dumping severity data, Father-son occupational mobility data, Independence, Symmetry, Unaided vision data, Uniform association

る各自由度の和に等しい。」がいえるとき、適当な有意水準 α (例えば $\alpha=0.05$) で、(i) モデル H_A が棄却されたときに、分解されたモデルの一方の H_B は棄却されるが、他方の H_C は採択される場合と、(ii) モデル H_A が採択されても、 H_B が (H_A より条件が緩いのにもかかわらず) 棄却され、 H_C が採択される場合である。

本論文の目的は、具体的にモデルの分解を与えて上記 (i) と (ii) のような場合が実際に起こり得る 3 種類のデータをモデルの分解を用いて解析し、モデルの分解を考えることが如何に有効的で重要であるかを示すことである。ここに 3 種類のデータのうち 2 つは行と列が同じ分類からなる、しかも順序づけ可能なカテゴリをもつ正方分割表データであり、表 1 と 2 に与えられている。残りの 1 つは一般の 2 次元分割表データであって表 3 に与えられている。表 1 は Stuart (1953, 1955) により最初に解析された 1943 年から 1946 年に英国の王立軍需工場で働いていた年齢 30 歳から 39 歳までの 7,477 名の女性の左右裸眼視力データからなる分割表である。表 2 は英国のある男子とその父親との職業関係からなる分割表である。行と列変数はそれぞれ父親とその息子の職業であり、社会的地位の低い方から高い方へ 5 段階に分けられた順序づけ可能なカテゴリから成っている。表 1 と 2 の特徴は、観測値が対角セル付近に集中していることである。即ち、表 1 では左右視力の変わらない人が多いこと、表 2 では父親の社会的地位と息子の社会的地位の近い組が多いことである。このようなデータにおいては、もはや通常の独立性のモデルをあてはめても無意味であり、それに代わって対称性に関するモデルをあてはめることに興味がある。例えば、左眼が右眼より良い人は逆に右眼が左眼より良い人と確率的に同じといえるかどうかが大変、興味が注がれる。また、表 3 は Agresti (1984, p. 81) から直接引用したデータであって、十二指腸潰瘍の手術直後に現われる、無気力、ふさぎ込みなどの“ダンピング (dumping) 症候群”の程度と手術の程度との分類から成っており、手術法は、A (排膿法と迷走神経切断)、B (十二指腸 25% 切除と迷走神経切断)、C (十二指腸 50% 切除と迷走神経切断)、D (十二指腸 75% 切

表 1. 英国の女性の裸眼視力 (Stuart (1953) より引用)

左眼	良	い	やや良い	やや悪い	悪	い	計
右眼	(1)	(2)	(3)	(4)			
良	1520	266	124	66			1976
い	234	1512	432	78			2256
(1)	117	362	1772	205			2456
(2)	36	82	179	492			789
(3)							
(4)							
計	1907	2222	2507	841			7477

表 2. 英国の父親と息子の職業の社会的地位 (Agresti (1984, p. 206) より直接引用)

息子の地位	低い				高い	計
父親の地位	(1)	(2)	(3)	(4)	(5)	
低	50	45	8	18	8	129
い	28	174	84	154	55	495
(1)	11	78	110	223	96	518
(2)	14	150	185	714	447	1510
(3)	3	42	72	320	411	848
(4)						
(5)						
高						
い						
(1)						
(2)						
(3)						
(4)						
(5)						
計	106	489	459	1429	1017	3500

表 3. 十二指腸潰瘍の手術とダンピング症候群との分類データ
(Agresti (1984, p. 81) より直接引用)

手術法	ダンピング症候群			計
	な し	軽 い	中程度	
A	61	28	7	96
B	68	23	13	104
C	58	40	12	110
D	53	38	16	107
計	240	129	48	417

除) の 4 通りに分類されている。尚, 表 3 のデータは Grizzle et al. (1969) により最初に解析された $4 \times 3 \times 4$ 分割表の 1 つの周辺和の表であることを付記しておく。

本論文の 2 節では, 表 1, 2, 3 のデータに適用するための具体的モデルとそれらの分解を与える。3 節では, これらのモデルの分解を用いてデータを解析し, モデルの分解を考えることが如何に有効的であるかを示す。

2. モデルの分解

はじめに, 表 1 と 2 のデータを解析するために用いる対称性に関するモデルと分解を考える。一般の $r \times r$ 分割表を考え, (i, j) セル確率を P_{ij} とする。対称性に関する色々なモデルの始まりである対称モデルは次のように与えられる (Bowker (1948), Bishop et al. (1975))。

$$H_S: P_{ij} = \phi_{ij} \quad (i=1, \dots, r; j=1, \dots, r), \quad \text{但し} \quad \phi_{ij} = \phi_{ji}.$$

もし, 表 1 のデータに対してこのモデルが成り立つならば, ある女性の右視力の階級が i で左視力の階級が j (但し $1 \leq i < j \leq 4$) である確率と逆に右視力の階級が j で左視力の階級が i である確率とが等しいことを意味している。実際のデータに対して対称モデル H_S の制約条件は, かなり厳しくもう少し条件の緩いモデルとして, 条件付対称モデル (McCullagh (1978), Bishop et al. (pp. 285-6)) が次のように考えられている。

$$H_{CS}: P_{ij} = \begin{cases} \phi \phi_{ij} (i < j), \\ \phi_{ij} (i \geq j), \end{cases} \quad \text{但し} \quad \phi_{ij} = \phi_{ji}.$$

ここで行と列変数をそれぞれ X と Y によって表わすと H_{CS} は $Pr(X=i, Y=j | X < Y) = Pr(X=j, Y=i | X > Y) (i < j)$ なる関係が成り立つことと同値になるので条件付対称モデルと呼ばれている。ここに, H_{CS} の下で $\phi \geq 1$ は $Pr(X < Y) \geq Pr(X > Y)$ と同値であり, 更に $Pr(X \leq k) \geq Pr(Y \leq k) (k=1, \dots, r)$ とも同値である。明らかに対称モデル H_S の下では, $Pr(X < Y) = Pr(X > Y)$ と $Pr(X \leq k) = Pr(Y \leq k) (k=1, \dots, r)$ が成り立つ。更に H_{CS} より条件の緩いモデルとして対角パラメータ対称モデルが次のように考えられている (Goodman (1979 a))。

$$H_{DP}: P_{ij} = \begin{cases} \phi_{j-i} \phi_{ij} (i < j), \\ \phi_{ij} (i \geq j), \end{cases} \quad \text{但し} \quad \phi_{ij} = \phi_{ji}.$$

特に, $\phi_1 = \dots = \phi_{r-1}$ ならば H_{CS} と一致する。

ところで Caussinus (1965) は対称モデル H_S の分解を次のように与えた。

定理 1 対称モデル H_S が成り立つための必要十分条件は、準対称モデル H_Q と周辺同等モデル H_M の両方が成り立つことである。

ここに、準対称モデル H_Q (Caussinus (1965)) と周辺同等モデル H_M (Stuart (1955)) は次のように定義される。

$$\begin{aligned} H_Q: P_{ij} &= \alpha_i \beta_j \phi_{ij} \quad (i=1, \dots, r; j=1, \dots, r), \quad \text{但し} \quad \phi_{ij} = \phi_{ji}; \\ H_M: P_{i\cdot} &= P_{\cdot i} \quad (i=1, \dots, r), \quad \text{但し} \quad P_{i\cdot} = \sum_{k=1}^r P_{ik}, \quad P_{\cdot i} = \sum_{k=1}^r P_{ki}. \end{aligned}$$

特に H_Q において $\alpha_1 = \dots = \alpha_r$ 且つ $\beta_1 = \dots = \beta_r$ のときは対称モデル H_S に一致する。また、準対称モデル H_Q はオッズ比を用いて

$$H_Q: \Theta(s < t, u < v) = \Theta(u < v, s < t) \quad (1 \leq s < t \leq r, 1 \leq u < v \leq r),$$

のようにも表わせる。ここに $\Theta(s < t, u < v) = P_{su}P_{tv} / (P_{tu}P_{sv})$ によって定義される。

次に、周辺同等モデル H_M の代わりにモデル H_{M1} と H_{M2} を次のように考える (Tomizawa (1984, 1985 a, 1987 c) 参照)。

$$\begin{aligned} H_{M1}: P_{i\cdot}^+ &= P_{\cdot i}^- \quad (i=1, \dots, r-1), \quad \text{但し} \quad P_{i\cdot}^+ = \sum_{k=i+1}^r P_{ik}, \quad P_{\cdot i}^- = \sum_{k=i+1}^r P_{ki}; \\ H_{M2}: P_{i\cdot}^+ &= P_{\cdot i}^- \quad (i=2, \dots, r), \quad \text{但し} \quad P_{i\cdot}^+ = \sum_{k=1}^{i-1} P_{ik}, \quad P_{\cdot i}^- = \sum_{k=1}^{i-1} P_{ki}. \end{aligned}$$

ここに、 H_{M1} または H_{M2} の下で $Pr(X < Y) = Pr(X > Y)$ が成り立つ。もし、表 1 のデータに対して H_M が成り立つならば、ある女性の右視力の階級が i ($i=1, \dots, 4$) である確率と左視力の階級が i である確率とが等しいことを意味している。それに対して、 H_{M1} はある女性の良い方の視力が右眼でその階級が k ($k=1, 2, 3$) である確率とそれが左眼で階級が k である確率とが等しいことを意味しており、 H_{M2} は逆に悪い方の視力が右眼でその階級が l ($l=2, 3, 4$) である確率とそれが左眼で階級が l である確率とが等しいことを意味している。ここに、対称モデル H_S の別な分解が Tomizawa (1984, 1985 a) により次のように得られている。

定理 2 $l=1, 2$ に対して、対称モデル H_S が成り立つための必要十分条件は、 H_Q と H_{Ml} の両方が成り立つことである。

次に条件付対称モデル H_{CS} の分解を考える。このモデルの分解は既に Tomizawa (1984, 1985 a, 1987 d) で考えられているが、ここでは新しい分解を考える。そのために次のようなモデルを導入する。

$$H_{M3}: P_{(d)}^+ = \Delta P_{(d)}^- \quad (d=1, \dots, r-1),$$

但し、

$$P_{(d)}^+ = \sum_{k=1}^{r-d} P_{k, k+d}, \quad P_{(d)}^- = \sum_{k=1}^{r-d} P_{k+d, k}.$$

このとき、次の定理を得る。

定理 3 条件付対称モデル H_{CS} が成り立つための必要十分条件は、対角パラメータ対称モデル H_{DP} とモデル H_{M3} の両方が成り立つことである。

ここに、定理 1, 2, 3 においてモデル H_S, H_{CS} の適合度カイ 2 乗検定に対する自由度は、分解された各モデルに対する自由度の和に等しいことに注意する (自由度の値は略)。

その他の種々の対称性に関するモデルまたは分解については Agresti (1983 b), 広津 (1984) として参考文献にある Tomizawa の論文を参照せよ。

次に, 表 3 のデータを解析するために用いられる独立モデルの分解を考える. 一般に $r \times c$ 分割表で隣接した行と列からなる 2×2 表に対するオッズ比 $P_{ij}P_{i+1,j+1}/(P_{i+1,j}P_{i,j+1}) = \Theta(i < i+1, j < j+1)$ を特に Θ_{ij} と略記する. このとき通常の独立モデル, $P_{ij} = P_{i.}P_{.j}$ ($i=1, \dots, r; j=1, \dots, c$) は次のように表わせる.

$$H_I: \Theta_{ij} = 1 \quad (i=1, \dots, r-1; j=1, \dots, c-1).$$

このモデルは別名, null アソシエーションモデルと呼ばれている (Goodman (1979 b)). 更に条件を緩めたモデルとして

$$H_U: \Theta_{ij} = \theta \quad (i=1, \dots, r-1; j=1, \dots, c-1),$$

が考えられており, uniform アソシエーションモデルと呼ばれている (Goodman (1979 b)). 尚, このモデルをスコアを用いてもう少し一般的に扱ったモデルは Agresti (1983 a, 1984) を参照されたい. ところでこれらのモデルは, セル確率に対しての乗積モデルとして

$$\begin{aligned} H_I: P_{ij} &= \alpha_i \beta_j & (i=1, \dots, r; j=1, \dots, c); \\ H_U: P_{ij} &= \alpha_i \beta_j \theta^{ij} & (i=1, \dots, r; j=1, \dots, c), \end{aligned}$$

のようにも表わせる. もちろん適当な制約, 例えば $\alpha_1 = 1$ が必要であることに注意する.

さて, 独立モデル H_I の分解は色々と考えられるが, ここでは単純な次のモデルを導入する.

$$H_O: \frac{P_{11}}{P_{r1}} = \frac{P_{1c}}{P_{rc}}$$

このとき, 独立モデル H_I に対する 1 つの分解定理を次のように得る.

定理 4 $r \times c$ 分割表において, 独立モデル H_I が成り立つための必要十分条件は, uniform アソシエーションモデル H_U とモデル H_O の両方が成り立つことである.

表 4. 表 1 のデータに対する尤度比カイ 2 乗値 G^2

モデル	自由度	G^2 値
H_S	6	19.25**
H_Q	3	7.27
H_M	3	11.99**
H_{M1}	3	11.99**
H_{M2}	3	11.97**
H_{CS}	5	7.35
H_{DP}	3	0.50
H_{M3}	2	6.86*

** : 対応する自由度のカイ 2 乗検定で 1% 有意となる場合

* : 対応する自由度のカイ 2 乗検定で 1% 有意でないが 5% 有意となる場合

表 5. 表 2 のデータに対する尤度比カイ 2 乗値 G^2

モデル	自由度	G^2 値
H_S	10	37.46**
H_Q	6	4.66
H_M	4	32.80**
H_{M1}	4	32.69**
H_{M2}	4	34.40**
H_{CS}	9	10.35
H_{DP}	6	6.4
H_{M3}	3	3.90

** : 対応する自由度のカイ 2 乗検定で 1% 有意となる場合

3. データ解析

表 1, 2, 3 のデータを解析するために, モデル H の適合度を検定するための尤度比カイ 2 乗統計量の値を $G^2(H)$ によって記す. 各モデルを表 1 (2) のデータに適合させたときの尤度比カイ 2 乗値 G^2 が表 4 (5) に与えてある.

3.1 表 1 のデータ解析

通常の対称モデル H_S は適合度は極めて悪く, 左右視力は対称的とはいえない. また対称モデル H_S の分解されたモデルに対しては準対称モデル H_Q は非常に良いというほどではないが適度に良く適合しており, それ以外の周辺の対称性に関するモデル H_M, H_{M1}, H_{M2} は全て適合度は悪い. 今, 準対称モデル H_Q が成り立つことを仮定して対称モデル H_S が成り立つかどうかの検定, 即ち, 同じ仮定の下で H_M (H_{M1} または H_{M2}) が成り立つかどうかの検定に対して, $G^2(H_S) - G^2(H_Q) = 11.98$ (自由度は $6 - 3 = 3$) となり, この仮説は有意水準 1% で棄却される. 即ち, 周辺の対称性に関するモデルは適切でないことがわかる. ここに, モデル H_M, H_{M1}, H_{M2} の不適切性は理論期待度数の最尤推定値 (MLE), G^2 値を直接に計算することなしに得られていることに注意する. このことは H_S の分解のおかげである.

他方, 対称モデル H_S の拡張したモデルである条件付対称モデル H_{CS} は $G^2(H_{CS}) = 7.35$ (自由度は 5) で有意水準 5% で採択され適合度は良い. そして H_{CS} の下で比率パラメータ ϕ の MLE は $\hat{\phi} = 1.159$ となり 1 より大きいので左眼は右眼より悪いといえる. また, 条件付対称モデル H_{CS} をより拡張した対角パラメータ対称モデル H_{DP} も $G^2(H_{DP}) = 0.50$ (自由度は 3) となり有意水準 5% で採択され, 適合度は良い. H_{DP} の下で $\hat{\phi}_1 = 1.165, \hat{\phi}_2 = 1.015, \hat{\phi}_3 = 1.833$ となり全て 1 より大きくなっているので, やはり左眼は右眼より悪いことを知ることが出来る. 今, H_{DP} が成り立ったという仮定の下で, $\phi_1 = \phi_2 = \phi_3$ の仮説検定, 即ち, 同じ仮定の下で条件付対称モデル H_{CS} が成り立つかどうかの検定に対して, $G^2(H_{CS}) - G^2(H_{DP}) = 6.85$ (自由度は $5 - 3 = 2$) となり有意水準 5% で有意になる. よって, 表 1 のデータに対しては, 条件付対称モデル H_{CS} よりも, むしろ, 対角パラメータ対称モデル H_{DP} の方が好まれるといえる (Goodman (1979 a) 参照). 即ち, モデル H_{CS} は適合度は良くても表 1 のデータに対して適切なモデルであるとは言い切れないことが示された. このことは, 定理 3 の分解を用いると一層明確になる. 即ち, モデル H_{CS} は有意水準 5% で採択されたのに, H_{CS} の条件を緩めたモデル H_{M3} に対して $G^2(H_{M3}) = 6.86$ (自由度は 2) となり, 有意水準 5% で棄却されてしまうという現象が起こる. したがって, モデル H_{CS} は適合度は良くても適切なモデルとはいえない原因の一つとして, 分解されたモデル H_{M3} の適合度の悪さがあげられよう. 一般にモデル H_A が採択されても, H_A の条件を緩めたモデル H_B が棄却される (理論上は H_A が成り立てば H_B は必ず成り立つ. よって $G^2(H_A) > G^2(H_B)$ となる) 場合が起こり得るので, モデルの分解を考えることは極めて重要といえる.

3.2 表 2 のデータ解析

表 1 のときと同じように表 2 のデータに対しても対称モデル H_S は, $G^2(H_S) = 37.46$ (自由度は 10) となり有意水準 1% で棄却され適合度は悪い. 即ち, 父親とその息子における職業の社会的地位が, それぞれ, i 階級と j 階級 ($i < j$) である確率は, それが逆に j 階級と i 階級である確率に等しいとはいえない. これに対して, 対称モデル H_S の分解されたモデルの一つである準対称モ

デル H_Q は, $G^2(H_Q)=4.66$ (自由度は 6) となり有意水準 5% で採択され, 表 2 のデータに非常に良く適合している. その上, 表 1 に対する H_Q より, はるかに良く適合している. しかしながら, 対称モデル H_S のもう一つの分解されたモデルである周辺同等モデル H_M においては, もはや $G^2(H_M)=32.80$ (自由度は 4) となり有意水準 1% で棄却され適合度は極めて悪い. 以上のことから, 対称モデル H_S の適合の悪さは H_S の分解されたモデルである H_Q と H_M のうちで特に H_M の適合の悪さが影響していると考えられる. 更に, モデル H_Q は適合度は極めて良く, 採択されたので, H_Q が成り立つことを仮定してモデル H_S が成り立つかどうかの仮説検定, 即ち, 同じ仮定の下でモデル H_M が成り立つかどうかの検定を考えると, この検定は $G^2(H_S)-G^2(H_Q)$ の値に基づいて検定できて, 表 2 のデータに対して 32.80 (自由度は $10-6=4$) となり, この仮説は有意水準 1% で棄却される. よって, $G^2(H_M)$ の値を直接計算することなしに, モデル H_M は表 2 のデータに対して不適切であることを知ることができる. このことは正にモデル H_S の分解のおかげである. 以上のことは周辺同等モデル H_M に代えてモデル H_{M1} または H_{M2} を用いても言える (同様なので略).

ところで, 対称モデル H_S は適合度は悪かったが, その拡張である条件付対称モデル H_{CS} は良く適合しており, 比率パラメータ ϕ の MLE は $\hat{\phi}=1.260$ となり 1 より大きいので息子の職業の社会的地位は, その父親の社会的地位より高いことがわかる. また, 条件付対称モデル H_{CS} の一層の拡張である対角パラメータ対称モデル H_{DP} も適合度は良く, 比率パラメータの MLE は, $\hat{\phi}_1=1.31$, $\hat{\phi}_2=1.11$, $\hat{\phi}_3=1.30$, $\hat{\phi}_4=2.67$ となり全て 1 より大きくなっている. 今, モデル H_{DP} が成り立つことを仮定して, $\phi_1=\phi_2=\phi_3=\phi_4$ の仮説検定, 即ち, 同じ仮定の下で条件付対称モデル H_{CS} が成り立つかどうかの検定に対して, $G^2(H_{CS})-G^2(H_{DP})=3.95$ (自由度は $9-6=3$) となって 5% 有意でない (表 1 のデータに対しては有意であった). よって表 1 のデータとは対照的に表 2 のデータに対しては, 条件付対称モデル H_{CS} は対角パラメータ対称モデル H_{DP} より好まれる.

更に, 条件付対称モデル H_{CS} の分解されたモデルのもう一つである H_{M3} に対しては, $G^2(H_{M3})=3.90$ (自由度は 3) となり有意水準 5% で採択され, 表 1 の場合とは対照的に表 2 のデータに対しては適合度は良い. そこで, モデル H_{M3} が成り立つことを仮定して条件付対称モデル H_{CS} が成り立つかどうかの検定を考えると, $G^2(H_{CS})-G^2(H_{M3})=6.45$ (自由度は $9-3=6$) となって 5% 有意でない. したがって, 条件付対称モデル H_{CS} の方がモデル H_{M3} より好まれる.

尚, 表 2 のデータに対する別な解析については, たとえば Bishop et al. (1975), Goodman (1981), Agresti (1984), 富澤 (1987 d) を参照せよ.

3.3 表 3 のデータ解析

表 3 のデータに対して, 独立モデル H_I は $G^2(H_I)=10.88$ (自由度は 6) (Agresti (1984, p. 81) より) となり有意水準 5% で採択され適合度は良い. しかし, 1 節でも述べたように, このことは独立モデル H_I が表 3 のデータに対して適切であるということではない. 実際, このデータに対して, 独立モデル H_I は適合度は良くても適切でない. 以下, このことを独立モデル H_I の分解を用いて示そう.

今, 独立モデル H_I よりも制約条件の緩い uniform アソシエーションモデル H_U は, $G^2(H_U)=4.59$ (自由度は 5) (Agresti (1984, p. 81) より) となり有意水準 5% で採択され適合度は良い. そこで, モデル H_U が成り立つことを仮定して, モデル H_I が成り立つかどうかの検定を行なうと $G^2(H_I)-G^2(H_U)=6.29$ (自由度は $6-5=1$) となるから, この仮説は有意水準 5% で棄却される.

よって、モデル H_1 は適合度は良くても表 3 のデータに対しては適切なモデルとは言い切れない。このことは、正に定理 4 の独立モデル H_1 に対する分解を考えれば、一層明確となる。即ち、モデル H_1 は有意水準 5% で採択されたのにもかかわらず、モデル H_1 の条件を緩めたモデル H_0 に対して $G^2(H_0)=4.17$ (自由度は 1) となり、有意水準 5% で棄却されるという現象が起こっている。よって、モデル H_1 は適合度は良くても適切なモデルとは言い切れない原因の一つとして、分解されたモデル H_0 の適合度の悪さがあげられよう。

以上のことから理解できるように、モデルの分解を考えることは極めて重要である。

4. お わ り に

3 節のデータ解析で示したように、分割表データにモデルを適用する際には、1 節で述べた (i) と (ii) の場合が起こり得るので意味のあるモデル分解を考えて、それを用いて種々の検定を行なうことは極めて重要である。

表 1 と 2 のような行と列が同じ分類からなる正方分割表データでは、対角セル上に観測値が集中する傾向にあるので、独立モデルは明らかに不適切であるが、表 3 のデータのような通常の 2 次元分割表データに対しては、通常の統計家は、まず独立モデルの適合度をカイ 2 乗統計量に基づいて調べるであろう。その際に、この論文でも示したように、独立モデルの適合度は非常に良いが、適切でない場合が当然起こり得るので注意を要する。そして、何故、独立モデルの適合度は良くても不適切なのかを知る手がかりの一つとして、独立モデルの分解を考えることが上げられる。とりわけ、2 つのモデルへうまく分解を考えると一方のモデルは非常に適合度は良いが、もう一方のモデルは非常に適合度が悪いという結果が得られるように思える。

独立モデルに限らず、分割表データをモデルを使って解析する際には、意味のあるモデルの分解を考えて、適用すると有効的な結果が得られるであろう。

謝 辞

本稿は日本数学会 1987 年度秋季総合分科会統計数学分科会で特別講演したときの内容の一部分をもとに新しい結果を加えてまとめたものです。また本稿をまとめるにあたり、審査員および編集委員会の方々、特に編集理事の鈴木義一郎博士には適切かつ示唆に富むご指摘を頂きました。ここに記して謝意を表します。さらに種々の面で御指導下さいました東京理科大学の國澤清典教授、牧野都治教授、清水邦夫博士には心からの感謝の意を表します。

参 考 文 献

- Agresti, A. (1983 a): A survey of strategies for modeling cross-classifications having ordinal variables. *J. Amer. Statist. Assoc.* 78, 184-198.
- Agresti, A. (1983 b): A simple diagonals-parameter symmetry and quasi-symmetry model. *Statistics and Probability Letters* 1, 313-316.
- Agresti, A. (1984): *Analysis of ordinal categorical data*. John Wiley & Sons, New York.
- Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975): *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass: MIT Press.
- Bowker, A.H. (1948): A test for symmetry in contingency tables. *J. Amer. Statist. Assoc.* 43, 572-574.
- Caussinus, H. (1965): Contribution à l'analyse statistique des tableaux de corrélation. *Ann. Fac. Sci. Univ. Toulouse* 29, 77-182.

- Goodman, L.A. (1979 a): Multiplicative models for square contingency tables with ordered categories. *Biometrika* 66, 413-418.
- Goodman, L.A. (1979 b): Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* 74, 537-552.
- Goodman, L.A. (1981): Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika* 68, 347-355.
- Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969): Analysis of categorical data by linear models. *Biometrics* 25, 489-504.
- 広津千尋 (1984): 順序カテゴリの分割表における準対称モデルについて. 応用統計学 13, 1-9.
- McCullagh, P. (1978): A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika* 65, 413-418.
- Stuart, A. (1953): The estimation and comparison of strengths of association in contingency tables. *Biometrika* 40, 105-110.
- Stuart, A. (1955): A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* 42, 412-416.
- Tomizawa, S. (1984): Three kinds of decompositions for the conditional symmetry model in a square contingency table. *J. Japan Statist. Soc.* 14, 35-42.
- Tomizawa, S. (1985 a): Analysis of data in square contingency tables with ordered categories using the conditional symmetry model and its decomposed models. *Environmental Health Perspectives* 63, 235-239.
- Tomizawa, S. (1985 b): Decompositions for odds-symmetry models in a square contingency table with ordered categories. *J. Japan Statist. Soc.* 15, 151-159.
- Tomizawa, S. (1987 a): Diagonal weighted marginal homogeneity models and decompositions for linear diagonals-parameter symmetry model. *Commun. Statist.-Theory Meth.* 16, 477-488.
- Tomizawa, S. (1987 b): Decompositions for 2-ratios-parameter symmetry model in square contingency tables with ordered categories, *Biometrical J.* 29, 45-55.
- Tomizawa, S. (1987 c): Analysis of square contingency tables with ordered categories by a new method of applying the 1-weight modified marginal homogeneity models. *Biometrical J.* 29, 667-674.
- 富澤貞男(1987 d): 分割表解析における対称性に関するモデルとそれらの分解. 日本数学会 1987 年度秋季総合分科会統計数学分科会講演アブストラクト, 123-146.

(1987 年 12 月 15 日受付 1988 年 2 月 2 日最終修正)

著者連絡先: 〒 278 千葉県野田市山崎東亀山 2641 東京理科大学理工学部情報科学科