

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 28 日現在

機関番号：17701

研究種目：基盤研究(C)

研究期間：2008 ～ 2011

課題番号：20540124

研究課題名（和文）：応答変数が二項反応である一般線型モデル推測の改良に関する研究

研究課題名（英文）：Improvement of inference for generalized linear model with binary response.

研究代表者：種市 信裕（TANEICHI NOBUHIRO）

鹿児島大学・大学院理工学研究科・教授

研究者番号：00207200

研究成果の概要（和文）：応答変数が二項反応である一般線型モデルにおいて2つの観点からモデル推測の改良をおこなった。一方は, Aranda-Ordaz によって提案されたリンク関数を拡張することによって得られるリンク関数の導入である。これにより種々のデータに柔軟に対応できるモデルが選択できるようになった。もう一方は, エッジワース展開による適合度検定統計量の分布の近似に基づく改良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と検出力はほとんど変わらないが, 小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 improved the inference for generalized linear model with binary response from two kinds of view point. On the one hand, we constructed family of parametric link functions by extending family of a link function proposed by Aranda-Ordaz. By the family of links, we can choose a model that can be fitted various data. On the other hand, we constructed an improved transformed statistic which is based on an approximation of goodness-of-fit test statistic by using Edgeworth expansion. Though the power of the test based on the transformed statistic is not so different from that of original statistic, the transformed statistic improves the speed of convergence to chi-square distribution very well.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,100,000	330,000	1,430,000
2009年度	800,000	240,000	1,040,000
2010年度	500,000	150,000	650,000
2011年度	800,000	240,000	1,040,000
年度			
総計	3,200,000	960,000	4,160,000

研究分野：数物系科学

科研費の分科・細目：数学一般（含確率論・統計数学）

キーワード：二項反応, 一般線型モデル, リンク関数, エッジワース展開, デビアンス, ϕ -ダイバージェンス適合度検定統計量, 漸近展開, 改良変換統計量

1. 研究開始当初の背景

研究代表者(種市信裕), 連携研究者(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を目指して基本的なモデルから始め, より一般的なモデルへとその研究を進めて来た。当初, 多項分布の単純帰無仮説の適合度検定, 分割表における独立性検定や一様性検定の研究をおこなっていた。本研究課題においては研究対象をより一般的分野へ向け, 医学, 工学等, 統計解析において非常に有効でありかつ頻繁に用いられるロジスティック回帰モデルに代表される応答変数が二項反応である一般線型モデルについての研究を開始した。

2. 研究の目的

(1) 二項反応モデルにおけるリンク関数の改良

二項反応モデルにおいて, そのリンク関数に着目する。データに対するモデルの適合度の良さは 0, 1 近辺にデータが多い時は, Probit リンクよりも Logit リンクの方が良く, また, ある点に対してのデータの形状の対称性が大きく崩れた時はこれらのリンクよりも Complementary log-log リンクのほうが勝ることは Agresti (Categorical Data Analysis 2nd ed., Wiley, 2002) 等でも指摘されているところである。また, Logit リンクと Complementary log-log リンクを包含するリンクの族が Aranda-Ordaz により考案されている。

本研究ではこれら現在用いられているリンク関数の適当な性質を維持しながらデータの種類に対応して柔軟に変化できるリンク関数の構築をおこなう。これにより, より適合度の良い二項反応モデルをつくることができる。

(2) 二項反応モデルにおける適合度検定統計量の分布の近似の改良

二項反応モデルとデータ間の適合度をデビアンス(対数尤度比統計量)やピアソンカイ二乗統計量などの適合度検定統計量を用いて検定をおこなう場合, これらは通常その漸近分布であるカイ二乗分布を用いておこなわれている。このことに関してはそれほど疑問もなく用いられているようであるが, 二項反応モデルにおいてはそれぞれの設定における観測値は少ない場合が多いため, このような大標本の理論を用いることは好ましくないとされる。また文献においても, McCullagh and Nelder (Generalized

linear model 2nd ed., Chapman and Hall, 1989)においては, 連続反応モデルにおけるデビアンスのバートレット変換に関しては言及されているが, 離散反応の場合にはよくわからないと記述されており Agresti (Categorical Data Analysis 2nd ed., Wiley, 2002) や Dobson (An Introduction to generalized linear models, 2nd ed., 2002) には通常用いられる統計量の連続補正について触れているにすぎない。そこで本研究においては, 構築されたモデルに対して小・中標本でも対応できるように, 適合度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の部分と, 離散修正項により与えられる。ここで, 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。

(3) 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統計量におけるバートレット変換を一般化した改良変換の理論に当てはめ, 二項反応の適合度検定のための変換統計量の構築をおこなう。このことにより, 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

二項反応モデルにおけるリンク関数の改良, 二項反応モデルにおける適合度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順を追って, 連携研究者(関谷氏)と議論し結果のチェックをうけながら, 以下のように研究を進めた。

(1) 二項反応モデルにおけるリンク関数の改良に関する資料収集

リンク関数として有名な Probit リンク, Logit リンク, Complementary log-log リンクは, それぞれある分布関数から派生してきたものである。そこで種市が多数の分布関数に関する資料の収集をおこなった。

(2) リンク関数の作成と性能評価

種市が(1)において収集した多数の分布関数を用いて, それぞれの分布関数より派生するリンク関数の作成をおこなった。また,

種市は多数の特徴あるデータセットを作成し、構築したリンク関数がどの種類のデータに対して優れているかの性能を調べた。この過程は関谷氏との研究打ち合わせのもとでおこなわれた。

(3) 種々のデータに対応できるリンク関数の構築

関谷氏とアイデアの情報交換をおこないながら Aranda-Ordaz によって提案されたリンク関数をベキ関数により拡張をおこない、種々の検定方式を作成した。

(4) 二項反応モデルにおける適合度検定統計量の分布の近似の改良の準備

構築したリンク関数を用いたモデルについて、その適合度検定統計量（ピアソンカイ二乗統計量、デビアンズ）の分布の漸近展開に基づく近似を与えるための方法論の資料収集および関谷氏と研究打ち合わせをおこないながら考察、考案をおこなった。

(5) 適合度検定統計量の分布の漸近展開に基づく近似の導出

連続性を仮定したエッジワース展開と離散修正項に分け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。

この部分に関しては、研究代表者および連携研究者による成果 Sekiya and Taneichi, J. Multivariate Anal. 91, 199-223, 2004 や Taneichi and Sekiya, J. Multivariate Anal. 98, 1630-1657, 2007 が本質的に役立った。

(6) 適合度検定統計量の分布の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。計算結果の考察のために研究打ち合わせをおこなった。

(7) 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く近似より改良変換を構築し、これによる変換統計量を作成した。また、数値計算によりこの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。

(8) 研究成果の公表

国内学会等への公表は進展に応じて毎年おこなった。外国への公表は最終年（2011）に invite された学会

International Conference on Advances in Probability and Statistics-Theory and Applications (Hong Kong)

および、

The 58th Session of the International Statistical Institute (Dublin, Ireland)

でおこなった。

4. 研究成果

研究の目的(1)に関して

データの対称性が崩れるにつれてリンク関数の適当な性質を維持しながら logit リンクから complementary log-log リンクのような非対称リンク関数へ連続的に変化するリンク関数の族の考案をおこない、3種類のリンク関数の構築をおこなった。

また、Aranda-Ordaz によって提案されたリンク関数をベキ関数により拡張することによって得られるリンク関数の導入をおこなった。これらのリンク関数が累積分布関数の逆関数として与えることが出来ることに基づき、これらの分布関数の skewness (歪度) と kurtosis (尖度) に基づく指標を作成し、拡張されたリンク関数ではモデルとして達成できるが Aranda-Ordaz によるリンク関数では達成出来ないデータの特性があることを示した。さらに、モデルが Aranda-Ordaz によるリンク関数を用いたもので十分なのか、それとも拡張されたリンク関数を用いたモデルが必要なかを評価するための近似スコア検定を構築し、拡張されたリンクを用いることが必要と判断された場合には、特定のモデルが適合しているかを判断する近似スコア検定を構築した。これにより種々のデータに柔軟に対応できるモデルが選択できるようになった。

研究の目的(2), (3)に関して

リンク関数が正準リンク関数である logit 関数である場合、つまりロジスティック回帰モデルの場合のデビアンズ (対数尤度比統計

量)の下側確率の漸近展開式を与えた。この漸近展開式は連続項および離散項よりなる。計算された漸近展開式の連続項に基づき、デビアンズにパートレット修正をほどこした変換統計量を導出した。この変換統計量の性能の評価をモンテカルロシミュレーションを用いておこなった結果、この変換統計量による検定の検出力はもとのデビアンズによる検定の検出力とほとんど変わらないにもかかわらず、変換統計量のカイ二乗分布への少ない標本数における収束の速さはもとのデビアンズより非常に改良されていること、つまり少ない標本数におけるカイ二乗分布を用いて検定することの妥当性が示された。

さらに、リンク関数が logit のみならず、complementary log-log リンクを包含する Aranda-Ordaz によるリンク関数族や probit リンクをも含む一般型リンク関数による一般線型モデルにおけるデビアンズの下側確率の漸近展開式を与えた。この漸近展開式の連続項に基づきデビアンズにパートレット修正をほどこしたデビアンズの変換統計量の構築をおこなった。一般型リンク関数による一般線型モデルにおけるデビアンズの変換統計量の性能の評価をモンテカルロシミュレーションを用いておこなった結果も logit リンクの場合と同様であり、変換統計量の有効性が示された。

また、適合度検定統計量として、デビアンズを特殊な場合として含むパワーダイバージェンス適合度検定統計量、さらにはより一般化をおこなった ϕ -ダイバージェンス検定統計量の下側確率の漸近展開式の導出をおこなった。また、漸近展開式の連続項を用いたパートレット修正の一般的検定統計量への拡張である改良変換統計量の構築をおこなった。 ϕ -ダイバージェンス検定統計量の具体的な例としてパワーダイバージェンス適合度検定統計量を用いて、変換統計量の性能の評価をモンテカルロシミュレーションによりおこなった結果も多くの場合 logit モデルにおけるデビアンズによる検定と同様の結果が得られた。

これらのことより、導出された変換統計量による検定はもとの統計量の検定と検出力はほとんど変わらないが、小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ことがわかり、研究の目的である優れた検定統計量による検定が可能となった。

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 3 件)

- ① Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, Improved transformed deviance statistic for testing a logistic regression model, Journal of Multivariate Analysis, 102, 1263 - 1279, 平成 23 年 (2011), 査読有 doi:10.1016/j.jmva.2011.04.010
- ② Pan Ei Htwe, Nobuhiro Taneichi, Yuri Sekiya, Improved approximations for the distributions of multinomial goodness-of-fit statistics based on ϕ -divergence under nonlocal alternatives, Journal of the Japan Statistical Society, 41(2), 121 - 142, 平成 23 年 (2011), 査読有 <http://www.terrapub.co.jp/journals/jjss/frame/41.html>
- ③ Nobuhiro Taneichi, Yuri Sekiya, Approximations of the distributions of test statistics for homogeneity of a product multinomial model, Communications in Statistics - Theory and Methods, 37, 1610-1631, 平成 20 年 (2008), 査読有 doi:10.1080/03610920801893947

[学会発表] (計 12 件)

- ① Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, On a transformed ϕ -divergence goodness-of-fit statistic for testing logistic regression model, International Conference on Advances in Probability and Statistics-Theory and Applications, (Invited speaker) 2011. 12. 28, Hong Kong. (Chinese University of Hong Kong)
- ② 種市信裕, 関谷祐里, 二項反応の一般線型モデルにおけるあるリンク関数族の性質について, 科学研究費シンポジウム「統計的推測の理論と方法論, 及び, 最近の動向」2011. 11. 8, 筑波大学自然系学系棟 D 棟 D509
- ③ 種市信裕, 関谷祐里, 外山淳, 二項反応の一般線型モデルにおけるデビアンズの改良について, 科学研究費シンポジウム「統計的推測の理論と方法論, 及び, 最近の動向」, 2011. 11. 8, 筑波大学自然系学系棟 D 棟 D509

- ④ Nobuhiro Taneichi, Yuri Sekiya, Jun Toyama, On performance of transformed ϕ -divergence goodness-of-fit statistic for testing logistic regression model, The 58th Session of the International Statistical Institute, 2011. 8. 26, Dublin, Ireland. (Dublin Convention Center)
- ⑤ 種市信裕, 関谷祐里, 外山淳, ロジスティック回帰モデルにおけるデビアンズ統計量の改良について, 科学研究費シンポジウム「生物情報解析の理論的基礎とその応用」, 2010. 12. 7, 東海大学代々木校舎4号館
- ⑥ 種市信裕, 関谷祐里, Pan Ei Htwe, 二項反応の一般線型モデルにおけるリンク関数族とその選択について, 科学研究費シンポジウム「生物情報解析の理論的基礎とその応用」, 2010. 12. 7, 東海大学代々木校舎4号館
- ⑦ 種市信裕, 関谷祐里, 多項分布の適合度検定統計量の漸近展開における離散項 J_2 の評価についての考察, 科学研究費研究集会「統計的推測方法の理論的展開とその応用」, 2010. 11. 18, 熊本大学大学院自然科学研究科研究棟
- ⑧ 種市信裕, 関谷祐里, Pan Ei Htwe, 二項反応の一般線型モデルにおけるリンク関数について, 科学研究費研究集会「統計モデルによる現象の解析, 並びに, その基礎理論」, 2010. 10. 22, 山形テルサ
- ⑨ 関谷祐里, 種市信裕, ロジスティック回帰分析における検定統計量のモーメント修正に基づく近似の改良について, 日本数学会 2010 年度年会, 2010. 3. 26, 慶応義塾大学(理工学部 矢上キャンパス)
- ⑩ 種市信裕, 関谷祐里, 外山淳, ロジスティック回帰分析の検定統計量の分布の漸近展開, 科学研究費研究集会「多変量モデル・時系列モデルにおける統計的推測の理論と応用」, 2009. 12. 3, 鹿児島大学理学部 理系先端研究棟 大会議室
- ⑪ 種市信裕, 関谷祐里, ロジスティック回帰分析の改良変換に基づく検定統計量について, 2009 年度日本数学会年会, 2009. 3. 28, 東京大学駒場キャンパス
- ⑫ 種市信裕, 関谷祐里, ロジスティック回帰分析における適合度検定統計量の分布の近似について, 2008 年度統計関連学会連合大会, 2008. 9. 8, 慶応義塾大学(理工学部 矢上キャンパス)

6. 研究組織

(1) 研究代表者

種市信裕 (TANEICHI NOBUHIRO)
 鹿児島大学・大学院理工学研究科・教授
 研究者番号：00207200

(2) 連携研究者

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