博士（工　学）　関　谷　祐　里

学 位 論 文 題 名

多項分布における検定統計量の分布の近似理論に関する研究

学位論文内容の要旨

本論文は、多項分布の適合度検定におけるパワーダイバージェンス統計量の分布に対する新たな３つの近似法とそれに基づく検出力近似法を提案し、かつ、その有効性を検証するために従来の検出力近似法との比較を行ったものである。その３つの近似法とは、統計量の正規化変換に基づく方法（Normalizing Transformation: 以下 NT 近似と記す）、局所対立仮説における漸近展開の連続項に基づく方法（Asymptotic Expansion: 以下 AE 近似と記す）、および局所的でない対立仮説における漸近展開の連続項に基づく方法（Normal Approximation based on Asymptotic Expansion: 以下 NAAE 近似と記す）である。さらに、NT 近似において提案した方法を分割表の独立性検定の検出力近似に対して応用した。

多項分布の適合度検定において、Cressie and Read はパワーダイバージェンスに基づく検定統計量 R^a の族を提案した。ここで a は実数値をとるパラメータであり、$a=1$ のときは Pearson のカイ２乗統計量、$a=0$ のときは対数尤度比統計量、$a=-0.5$ のときは Freeman-Tukey 統計量、$a=-1$ のときは修正対数尤度比統計量である。また、$a=2/3$ に基づく統計量は Cressie and Read によって推奨された統計量である。これらの統計量 R^a はすべて、帰無仮説のもとで漸近的に同一のカイ２乗分布に従うこと、また局所対立仮説のもとで漸近的に同一の非心カイ２乗分布に従うことが知られている。さらに Read and Cressie は、分割表の独立性検定や複数の多項母集団の一様性検定においてもパワーダイバージェンスに基づく検定統計量を提案した。

多項分布の適合度検定におけるパワーダイバージェンス統計量の分布の近似に関するこれまでの主な研究として、単純帰無仮説のもとでの統計量の漸近展開 (Yarnold, Siotani and Fujikoshi, Read)、対立仮説のもとでの統計量の分布に対する正規近似 (Broffitt and Randles, Read and Cressie)、対立仮説のもとでの統計量の展開式における２次までの項に基づく２種類の検出力近似 (Drost et al.) があげられる。Drost et al. の研究によって非心カイ２乗近似より優れた検出力近似が与えられたということができるが、彼らの検出力近似法は互いに独立な非心カイ２乗確率変数の一次結合により構成されているため多くの計算量が必要となり、さらに a の値が１から減少するにつれて正確でなくなるという難点をもっている。

これに対して本論文で提案する NT 近似は、正規分布に基づいた近似法であるため計算が容易であり、$a<0$ の時の近似として有効であるという利点を持っている。正規化変換は統計量の分布に対する簡単で正確な近似を得るための有用な方法として知られており、NT 近似を導く上で考え出された統計量の具体的な正規化変換は、適合度検定統計量の分布に限らず多くの統計量に対して適用可能な方法である。実際、分割表の独立性検定や多項母集団の一様性検定の検出力近似にも用いられている。また漸近展開は、一般に統計量の分布を近似

するための有効な手法として知られているが、多項分布の適合度検定統計量は非常に複雑な離散分布であるため、その漸近展開は単純帰無仮説の場合にしか求められていなかった。しかし、本論文で初めて局所対立仮説の場合の漸近展開が与えられた。この場合の不連続項は帰無仮説の場合と違って簡単な式で表現することが非常に困難であって、実用的ではない。そこで、漸近展開の連続項に基づく検出力近似法として AE 近似を提案した。AE 近似は多くの a の値に対する共通の検出力近似法として優れており、不連続項の除去による影響はそれ程大きくないことが数値的に示された。さらに本論文では、局所的でない対立仮説の場合の漸近展開の連続項の式も与え、これに基づく検出力近似法として NAAE 近似を提案した。NAAE 近似は正規分布に基づく近似法であり、対数尤度比統計量（$a=0$ のとき）の検出力近似法として他の検出力近似法より優れているという利点を持っている。

本論文は以下の 9 章より構成される。

第 1 章では、本研究の背景と目的、および論文の構成について述べている。

第 2 章では、多項分布の適合度検定のためのパワーダイバージェンス統計量について述べ、帰無分布の近似および検出力近似に関するこれまでの主要な研究結果をまとめた。

第 3 章では、統計量の正規化変換に基づく検出力近似法（NT 近似）を与えた。初めに、Konishi によって与えられた正規化変換を構成するための一般的な枠組みに基づいて、本論文で提案する具体的な正規化変換を公式として与えた。次に、多項分布の適合度検定において、局所的でない対立仮説のもとでのパワーダイバージェンス統計量の分布に対して正規化変換の公式を適用することによって、新たな検出力近似である NT 近似を与えた。

第 4 章では、局所対立仮説における漸近展開の連続項に基づく検出力近似法（AE 近似）を与えた。初めに、局所対立仮説のもとでの確率に対する局所エッジワース近似を定理として与え、次にその局所エッジワース近似をもとに検定統計量の局所対立仮説における漸近展開の連続項と不連続項を与えた。そして、連続項である連続分布に対する多変量エッジワース展開の項を使って、新たな検出力近似である AE 近似を与えた。

第 5 章では、局所的でない対立仮説における漸近展開の連続項に基づく検出力近似法（NAAE 近似）を与えた。初めに、局所的でない対立仮説のもとでの確率に対する局所エッジワース近似について述べ、次にそれをもとに、局所的でない対立仮説のもとでのパワーダイバージェンス統計量の分布に対して、連続分布を仮定した多変量エッジワース展開を与え、それに基づく検出力近似である NAAE 近似を与えた。

第 6 章では、3 章、4 章、および 5 章で与えた 3 つの検出力近似法の性能を、従来の検出力近似法とともに数値的に比較した。数値計算結果により、従来の検出力近似法より優れた近似法として、$a=0$（対数尤度比統計量）に対しては NAAE 近似、$a=2$ に対しては AE 近似、$a=-0.5$ に対しては NT 近似であることが示された。

第 7 章では、第 3 章で述べた正規化変換に基づく近似法（NT 近似）の応用として、$s \times r$ 分割表の独立性検定の検出力近似、複数の多項母集団における一様性検定の検出力近似、およびスパース多項分布の適合度検定統計量の改良について述べた。

第 8 章では、第 4 章で述べた局所対立仮説における漸近展開の連続項に基づく検出力近似法（AE 近似）の多項分布の適合度検定のための ϕ-ダイバージェンス統計量の検出力近似への拡張について述べた。

第 9 章では本研究で得られた成果をまとめ、今後の展開について述べた。

学位論文審査の要旨

主　査	教　授	佐藤義治
副　査	教　授	宮腰政明
副　査	教　授	工藤峰一
副　査	教　授	栗原正仁
副　査	教　授	水田正弘

学位論文題名

多項分布における検定統計量の分布の近似理論に関する研究

諸現象の解析において，ある属性に関する「生起または非生起」あるいは「効果の有無」などが測定されたデータに基づく分析では，確率モデルとして多項分布が用いられることが多い。多項分布の検定問題は離散多変量解析の中心をなすものであり，その応用分野の広がりとともに今後さらに重要性が増す分野であると考えられる。

多項分布における検定法として最も基本的なものは適合度検定であり，その拡張として分割表の独立性検定や複数の多項母集団の一様性性検定などがある。

多項分布の適合度検定は，観測度数と帰無仮説のもとで得られる期待度数との差を評価することによって行われる検定法であり，代表的な適合度検定統計量としてはピアソンのカイ二乗統計量および対数尤度比統計量，Cressie & Read による検定統計量などがあるが，これらを含む一般的な統計量の族としてパワーダイバージェンス統計量 R^a がある。実数値パラメータ a の値を具体的に与えることによって，それに対応するパワーダイバージェンス統計量 R^a が定まる。さらに a の値にかかわらずこれらの統計量 R^a はすべて，帰無仮説のもとで漸近的に同一のカイ二乗分布に従うことが知られている。しかし，異なる a の値に対応する検定統計量の良さを評価するためには，対立仮説の検出力（即ち，対立仮説が正しいときに帰無仮説を棄却する確率）を計算する必要がある。

以上のような背景の下で，本論文は多項分布の適合度検定におけるパワーダイバージェンス統計量の分布に対する新たな三つの近似法とそれに基づく検出力近似法を考案し，それらの近似法の有効性を検証するために従来の検出力近似法との比較を行ない，さらに提案した近似法の一つの応用として分割表の独立性検定の新たな検出力近似法を考案したものである。三つの近似法とは，統計量の正規化変換に基づく方法（NT 近似）および局所対立仮説における漸近展開の連続項に基づく方法（AE 近似），局所的でない対立仮説における漸近展開の連続項に基づく方法（NAAE 近似）である。

多項分布の適合度検定におけるパワーダイバージェンス統計量の局所対立仮説のもとでの極限分布は，a の値にかかわらず全て同一の非心カイ二乗分布であるため，この近似に基づくと検出力は a の値によらず一定であり，検定統計量による差はないものと評価される。そ

こで, Drost et al. は対立仮説のもとでの統計量の漸近展開式における二次までの項に基づく検出力の近似法を与えたが, 互いに独立な非心カイ二乗確率変数の一次結合により構成されているためアルゴリズムが複雑であり, 数値的精度を上げることが困難である。

本論文で提案されたNT近似は, 正規分布に基づいた近似法であるため計算が容易であり, 特に $a<0$ の場合の近似として有効である。また, NT近似を導く上で考案された統計量の正規化変換は, 適合度検定統計量の分布に限らず多くの統計量に対して適用可能な方法であり, 実際, 分割表の独立性検定や多項母集団の一様性検定の検出力近似にも用いられている。また, 多項分布の適合度検定統計量は非常に複雑な離散分布であるためその漸近展開は単純帰無仮説の場合にしか求められていなかったが, 本論文で初めて局所対立仮説の場合の漸近展開が与えられた。この場合の不連続項は帰無仮説の場合と異なり簡単な式で表現することが非常に困難であり, 実用的ではない。そこで, 漸近展開の連続項に基づく検出力近似法としてAE近似が提案された。AE近似は多くの a の値に対する共通の検出力近似法として優れており, 不連続項の除去による影響はそれ程大きくないことが数値的に示されている。さらに本論文では, 局所的でない対立仮説の場合の漸近展開の連続項の式も与え, これに基づく検出力近似法としてNAAE近似が提案された。NAAE近似は正規分布に基づく近似法であり, 対数尤度比統計量（$a=0$ のとき）の検出力近似法として他の検出力近似法より優れている。

これを要するに, 著者は離散多変量確率分布として代表的な多項分布における仮説検定統計量の標本分布に関する近似理論に関して, 従来の検定統計量を包含するより一般的な検定統計量の族に対して, 帰無仮説のみならず対立仮説の下での優れた近似法を考案したものであり, 情報解析学, 計算機統計学の発展に寄与するところ大なるものがある。

よって, 著者は北海道大学博士（工学）の学位を授与される資格あるものと認める。