

機関番号：32660

研究種目：若手研究（B）

研究期間：平成21年度 ～ 平成22年度

課題番号：21740084

研究課題名（和文）

分割表解析における対称性のモデル構築とその直交分解の研究

研究課題名（英文）

Modeling and orthogonal decompositions of symmetry for contingency tables

研究代表者

田畑 耕治 (TAHATA KOUJI)

東京理科大学・理工学部情報科学科・助教

研究者番号：30453814

研究成果の概要（和文）：本研究では、正方分割表データの解析に用いる新しい対称性のモデルを導入した。また、あるモデルを制約の弱い複数のモデルへ分解する方法を与えた。これらの成果は、正方分割表データを解析するうえで大変有用であることを実際のデータを用いて示した。

研究成果の概要（英文）：In this research, for the analysis of square contingency tables, we proposed some symmetry models. Also, we gave the orthogonal decompositions of model into some models that have weaker restrictions than the model. These results will work well for the analysis of square contingency tables.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
平成21年度	800,000	240,000	1040,000
平成22年度	700,000	210,000	910,000
年度			
年度			
年度			
総計	1500,000	450,000	1950,000

研究分野：数物系科学

科研費の分科・細目：数学・数学一般（含確率論・統計数学）

キーワード：統計数学、分割表解析

1. 研究開始当初の背景

現在、カテゴリカルデータ解析の方法は、医学、薬学、政治哲学など幅広い分野で利用されている。カテゴリカルデータ解析の目的は、分類間にどのような相互関連性が存在しているのかを得られた観測度数から高い信頼度で推測し、データに対して有用な解釈を与えることである。本研究の目的は、新しい分割表解析の手法を提案し、実社会における問題解決に貢献することである。

本研究は、行と列が同じ分類からなる正方分割表の解析を扱う。正方分割表の解析には、対称性に関する統計モデルを用いることが多い。対称性のモデルとしては、Bowker

(1948) の対称モデルなど種々のモデルが導入されている。一方、非対称性のモデルも多く導入されている。また、Caussinus (1965) は「対称モデルが成り立つための必要十分条件は、準対称モデルと周辺同等モデルの両方が成り立つことである」という対称モデルの分解定理を与えた。さらに、Tomizawa (1994) は対称モデルからの隔たりを測る尺度を開発した。この他にも多くの研究者によって種々の統計解析手法が導入されている。しかし、分割表解析の手法は未だ完全に整備されておらず、解決すべき問題が数多く存在する。したがって、それらの問題に対する解決策の提案が重要な研究課題である。

2. 研究の目的

(1) 検定統計量の直交分解に関する情報幾何学的アプローチの研究：近年、統計学において重要な推定や検定、その他の手法に関して情報幾何学を用いた新しい研究が行われている (Amari, 1990)。一方、これまで対称性のモデルに関する分解の研究は数多くおこなわれているが、情報幾何学を応用した研究はまだない。この研究の目的は、情報幾何学の理論をモデル分解に応用することにより、俯瞰的に各モデルの関連性を把握することが目的であり、この結果は今後のモデルの直交分解を理解するうえで大いに役立つことが期待される。

(2) モデル分解におけるグラフィカルモデリングの応用：多元分割表において対数線形モデルを考えると、階層モデルはあまりにおおきなクラスであるため、グラフィカルモデルや分解可能モデルの利用が有効である (宮川, 1997)。これらは独立性 (条件付き独立性) のモデルを考える上で、絶大な効果を発揮する。そこでこれらの理論を対称性のモデルへ応用することを目標とした研究を行い、モデル選択への応用や、モデルの分解定理のグラフ理論的解釈など大いに役立つことが期待される。

(3) 周辺同等モデルからのベクトル型隔たり尺度の開発研究：周辺同等モデルは、薬効の同等性に関する解析などに用いられる。実際のデータ解析において、周辺同等モデルが成り立たない場合、我々は周辺分布がどのような非対称構造になっているのか、また周辺同等性からの隔たりがどの程度なのかを測ることに興味がある。本研究では、新しい周辺分布の非対称構造を示すモデルの構築、及び、解釈が明確で有用な尺度の開発を行う。

(4) 一般的な対数線形モデルの分解の研究：モデルの分解定理は、実際のデータ解析において、モデルが成り立たない原因を考察するのに役立ち、種々のモデルに対する分解定理が導出されている。しかし、これまでに導入されてきた分解定理は、全て個別に開発されているのが現状である。そこで、本研究では一般的な対数線形モデルに対する分解の研究に取り組み、一般化された分解定理を導出する。

3. 研究の方法

(1) 検定統計量の直交分解に関する情報幾何学的アプローチの研究：情報幾何学の基礎と一般化線形モデルの幾何学等について Amari and Nagaoka (2008)、Amari (1990)、

竹内ら (2003)、甘利ら (2000)、Kass and Vos (1997) などを調べる。そこで得た方法論を検定統計量の直交分解に応用し、情報幾何学における解釈を考察する。そして、検定統計量の直交分解における情報幾何学的見方の有用性を考察し、その結果が他の研究に応用可能か熟慮する。

(2) モデル分解におけるグラフィカルモデリングの応用：グラフィカルモデルの基礎理論および多元分割表における種々の独立性のモデルについて宮川 (1997)、Bishop et al (1975) や Agresti (2002) を調べる。また、Bowker (1948)、Caussinus (1965)、McCullagh (1978)、Goodman (1979) などを調べ種々の対称性のモデルへグラフィカル対数線形モデリングが応用可能かどうかを考察し、独自の理論を開発する。

(3) 周辺同等モデルからのベクトル型隔たり尺度の開発研究：周辺同等モデルからの隔たり尺度は、これまでにいくつか導入されている。そこで、これらの尺度を改良し、より解釈が容易で有用な尺度の開発・検討を行う。さらに実際のデータへ応用し、従来の尺度との比較を行う。周辺同等モデルからの隔たりを測る尺度に関して、Tomizawa (1995)、Tomizawa and Makii (2001)、Tomizawa, Miyamoto and Ashihara (2003) を調べる。過去に導入されている尺度は、解釈が難解であるという問題がある。そこで解釈が容易なベクトル型の尺度としてどのようにしたらよいかを考察する。

(4) 一般的な対数線形モデルの分解の研究：すでに提案されている分解定理の数理的性質を調べ、その本質的な共通点を見つけることにより一般的な分解定理を導出する。モデルの分解に関しては、個々のモデルに対して個別に導出されている。例えば、Caussinus (1965)、Tomizawa (1991) などがある。これらの論文をよく吟味し、分解定理の核となる共通部分を見つけ出し一般化された分解定理を導出する。一般的な分解定理が得られた場合、今まで個々に対応していた解析を一度にまとめた統合的かつ詳細な解析が可能になるため、その応用方法に関しても考察する。

4. 研究成果

下記の (1) から (8) の各項目は主な発表論文等の雑誌論文に記載されている①から⑧に対応しています：

(1) 対称 (S) モデル、点対称 (PS) モデル、二重対称 (DS) モデルに対して、各々のモデルのもとでの最尤推定量 (MLE) のもつリスク (期待平均二乗誤差) を導出

した。これらのモデルの間には、DS \Rightarrow S、DS \Rightarrow PS という関係が成り立つ。このような場合には、DS モデルが成り立つという条件の下で、DS モデルにおける MLE のリスクが S モデルと PS モデルにおける MLE のリスクより小さくなることを示した。

- (2) 二変量正規分布に従う変数を任意のカットポイントを用いて離散化したときに得られるような分割表データに対してよく適合すると思われるモデルを構築した。また、このモデルを用いて二重対称モデルの分解定理を導出し、さらにそれらのモデルの適合度検定統計量の直交分解を与えた。表 1 に示すデータに対して、提案手法を用いてこれまでになかった新しい結果と解釈が得られた。

表 1. デンマークの健康調査

	1965			計
	1962	A	B	
A	168	51	9	228
B	42	73	23	138
C	5	17	23	45
計	215	141	55	411

A:良好、B:普通、C:悪い

- (3) 多元分割表解析において、周辺点対称モデルの適合度検定をおこなうために必要となる検定統計量を導出した。この統計量は、コンピュータを用いた繰り返し計算を必要としないため、従来法に比べて容易に利用出来る。また、繰り返し計算に比べて早い時間で計算可能というメリットもある。
- (4) プロ野球の勝敗表や、5種類の日本酒を2種類ずつ比較して得られるデータを一対比較データという。そのようなデータの解析に、Bradley-Terry (BT) モデルが良く用いられる。本論文では BT モデルが成り立たない場合に、そのモデルからの隔たりを測る尺度を導入した。この尺度は複数の分割表を比較するのに有効である。また勝敗表などの解析において、どの3チーム間で三すくみが起きているかを推測するのにも役立つ。
- (5) 分割表解析において、独立モデルの分解定理が導出されている。正方分割表解析において、対角パラメータ対称 (DPS) モデルが導入されており、このモデルは

独立モデルと関連がある。このことから、独立モデルの分解定理を応用することにより、DPS モデルの分解定理を導出した。さらに DPS モデルの適合度検定統計量の直交分解を導出した。

- (6) 対称モデルは制約の強いモデルであり、データに適合することは多くない。したがって、その制約を緩めた新しいモデルを導入した。また、対称性からの最大の隔たりを「完全上側非対称」、「完全下側非対称」と二種類定義し、そのどちらの方向に向かって対称性からの隔たりがあるのかを測ることの出来る尺度を導入した。表 2 は完全上側非対称 (左目より右目の方が良い傾向) に向かって、表 3 は逆に完全下側非対称 (右目より左目の方が良い傾向) に向かって対称性から隔たっていることが推測できた。

表 2. イギリス人女性の視力データ

	左目			
	右目	良い	やや良い	やや悪い
良い	1520	266	124	66
やや良い	234	1512	432	78
やや悪い	117	362	1772	205
悪い	36	82	179	492

表 3. 日本人大学生の視力データ

	左目			
	右目	良い	やや良い	やや悪い
良い	1291	130	40	22
やや良い	149	221	114	23
やや悪い	64	124	660	185
悪い	20	25	249	1429

- (7) 多元分割表において、新たな非対称性のモデルを導入した。そしてそのモデルを用いて対称モデルの分解定理および検定統計量の直交性を導出した。
- (8) 連続型分布である二変量正規分布に従う確率変数をいくつかのカットポイントで離散化して得られる分割表データに対して、うまく適合すると思われるモデルを構築した。このモデルのパラメータを推定することにより今まで得られていない結果と解釈が得られた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計8件)

- ① Kouji Tahata, Takahiro Inoue, Sadao Tomizawa, Risk of estimators for point symmetry and double symmetry models for square contingency tables, *International Journal of Pure and Applied Mathematics*, 査読有, Vol. 66, 2011, 25-32.
- ② Kouji Tahata, Sadao Tomizawa, Double linear diagonals-parameter symmetry and decomposition of double symmetry for square tables, *Statistical Methods and Applications*, 査読有, Vol. 19, 2010, 307-318.
- ③ Kouji Tahata, Hiroki Tokuno, Sadao Tomizawa, On test of marginal point-symmetry in multi-way tables, *Advances and Applications in Statistical Sciences*, 査読有, Vol. 3, 2010, 1-12.
- ④ Kouji Tahata, Akihito Tomita, Sadao Tomizawa, Measures of departure from Bradley-Terry model for square tables: application to baseball, *International Journal of Sports Science and Engineering*, 査読有, Vol. 4, 2010, 195-208.
- ⑤ Kouji Tahata, Sadao Tomizawa, Decomposition of diagonals-parameter symmetry model and orthogonality of its test statistic for square contingency tables with ordered categories, *Journal of Statistical Theory and Applications*, 査読有, Vol. 8, 2009, 406-415.
- ⑥ Kouji Tahata, Kouji Yamamoto, Noriyuki Nagatani, Sadao Tomizawa, A measure of departure from average symmetry for square contingency tables with ordered categories, *Austrian Journal of Statistics*, 査読有, Vol. 38, 2009, 101-108.
- ⑦ Kouji Tahata, Sadao Tomizawa, Decomposition of symmetry using two ratios-parameter symmetry model and orthogonality for square contingency tables, *Journal of Statistics: Advances in Theory and Applications*, 査読有, Vol. 1, 2009, 19-33.
- ⑧ Kouji Tahata, Kouji Yamamoto, Sadao Tomizawa, Normal distribution type symmetry model for square

contingency tables with ordered categories, *Open Statistics and Probability Journal*, 査読有, Vol. 1, 2009, 32-37.

[学会発表] (計8件)

- ① 田畑耕治、富澤貞男、正方分割表における二重対称性の分解について、日本数学会、2011年3月21日、早稲田大学。
- ② 田畑耕治、富澤貞男、順序カテゴリをもつ正方分割表における対称性の分解について、日本数学会、2010年9月24日、名古屋大学。
- ③ 田畑耕治、富田聡人、富澤貞男、分割表における準対称性からの隔たりを測る距離尺度、統計関連学会連合、2010年9月8日、早稲田大学。
- ④ 田畑耕治、富田聡人、富澤貞男、Bradley-Terry モデルからの隔たりを測る距離尺度、応用統計学会、2010年5月20日、統計数理研究所。
- ⑤ 田畑耕治、対称性と非対称性のモデルに基づく分割表解析、日本数学会、2009年9月26日、大阪大学。
- ⑥ 田畑耕治、富澤貞男、多元分割表における点対称モデルの分解、統計関連学会連合、2009年9月7日、同志社大学。
- ⑦ 田畑耕治、山田章史、山本紘司、富澤貞男、正方分割表における非対称性を測る尺度とその応用、応用統計学会、2009年8月1日、統計数理研究所。
- ⑧ 田畑耕治、山本紘司、富澤貞男、分割表における点対称性からの隔たりを測る尺度、日本計算機統計学会、2009年5月17日、福岡女子大学。

[その他]

ホームページ等

http://www.sut.ac.jp/fac_grad/p/index.php?A15578

6. 研究組織

(1) 研究代表者

田畑 耕治 (TAHATA KOUJI)
東京理科大学・理工学部・助教
研究者番号：30453814

(2) 研究分担者

なし

(3) 連携研究者

なし