氏名（本籍）	三枝祐輔（千葉県）
学位の種類	博士（理学）
学位記番号	甲第1137号
学位授与の日付	平成29年3月18日
学位授与の要件	学位規則第4条第1項該当
学位論文題目	**Modeling and Measure of Symmetry for Contingency Table Analysis** **（分割表解析における対称性のモデリングと尺度）**

論文審査委員　（主査）教授　富澤　貞男

教授　戸川　美郎　　准教授　宮本　暢子

教授　平場　誠示　　教授　瀬尾　隆

論文内容の要旨

カテゴリカルデータに対する統計解析手法の開発は20世紀初頭より行われてきた．それらの解析手法の適用範囲は，医学，薬学，疫学，公衆衛生学，行動計量学，生態学，経済学など多岐に渡っている．

分割表データ解析においては，得られた観測度数から，その観測度数の分布に対する潜在的な未知の確率分布を高い精度で推定することが主要な問題である．そのためには，データによく適合し，分かりやすい解釈を与えるモデルの導入が重要である．さらに，データに対して適用したモデルの当てはまりが悪いとき，モデルの分解はその原因を探ることにおいて有用である．また，未知の確率分布が適用したモデルからどの程度隔たっているかを測ることにも関心がある．

一般に，2元分割表解析における主な関心の一つは，行変数と列変数が独立かどうかである．しかし，行と列が同じ分類からなる正方分割表においては，観測度数が分割表の主対角線上のセルに集中する傾向がある．このことは，行変数と列変数間に関連がある，すなわち統計的独立性が成り立たないことを示している．また，行と列が同じ分類からなる分割表においては，独立性に代わって，分割表の主対角線に関する対称性や非対称性を示すモデルが数多く提案されている．

カテゴリカル変数は2種類に区別することができる．自然な順序を持たないカテゴリによって分類された変数は名義変数と呼ばれる．また，名義変数からなる分割表を名義カテ

ゴリ分割表と呼ぶ．名義変数の例として，髪の色（赤，青，黒，ブロンド）や血液型（A, B, O, AB），海域（太平洋，大西洋，インド洋，北極海，南極海）などが挙げられる．一方，自然な順序を持つカテゴリによって分類された変数は順序変数と呼ばれる．また，順序変数からなる分割表を順序カテゴリ分割表と呼ぶ．例えば，視力（良い，やや良い，やや悪い，悪い）や患者の容体（良好，普通，深刻）などが挙げられる．

本論文は 4 章からなる．第 1 章では分割表解析における研究の歴史的背景および理論的背景について述べた．また，先行研究において与えられたモデルやモデルの分解およびモデルからの隔たりの程度を測るための尺度について述べた．さらに，第 2 章以降の各章の概要を述べた．

第 2 章では，累積確率の非対称構造を示すモデルを提案した．また，提案したモデルを用いたセル確率の対称性の分解を与えた．1948 年に正方分割表のセル確率の対称構造を示す対称モデルが与えられた．1978 年にセル確率の和で表される累積確率に関する構造を示すパリンドロミック対称モデルとそのモデルの制約を弱めた一般化パリンドロミック対称モデルが提案された．1984 年には，正方分割表の行変数と列変数の周辺分布の同等性を拡張した構造を示す拡張周辺同等モデルが提案された．また，1989 年にパリンドロミック対称モデルが成り立つための必要十分条件は，拡張周辺同等モデルと一般化パリンドロミック対称モデルの両方が成り立つことであるという定理が与えられた．本章では，パリンドロミック対称モデルのもつ拡張周辺同等モデルの制約を，さらに一般化されたモデルである多母数一般化周辺同等モデルの制約で置き換えることによって，新たに多母数一般化パリンドロミック対称モデルを提案した．提案モデルは，順序カテゴリ正方分割表における累積確率の非対称構造を示す．また，提案モデルのもつパラメータは行と列の累積周辺確率の大小関係を推測することにおいて有用である．さらに本章では，対称モデルが成り立つための必要十分条件は，多母数一般化パリンドロミック対称モデル，行変数と列変数の原点まわりのモーメント一致構造，累積確率の部分対称性を示すモデルのすべてが成り立つことであるということを示した．このモデルの分解は，実際のデータ解析において，対称モデルの当てはまりが悪いとき，その原因を探ることにおいて有用である．さらに，提案モデルを米国黒人の両親の学歴に関する分割表データに適用した例を示した．また，英国人の父親とその息子の組を対象とした職業的地位に関する分割表データに各モデルを適用し，本章で与えたモデルの分解が，対称モデルの当てはまりが悪いときに，その原因を調べることにおいて有用であることを示した．さらに，カテゴリの順序の並べかえおよびカテゴリの併合の下でのモデルの不変性について議論した．

第 3 章では，f-ダイバージェンスに基づく順序準対称モデルを用いた対称モデルの分解を与えた．また，本章で与えたモデルの分解に関して，モデルの直交性が成り立つことを示した．2007 年に f-ダイバージェンスに基づく順序準対称モデルが提案された．このモデルと周辺分布の同等性を示すモデルを用いた対称モデルの分解が先行研究で与えられている．本章では，正方分割表に対する f-ダイバージェンスに基づく順序準対称モデルの制約数と周辺同等性を示すモデルの制約数の和が，対称モデルの制約数より大きいという意味で，先行研究の分解に問題があることを指摘した．さらに本章では，モデルの制約数を考慮した

新たな対称モデルの分解として，対称モデルが成り立つための必要十分条件はf-ダイバージェンスに基づく順序準対称モデルと行変数と列変数の平均一致構造が同時に成り立つことであるということを示した．加えて，本章で与えた新しいモデルの分解に関して，モデルの直交性が成り立つことを示した．すなわち，対称モデルの適合度を検定するための尤度比カイ二乗統計量は，f-ダイバージェンスに基づく順序準対称モデルと平均一致モデルに対する各尤度比カイ二乗統計量の和に漸近的に同等であることを示した．また，日本人大学生の視力に関する分割表データに各モデルを適用し，新しいモデルの分解がデータ解析において有用となることを示した．

第4章では，2次周辺対称モデルからの隔たりの程度を測るための尺度を提案した．1990年に多元分割表の1次周辺分布の同等性を示す1次周辺対称モデルと2次周辺分布の対称かつ同等性を示す2次周辺対称モデルが提案された．2001年に1次周辺対称モデルからの隔たりの程度を表す尺度が提案された．本章では，多元分割表データに対する潜在的な確率分布が2次周辺対称モデルからどの程度隔たっているのかを表すための尺度を提案した．その尺度は異なる複数の分割表間において，それぞれの2次周辺対称性からの隔たりの程度を比較することにおいて有用である．提案尺度はシャノンエントロピーの重み付き和で表され，0以上1未満の値をとる．提案尺度の値が0となるための必要十分条件は2次周辺対称モデルが成り立つことであり，2次周辺対称モデルからの隔たりの程度が大きくなるにつれて，提案尺度の値も大きくなる．また，本章では提案尺度の推定量の近似分散および近似信頼区間を与えた．さらに，2004年と2014年の米国の政府支出に対する意見のデータに提案尺度を適用し，尺度が複数の分割表間の2次周辺対称性からの隔たりの程度の比較において有用となることを示した．また，提案尺度と正規分布の関係を調べ，提案尺度が2次周辺対称性からの隔たりを測る尺度として妥当であることを示した．

論文審査の結果の要旨

統計学における分割表解析は，医学，薬学，理学，工学，心理学，教育学，社会学など幅広い分野で利用されている．ある人間の集団を高血圧の有無と心臓病の有無によって4つの小集団に分けると，各集団の人数からなる2×2分割表を得ることができる．また，2種類の分類が同じ分類からなる場合，たとえば，人間の左右裸眼視力を，悪い，やや悪い，やや良い，良い，の4つのカテゴリに分けると4×4分割表を得ることができる．このような2種類の同じ分類からなる分割表を正方分割表と呼んでいる．さらに，多種類の分類からなる場合は，多元分割表と呼んでいる．

分割表解析において，我々は全体（母集団）の一部分として得られた標本に基づく観測度数しか得ることができない．我々の関心は，母集団における未知の確率分布がどのような構造になっているのかを得られた観測度数から，高い信頼度で推測することにある．そのためには，データに良く適合し，かつ解釈が容易な確率分布に関する統計モデル（仮説）を導入する必要がある．さらに，モデルの未知パラメータの推定法，モデルの分解，モデ

ルからの隔たりを測る尺度の開発など多くの解決すべき問題点がある．

一般の2元分割表の解析において，多くの人の関心の一つは，分類間が独立（関連性がない）かどうかである．しかし，同じ分類からなる正方分割表や多元分割表の解析においては，分類間の関連性は極めて強く，統計的独立性は成り立たない．代わって，分類間の対称性に関心がある．たとえば，人間は，幼児の頃は左右裸眼視力は対称的であるが，成長とともに対称性は崩れる傾向にあり，左右どちらが良い目なのか，また，どのように非対称になるのかに関心がある．そのために，対称性あるいは非対称性に関する統計モデルを用いた解析が行われる．歴史的には，Bowker (1948) が対称モデルを導入したのが始まりである．その後，対称モデルを拡張したモデルが提案されている．たとえば，Stuart (1955) の周辺同等モデル，McCullagh (1978)のパリンドロミック対称モデル，そして Kateri and Agresti (2007)の f-ダイバージェンス型順序準対称モデルなどがある．

本論文は，同じ分類からなる正方分割表そして多元分割表において，対称性のモデルの拡張と分解，そして尺度の提案をしている．本論文は，4つの章から構成されている．

第1章では，本研究に関する歴史的な背景と各章の概要について述べている．

第2章では，順序カテゴリを持つ正方分割表において，周辺同等モデルとパリンドロミック対称モデルの両方を特別な場合に含むような非対称モデルとして，多母数一般化パリンドロミック対称モデルを提案している．また，「対称モデルが成り立つための必要十分条件は，多母数一般化パリンドロミック対称モデル，行変数と列変数の積率一致モデル，そして累積確率の部分対称モデルの全てが成り立つことである」というモデルの分解定理を与えている．さらに，モデルの推定法と検定法，そして具体的応用例を示し，提案したモデルと得られた分解定理の有用性を示している．特に，対称モデルが成り立たない場合，得られた分解定理はどのような確率構造が原因なのかを探るのに極めて有用である．

第3章では，順序カテゴリを持つ正方分割表において，「対称モデルが成り立つための必要十分条件は，f-ダイバージェンス型順序準対称モデルと，行変数と列変数の平均一致モデルの両方が成り立つことである」というモデルの分解定理を与えている．さらに，「対称モデルの適合度を検定するための尤度比カイ2乗検定統計量は，f-ダイバージェンス型順序準対称モデルと，行変数と列変数の平均一致モデルの各尤度比カイ2乗検定統計量の和に漸近的に同等である」という直交定理を与えている．そして実際の応用例と共に得られた直交定理の有用性を示している．

第4章では，多元分割表において，複数得られる2次元周辺分布に関して対称性と同等性を示す2次元周辺対称モデルが成り立たないとき，その隔たりを測るための尺度を提案している．よく知られた変数間の相関係数は，変数間の独立性からの隔たりの程度を測るのに有用であるが，本論文で提案されている尺度は，2次元周辺対称性からの隔たりがどの程度あるのかを測るのに有用である．提案された尺度は未知であるが，その推定法や推定した尺度の近似的分散，そして尺度の近似的信頼区間を理論的に与えている．提案した尺度を用いると，複数の分割表データにおいてモデルからの隔たりの程度が比較可能となり，従来の解析法では得られない詳細な解析が可能であり，極めて有用である．

以上，本論文は理論面と応用面の両面において大変高く評価できるものであり，分割表統計解析の分野に，独創的な新しい解析法を与えており，この分野に大きな貢献をしてい

る．よって理学的に価値ある知見と成果を得たもので博士（理学）の学位論文として十分価値あるものと認める．