

令和元年6月18日現在

機関番号：32665

研究種目：若手研究(B)

研究期間：2017～2018

課題番号：17K14240

研究課題名(和文) カテゴリカルデータ解析における潜在分布の推定に関する研究

研究課題名(英文) Research on estimation of underlying distribution in categorical data analysis

研究代表者

生亀 清貴 (Iki, Kiyotaka)

日本大学・経済学部・講師

研究者番号：30711593

交付決定額(研究期間全体)：(直接経費) 1,200,000円

研究成果の概要(和文)：潜在分布として、二変量対数正規分布が想定されるときに、当てはまりが良いとされるモデルの提案を行った。また、提案モデルを用いて対称モデルの分解定理を示した。さらに提案モデルがどのような条件下で当てはまりが良いかを、シミュレーションを通じて示すことができた。他にも、直接的な研究成果と呼ぶことはできないが、当該研究に取り組む過程で計7編の論文が査読付にジャーナルに採択・掲載されることとなった。

研究成果の学術的意義や社会的意義

連続型の変数を、いくつかの切断点を設けて離散型の変数に変換し解析を行うことは、医学や経済学など様々な分野で行われているが、潜在分布に焦点を当てた研究はこれまで数少なかった。本研究で得られた成果をそれらの分野のデータに対して適用することで、これまでより詳細な解析が可能となる。

研究成果の概要(英文)：When bivariate log-normal distribution is assumed as the underlying distribution, we proposed a model that fits well. We also show the decomposition theorem of the symmetric model using the proposed model. Furthermore, it was possible to show through simulation that under what conditions the proposed model is good. Other than this, although it can not be called direct research results, a total of seven papers have been accepted and published in peer-reviewed journals in the process of working on the research.

研究分野：数理統計学

キーワード：正方分割表 潜在分布 対称性

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

分割表の解析における関心の一つは、母集団における未知の確率に対して、行変数と列変数が独立かどうか、あるいは二つの変数間にどのような関連性があるかどうかを適合度検定によって調べることである。一方、行と列が同じ分類からなる正方分割表の解析においては、分類間の関連性が極めて強い場合が多く、統計的独立性は成り立たないことが多い。したがって正方分割表の解析においては、統計的独立性に代わって分類間の対称性や非対称性に関心が移る。たとえば左右の裸眼視力を四つのカテゴリに分類し集約した正方分割表のように、行変数と列変数が元々連続量で、適当なカットポイントを用いて作成された分割表データの解析を考える。これまで分割表の独立性や対称性に関する研究は数多く行われているが、裸眼視力のデータのように背後に連続分布が存在するとき、その分布の特徴と分割表解析を関連づけた研究は数少ない状況にある。したがってこれまでの研究成果をさらに発展させ、分割表だけでなく背後の分布に対しても解釈の得られる新たな解析手法を提案することは必須である。また、データ変換を行う際の適切なカットポイントについての研究も不十分である。医学、社会学、スポーツ科学などの諸分野で連続量のデータを離散型に変換することは、昔から数多く行われてきているが、そのグループ化の基準に関しては、医学的な根拠があるものもあれば、経験的あるいは解析者の独断によるものも多い。したがって、解析者本人が適切なカットポイントを探したいとき、分割表解析を通じてカットポイントの候補を提示する手法の提案が期待される。

2. 研究の目的

(1) 分割表データに対しての解析を通じて、背後の分布にどのような構造が成り立つかを調べ、さらに背後の分布のもつパラメータを推定することを目指す。様々な条件下でシミュレーションを行い、各条件下でどの程度近似が成り立つかを調べる。

(2) 連続型のデータをいくつかのグループに分類することを考えるとき、分割表解析の観点から適切なカットポイントを示すことも試みる。カットポイントの個数や位置に関して、適合度検定の当てはまりの良さを調べることによって候補点を探る。

3. 研究の方法

(1) モデルの提案を考える上で、これまでにどのようなモデルが提案されてきたかを調べ整理する。モデルの有用性を示す手段として、どのような方法が採られているかも同時に調べる。また、日本数学会、統計関連学会連合大会などで関連する研究分野の講演を聴講し、最近の話題を調査する。さらに、連続変数を離散変数に変換する場合に、どのような手法が存在するかを調べる。統計学の分野だけでなく数学全般から幅広く調査し、解析手法に応用できそうな方法が存在するかを精査する。調べた結果を踏まえて、分割表への応用を考える。実際の現場で連続型のデータを離散型に変換する際、どのようなプロセスを経て行われているのかを製薬会社やCRO(受託臨床試験実施機関)などで働く社会人に直接話を聞く。離散データが得られたとき、潜在分布がどのような形状をしているかを、統計的に推測する方法を調べる。

(2) これまでの研究で得られた知見を総合して、分割表解析における新たなモデルの提案を行う。離散的なセル確率に対する解釈だけでなく、背後の連続分布に関しても、その分布がどのような性質を有するのか解釈できるようなモデルを導入する、さらにデータ変換の際にも、統計的に望ましい性質を満たすカットポイントの候補点を提示する方法を示す。これまでの研究成果を活かして、今までにない解析手法を提案する。企業などの現場で得られた実データをもとに、提案した手法の有用性を調べる。分布の種類、サンプル数、カテゴリ数などの条件をさまざまに設定することで種々の分割表データを作成し、提案モデルの当てはまりの良さや背後の分布の関係を調べ、様々な角度から研究成果を考察する。思いどおりの結果が得られない場合、モデルの修正や人工データ作成条件の変更を行い、再度結果を考察する。

4. 研究成果

(1) Iki (2018)では正方分割表を45度回転させたダイヤモンド型の表に対して、共分散が0であるという構造を定義し、定理「ダイヤモンドモデルが成り立つための必要十分条件は、一様連関ダイヤモンドモデルおよび共分散が0であるという構造の両方が成り立つことである」を与えた。

(2) Iki and Tomizawa (2018)では正方分割表に対して対称モデルからの隔たりを測る尺度を提案した。先行研究と提案尺度の異なる点は、主対角セルの情報を尺度に反映させることが可能な点である。提案尺度に対してデルタ法を用いて正規近似を行い、近似の信頼区間を導出した。さらに人工データを用いて先行研究との尺度のふるまいの比較を行った。

(3) Iki and Tomizawa (2018)では正方分割表の潜在分布として、2変量対数正規分布が想定されるときに当てはまりの良いモデルを提案した。さらに提案モデルを用いて対称モデルの分解定理を与えた。対数正規分布の乱数を用いたシミュレーションにより、提案モデルがどのような条件下で当てはまりが良いかの検証を行った。

(4) Iki, Sato and Tomizawa (2018)では二元分割表に対して儉約な独立モデルを定義し、定理「儉約な独立モデルが成り立つための必要十分条件は、儉約な線形線形連関モデルと相関係数 (Pearson や Kendall のタウ, Spearman のロー) が0という構造の両方が成り立つことである」を与えた。さらに定理「儉約な独立モデルに対する尤度比カイ二乗統計量が、儉約な線形線形連関モデルに対する統計量および相関係数が0の構造に対する統計量の和に漸近的に同等である」も与えた。

(5) Iki, Okada and Tomizawa (2018)では順序カテゴリ正方分割表において、分割表の背後に潜在的な等分散性を仮定しない2変量 t 分布が想定される場合に当てはまりの良いモデルを提案した。2変量 t 分布に従う乱数を発生させ、カットポイントを用いて分割表の作成を行い、様々なパラメータ設定の下でモデルの当てはまりをみるシミュレーションにより、提案モデルの有用性を示した。

(6) Iki and Tomizawa (2018)では、連続型の確率変数に対して、その確率密度関数の二重対称性、準二重対称性、周辺二重対称性を定義し、定理「確率密度関数に関して二重対称性が成り立つための必要十分条件は、準二重対称性および周辺二重対称性の両方が成り立つことである」を与えた。さらに定理を多変数の確率密度関数に拡張し、具体的な分布 (正規分布やベータ分布) に対して定理を適用し、各二重対称性をパラメータ制約にて表すことに成功した。

(7) Iki and Tomizawa (2017)では、周辺同等モデルからの隔たりを測る尺度 (Iki, Tahata and Tomizawa, 2011) を改良し、新しい尺度を提案した。モデルからの隔たりを測る尺度は未知であり、観測度数から尺度を推定する必要がある。新尺度ではこれまでより小サンプルの分割表データにおいて、より精度の良い推定値を得ることが可能となった。シミュレーションにより実際に推定精度が向上していることを確認し、有用性を示した。

5. 主な発表論文等

[雑誌論文](計7件)

Iki, K. (2018). Decomposition of diamond model for square contingency tables with ordered categories. *SUT Journal of Mathematics*, 54, 99-107. (査読有)

Iki, K. and Tomizawa, S. (2018). Measure of departure from symmetry using cumulative probability for square contingency tables with ordered categories. *Advances and Applications in Statistics*, 52, 401-412. (査読有)

Iki, K. and Tomizawa, S. (2018). Log-normal distribution type symmetry model for square contingency tables with ordered categories. *Austrian Journal of Statistics*, 47, 39-48. (査読有)

Iki, K., Sato, S. and Tomizawa, S. (2018). Decomposition of Parsimonious Independence Model Using Pearson, Kendall and Spearman's Correlations for Two-Way Contingency Tables. *International Journal of Statistics and Probability*, 7, 105-111. (査読有)

Iki, K., Okada, M. and Tomizawa, S. (2018). An extended bivariate t-distribution type symmetry model for square contingency tables. *Open Journal of Statistics*, 8, 249-257. (査読有)

Iki, K. and Tomizawa, S. (2018). Double symmetric multivariate density function and its decomposition. *Journal of Statistics Applications and Probability*, 7, 1-8. (査読有)

Iki, K. and Tomizawa, S. (2017). Improved estimator of measure for marginal homogeneity using marginal odds in square contingency tables. *Journal of Advanced Statistics*, 2, 71-77. (査読有)

[学会発表](計13件)

生亀清貴. (2019年3月). 正方分割表における潜在分布に基づく対称性のモデル. 日本数学会, 東京工業大学(特別講演).

吉本拓矢, 田畑耕治, 生亀清貴, 富澤貞男. (2019年3月). 多元分割表におけるモーメント対称モデルと周辺対称モデルの分解. 日本数学会, 東京工業大学.

篠田覚, 田畑耕治, 生亀清貴, 富澤貞男. (2019年3月). 多元分割表における補対数対数変換に基づく周辺非同等性について. 日本数学会, 東京工業大学.

齋藤健, 生亀清貴, 石井晶, 富澤貞男. (2019年3月). $R \times C$ 分割表における周辺点対称モデルに対する尺度, ポスターセッション. 日本統計学会, 日本大学.

吉本拓矢, 田畑耕治, 生亀清貴, 富澤貞男. (2018年3月). Covariance symmetry model and decomposition of symmetry for square contingency table, ポスターセッション. 応用統計学会, 統計数理研究所.

竹内貴志, 安藤宗司, 生亀清貴, 富澤貞男. (2018年3月). 正方分割表における累積周辺確率を用いた周辺同等モデルからの隔たりを測るベクトル尺度, ポスターセッション. 応用統計学会, 統計数理研究所.

吉本拓矢, 田畑耕治, 生亀清貴, 富澤貞男. (2018年3月). 正方分割表における共分散対称モデルと対称モデルの分解. 日本数学会, 東京大学.

池澤友哉, 生亀清貴, 山本紘司, 富澤貞男. (2018年3月). 正方分割表における併合した 3×3 表を用いた点対称性からの隔たりを測る尺度. 日本数学会, 東京大学.

佐藤駿, 生亀清貴, 富澤貞男. (2018年3月). 二元分割表における既知のスコアを用いた儉約な独立モデルの直交分解, ポスターセッション. 日本統計学会, 早稲田大学.

生亀清貴. (2017年9月). 順序カテゴリ正方分割表におけるダイヤモンドモデルの分解. 日本数学会, 山形大学.

生亀清貴. (2017年9月). 二元分割表における周辺点対称性に関する方向付き尺度. 統計関連学会連合大会, 南山大学.

Iki, K. (Dec, 2017). Measure of departure from marginal average point-symmetry for two-way contingency tables. 10th Conference of the IASC-ARS/68th Annual NZSA Conference, The University of Auckland, Auckland, New Zealand.

中野弘, 生亀清貴, 富澤貞男. (2017年9月). Directional measure for marginal homogeneity in square contingency tables. 統計関連学会連合大会, 南山大学.

6 . 研究組織

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。