

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 16 日現在

機関番号：17701

研究種目：基盤研究(C) (一般)

研究期間：2012～2016

課題番号：24540133

研究課題名(和文) 多次元分割表における独立性検定の改良に関する研究

研究課題名(英文) Improvement of test for independence of multidimensional contingency table.

研究代表者

種市 信裕 (Taneichi, Nobuhiro)

鹿児島大学・理工学域理学系・教授

研究者番号：00207200

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：多項モデルを想定する多次元分割表において、一般次元の分割表における完全独立性検定、3次元分割表における一要因対二要因に関する独立性検定、3次元分割表の条件付独立性検定に対する改良をおこなった。具体的には、エッジワース展開による検定統計量の分布の近似に基づく対数尤度比統計量におけるパートレット修正を一般化した変換統計量の構築である。この変換統計量による検定はもとの統計量の検定の検出力はほとんど変わらないが、小中標本における適合度検定統計量のカイ二乗分布への収束の速さがもとの統計量より大幅に改良されている。

研究成果の概要(英文)：We consider multi-dimensional contingency table whose underlying distribution is assumed to be multinomial distribution. We improve the test for complete independence in contingency table of general dimension, test for one factor independence of the other two in three dimensional contingency table and test for conditional independence in three dimensional contingency table. To put it concretely, we constructed improved transformed statistics which are based on approximations of test statistics for independence by using Edgeworth expansion. When the statistic is the log-likelihood ratio statistic, the transformed statistic corresponds to Bartlett adjusted statistic. Nevertheless, the power of the test based on the transformed statistic is not so different from that of an original statistic, the transformed statistic improves the speed of convergence to chi-square distribution very well.

研究分野：数物系科学

キーワード：多次元分割表 完全独立性検定 一要因対二要因独立性検定 条件付独立性 エッジワース展開 変換統計量

1. 研究開始当初の背景

研究代表者(種市信裕), 研究分担者(関谷祐里)は離散モデルの適合度検定の漸近展開に基づく改良に関して, 基本的なモデルである多項分布の単純仮説の適合度検定より始め, 2次元分割表における独立性検定や, 一様性検定の改良の研究をおこなってきた。引き続き, 科研費基盤研究(C) 課題番号 20540124 研究課題名「応答変数が二項反応である一般線型モデル推測に関する研究」においては, 二項反応における一般化線型モデルへの適合度検定の改良をおこなった。本研究課題においては, 分割表における検定の改良をもう一步深め, 一般的な多次元分割表のモデルであるログリニアモデルの改良につなげるため, 多次元分割表の独立性モデルの検定の改良についての研究を開始した。

2. 研究の目的

(a) 多次元分割表の種々の独立性モデルにおける検定統計量の分布の近似の改良

カテゴリカルデータ解析において, いくつかの要因により非常に複雑に関係しあったデータを表現するためには多次元分割表を用いる必要がある。多次元分割表はセル数が非常に多くなるため一つのセル当たりのデータ数が稀となることが多い。このような場合に分析をおこなうと, 検定統計量の正確な分布とその漸近分布であるカイ二乗分布との乖離が大きくなり, 検定の信頼性が疑わしくなる。そこで, 本研究においては多次元分割表の種々の独立性モデル, 具体的には, 一般次元の分割表の完全独立性モデル, 三次元分割表における一要因対二要因の独立性モデル, 三次元分割表における条件付独立性モデルの独立性検定統計量の分布の漸近展開に基づく近似の構築をおこなう。この近似は確率変数の連続性を仮定したエッジワース展開による近似と離散項により与えられる。この近似において連続性を仮定したエッジワース展開による近似のみでも, 漸近分布に基づく近似より小・中標本において性能が優れていることが予想できる。

(b) エッジワース展開に基づく近似からの変換統計量の構築

本研究においては, 上述のエッジワース展開に基づく近似を尤度比検定統計量におけるバートレット修正を一般化した変換統計量の理論に当てはめ, 種々の多次元分割表の独立性検定のための変換統計量の構築をおこなう。これにより, 小・中標本においてもカイ二乗分布への近似のすぐれた統計量による検定が可能になる。

3. 研究の方法

多次元分割表の種々の独立性モデルにおける独立性検定統計量の分布の近似の改良, エッジワース展開に基づく近似からの変換統計量の構築に関して研究分担者(関谷氏)と議論し, 結果のチェックを受けながら, 以下のように研究を進めた。

(1) 多次元分割表に関する資料収集

研究の準備段階として, 多次元分割表の種々の独立性に関する資料, およびそれらの独立性とログリニアモデルとの関係に関する資料等の収集をおこなった。また, 現実のデータの考察のため多次元分割表の多数のデータセットの収集をおこなった。

(2) 多次元分割表における独立性検定統計量の漸近展開に基づく近似の導出

一般次元の分割表の完全独立性モデル, 三次元分割表における一要因対二要因の独立性モデル, 条件付独立性モデルそれぞれに, 連続性を仮定したエッジワース展開と離散項に分け, 種々の独立性検定統計量の漸近展開式の導出をおこなった。この間研究打ち合わせにより関谷氏に理論的アドバイスを受けた。導出に関しては, 研究代表者および研究分担者によるこれまでの成果 Taneichi and Sekiya, J. Multivariate Anal. 98, 1630-1657, 2007 や Taneichi, Sekiya and Toyama, J. Multivariate Anal. 102, 1263-1279, 2011 が本質的に役立った。

(3) 漸近展開に基づく近似の性能評価

数値計算により統計量の分布の漸近展開に基づく近似, 連続性を仮定したエッジワース展開による近似の性能評価をおこなった。関谷氏は性能評価のためのプログラミングのチェックをおこなった。計算結果のチェックのために研究打ち合わせをおこなった。

(4) エッジワース展開に基づく近似からの変換統計量の構築

エッジワース展開に基づいて, カイ二乗分布への収束の速さを改良する変換を構築し, これによる変換統計量を作成した。変換統計量の選択に関しては, 文献を収集し今まで用いてこなかったタイプの変換を選択肢に入れた。また, 数値計算により変換統計量の性能評価をおこなった。関谷氏は性能評価のためのプログラミングのチェックをおこなった。プログラミングのチェックおよび計算結果の考察のために研究打ち合わせをおこなった。

4. 研究成果

本研究においては、分割表の種々の独立性検定統計量としてはすべて従来の検定統計量(ピアソンカイ二乗統計量, 対数尤度比統計量, パワーダイバージェンス統計量の族)を特殊な場合として含む α -ダイバージェンス統計量の族を考えている.

本研究では、まず三次元分割表において、完全独立なモデルのもとでの検定統計量の局所エッジワース近似の導出、完全独立性検定統計量の分布の多変量エッジワース展開の導出、離散項の評価をおこなった. さらに、これらの結果のすべてについて、一般の M 次元分割表へ拡張をおこなった.

これらの研究成果を用いて対数尤度比統計量におけるパートレット修正の一般化であるカイ二乗分布への収束を速める変換統計量の構築をおこなった. 本研究においては、今までの研究において用いてきた改良変換と呼ばれるものに加えて Cordeiro-Ferrari 型の変換の構築もおこなった. 検定統計量をパワーダイバージェンス統計量の族とした場合に、モンテカルロシミュレーションにより、元の統計量と変換統計量における標本分布とカイ二乗分布との乖離を調べる性能評価をおこなった. その結果、変換統計量によりカイ二乗分布への収束の速さが改良されていることを示すことができた. 特に、Cordeiro-Ferrari 型の変換統計量の性能が良かった. さらに、モンテカルロシミュレーションを用いて検出力の評価をおこなった. この結果、変換統計量は元の統計量とほぼ同等の検出力を持つことがわかり、これらの結果より変換統計量は小・中標本でも信頼度が高い(つまり、正確な検定統計量の標本分布を用いた検定結果と、漸近分布を用いて検定統計量の分布の近似をおこなって得られる検定結果が相反することが少ない)検定統計量であることが確認できた.

次に、三次元分割表における一要因対二要因に関する独立性検定統計量および条件付独立性検定統計量の局所エッジワース近似の導出、独立性検定統計量の分布の多変量エッジワース展開の導出、離散項の評価をおこなった. これらの研究成果を用いて、変換統計量の構築をおこない、検定統計量をパワーダイバージェンス統計量の族とした場合にモンテカルロシミュレーションにより、カイ二乗分布への収束の速さを調べる性能評価をおこなった. その結果、完全独立性検定の場合と同様に、変換統計量によりカイ二乗分布への収束の速さが改良され、特に Cordeiro-Ferrari 型の変換統計量による改良が著しいことがわかった. また、検出力の評価をおこなった結果においても変換統計量と元の統計量はほぼ同等であることが確認され、これらの独立性検定においても変換統計量は小・中標本でも信頼度が高い検定統計量であることが確認できた.

これらの結果から、三次元分割表においてはパートレットの最終独立性検定以外のすべての独立性検定のための変換統計量が構築されたことになる. また、四次元以上の分割表においても、本研究の成果を適用することによって、多様な独立性検定の変換検定統計量の構築が可能となった. また、本研究をおこなったことによって、多次元分割表のログリニアモデルへの理論の拡張のために、以前おこなった二項反応のロジスティック回帰モデルにおける方法論を使用できることが明確となり、ログリニアモデル研究への見通しがついた.

5. 主な発表論文等

(研究代表者, 研究分担者には下線.)

(雑誌論文)(計 5 件)

Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, Improved transformation of α -divergence goodness-of-fit test statistics based on minimum α -divergence estimator for GLIM of binary data, *Sut Journal of Mathematics*, 52(2), 193-215, 平成 28 年 (2016), 査読有

<http://www.rs.tus.ac.jp/sutjmath/>

関谷祐里, 種市信裕, 漸近展開の不連続項を利用した離散パートレット型変換統計量, *日本統計学会誌*, 45(1), 1-17, 平成 27 年 (2015), 査読有

<http://www.terrapub.co.jp/journals/jjssj/pdf/4501/45010001.pdf>

Takasumi Kobe, Nobuhiro Taneichi, Yuri Sekiya, Improved transformed statistics for the test of one factor independence from the other two in an $r \times s \times t$ contingency table, *Journal of the Japan Statistical Society*, 45(1), 77-98, 平成 27 年 (2015), 査読有

<http://www.terrapub.co.jp/journals/jjssj/pdf/4501/45010077.pdf>

Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, A new family of parametric links for binomial generalized linear models, *Journal of the Japan Statistical Society*, 44(2), 119-133, 平成 26 年 (2014), 査読有

<http://www.terrapub.co.jp/journals/jjssj/pdf/4402/44020119.pdf>

Nobuhiro Taneichi, Yuri Sekiya, Jun

Toyama, Transformed goodness-of-fit statistics for a generalized linear model of binary data, Journal of Multivariate Analysis, 123, 311-329, 平成 26 年 (2014), 査読有
<http://dx.doi.org/10.1016/j.jmva.2013.09.014>

〔学会発表〕(計 17 件)

Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, On improvement of generalized χ^2 -divergence goodness-of-fit statistics for GLIM of binary data, Hokkaido International symposium, Recent Development of Statistical Theory in Statistical Science, 2016. 10. 28, Hokkaido University (Hokkaido, Sapporo)

種市信裕, 関谷祐里, 外山淳, 二項反応における拡張されたリンク関数族に基づく一般化線型モデルについて, 2016 年度統計関連学会連合大会, 2016. 9. 5, 金沢大学 (石川県・金沢市)

Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, Improved χ^2 -divergence test statistics based on minimum χ^2 -divergence estimator for GLIMs of binary data, 22nd International Conference on Computational Statistics, 2016. 8. 26, Oviedo (Spain).

種市信裕, 多項分布の適合度検定統計量の分布の近似と離散統計モデルへの応用, 日本数学会 2016 年度年会, 特別講演(招待講演), 2016. 3.18, 筑波大学 (茨城県・つくば市)

Nobuhiro Taneichi, Improvement of deviance for binary models, Kumamoto International Symposium, High dimensional statistical analysis & quantile analysis for time series, 2016. 3. 4, Kumamoto University (Kumamoto, Kumamoto).

種市信裕, 関谷祐里, 外山淳, 二項反応における一般化線型モデルのリンク関数の拡張, 科研費シンポジウム「多様な分野における統計科学の新展開」, 2015. 10. 26, 富山県民会館 (富山県・富山市)

関谷祐里, 種市信裕, 漸近展開の不連続項を利用した離散バートレット型変換統計量の性質について, 科研費シンポジウム「多様な分野における統計科学の新展

開」, 2015. 10. 26, 富山県民会館 (富山県・富山市)

小部敬純, 種市信裕, 関谷祐里, 3次元分割表における1要因対2要因の独立性検定による改良変換統計量, 科研費シンポジウム「多様な分野における統計科学の新展開」, 2015. 10. 25, 富山県民会館 (富山県・富山市)

関谷祐里, 種市信裕, 多項分布の適合度検定における不連続項を考慮した統計量の変換について, 2015 年度統計関連学会連合大会, 2015. 9. 8, 岡山大学 (岡山県・岡山市)

関谷祐里, 種市信裕, 多項分布の適合度検定統計量の分布に対する漸近展開の不連続項の利用について, 日本数学会 2015 年度年会, 2015. 3. 23, 明治大学 (東京都・千代田区)

関谷祐里, 種市信裕, 漸近展開の不連続項を考慮した多項分布の適合度検定統計量について, 2014. 12. 2, 科研費シンポジウム「統計的推測の理論的基礎とその応用」, 2014. 12. 2, 筑波大学 (茨城県・つくば市)

種市信裕, 関谷祐里, 外山淳, 二項反応の一般化線型モデルの検定, 九州大学統計科学セミナー, 2014. 10. 9, 九州大学数理学府 (福岡県・福岡市)

Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, Bartlett adjustment of deviance statistic for three types of binary response models, 21st International Conference on Computational Statistics, 2014. 8. 22, Geneva (Switzerland)

種市信裕, 関谷祐里, 外山淳, 二項反応一般化線型モデルにおける変換適合度検定統計量, 科研費シンポジウム「一般化線形モデルの最新の展開とその周辺」, 2013. 11. 9, 千葉大学西千葉キャンパス (千葉県・千葉市)

種市信裕, 関谷祐里, 外山淳, 二項反応の一般化線形モデルにおけるパワーダイバージェンス適合度検定統計量の改良, 科研費シンポジウム「統計推測理論の展開と諸モデルへの応用」, 2012. 10. 3, 釧路市生涯学習センター (北海道・釧路市)

種市信裕, 関谷祐里, ロジスティック回帰モデルにおける χ^2 -divergence 適合度検定統計量の帰無分布の漸近展開, 日本数学会 2012 年度秋季総合分科会, 2012. 9. 20, 九州大学 (福岡県・福岡市)

種市信裕, 関谷祐里, 外山淳, 二項反応の一般化線型モデルにおけるパワーダイバージェンス適合度検定統計量の漸近近似について, 2012 年度統計関連学会連合大会, 2012. 9. 10, 北海道大学 (北海道・札幌市)

6 . 研究組織

(1)研究代表者

種市信裕 (TANEICHI NOBUHIRO)
鹿児島大学・理工学域理学系・教授
研究者番号 : 00207200

(2)研究分担者

関谷祐里 (SEKIYA YURI)
北海道教育大学・教育学部・釧路校・教授
研究者番号 : 10226665

(3)連携研究者

なし ()