

令和 4 年 5 月 30 日現在

機関番号：32660

研究種目：基盤研究(C) (一般)

研究期間：2018～2021

課題番号：18K11200

研究課題名(和文)絶対尺度のモデル評価基準および予測分布の優越性評価法の開発

研究課題名(英文)Absolute model evaluation measures and evaluations of predictive densities

研究代表者

黒沢 健 (Kurosawa, Takeshi)

東京理科大学・理学部第一部応用数学科・准教授

研究者番号：80582303

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：線形モデルの一般化である一般化線形モデルと呼ばれる統計モデルが存在し、ロジスティック回帰モデル、多項ロジットモデル、ポアソン回帰モデル等の多種多様な統計モデルを含む。これらの統計モデルは幅広い分野で用いられている。モデルはデータを用いて構築されるが、その出来上がったモデルのもっともらしさを評価することは非常に重要な問題である。一般化線形モデルには一般的に普及しているモデル評価尺度は存在しないため、提案されているモデル評価尺度の性質を明らかにした。

研究成果の学術的意義や社会的意義

世の中の生活の中には様々な統計モデルが潜んでいる。インターネットで買い物をする際、お勧めの商品をWEBサイトが推薦してくれる。これらは統計的なモデルが使われていることがあり、その良し悪しを評価する尺度を開発することは学術的な意味だけではなく、より便利な社会に役立てる一般的な社会的な意義を持ち合わせている。

研究成果の概要(英文)：Generalized linear models that are generalizations of the linear models include various types of models such as logistic models, multinomial logit models, and Poisson regression models. These statistical models are widely used in various practical fields. The models are constructed based on data set. It is an important problem that the models are evaluated appropriately by using a goodness-of-fit measure. There do not exist widely used goodness-of-fit measures such that the measure can evaluate the generalized linear models. We have clarified the statistical properties of the proposed model evaluation measures.

研究分野：統計学

キーワード：モデル評価 予測分布

1. 研究開始当初の背景

線形モデルの一般化として、一般化線形モデル(GLM) と呼ばれる統計モデルのクラスが存在する。GLM にはロジスティック回帰モデル、多項ロジットモデル、ポアソン回帰モデル等の多種多様な統計モデルを含み、医学、交通工学、生態系モデル、計量経済学等、幅広い分野で用いられている。そのモデルのもっともらしさや当てはまり度を評価することは非常に重要な問題である。一般化線形モデルの特殊なモデルである、線形モデルには決定係数と呼ばれる絶対尺度のモデル評価尺度が存在する。しかしながら GLM に関しては、一般的に普及した絶対尺度のモデル評価尺度は存在しない。AIC は広く統計学で使用されているモデルを比較選択するための尺度である。これは複数のモデルが与えられた時に、それらの AIC の値を計算することで、どのモデルが優れているのかという相対的な比較をするために用いられる尺度であるため、モデルそのものの絶対的なもっともらしさを評価する尺度ではない。モデル選択という観点においては、AIC やその亜種を用いることで十分ではある。一方でモデルを絶対的に評価する尺度と観点から考えた場合、GLM に対する一般的に普及したモデル評価尺度が存在しないという大きな問題がある状況にある。

もう一つの検討課題は予測分布の評価である。統計モデルを構築する一つの目的として、観測していない未観測データの予測、つまり予測分布を構成することにある。しばしば観測したデータと未観測データの観測データが独立ではなく、従属的な関係にあることがある。従属的な関係にあれば、観測データの条件付き分布の予測分布を構成することができ、条件付けされていない予測分布よりもその予測精度の向上が見込まれる(条件付けされていない場合は、観測データからパラメータを推定することで予測分布を構成する)。例えば、従属的な関係がある観測データの例として、故障データや生存時間解析に用いられる打ち切りデータがあり、打ち切る前のデータ(観測データ)と打ち切られなかったデータ(未観測データ)は順序統計量で表され、従属的な関係にある。また時系列データは確率変数の列として観測され、各時点間の関係性を持つ。そこで、観測データと未観測データの間に従属関係がある場合の予測分布を適切に選べないか?といった問題が沸き上がる。例えば、打ち切りデータに関するベイズ理論を利用した予測分布として、多くの予測分布が提案されているがその評価は十分とはいえない。

2. 研究の目的

AIC などその他の多くの有用なモデル選択基準が存在する。それらの多くの目的は「モデル選択」である。モデル選択が目的であるならば AIC のように相対尺度であることは問題にはならないが、本研究で注目している問題は状況が異なり、線形モデルに対する決定係数のように、それらの値からモデルの適合度を評価できる絶対尺度のモデル評価基準に着目していることにある。また絶対尺度のモデル評価基準は、実際の現場における統計解析において、直感的に解釈がしやすい適合性の評価であり、実用上の意義は大きい。更には、決定係数と親和性を持つようなものが望ましい。しかしながら決定係数のように GLM に汎用的に使われているモデル評価基準は存在しない。実際はこれらの要求に沿った汎用的なモデル評価基準がいくつか提案されているが、まだ理論的な性質は十分検討されておらず、これらの評価基準を利用した検定統計量等は提案されていない。よって数少ない提案されている要求に沿ったモデル評価基準の理論的性質を解明することが目的となる。

もう一点目の予測の問題においては、例えば、打ち切りデータに関する問題において、ベイズ理論を利用した予測分布として、様々な状況下と仮定下で事後予測分布が提案されている。しかしながら、事前分布の選び方など、恣意的な部分も否めなく、事前分布のパラメータの選び方については、多くの場合何も示唆されていない。そのような意味で予測分布を適切に選べているとは考えにくい。しかし、ベイズ決定理論を用いることで、分布間の優越性を評価する方法が利用されつつある。近年、このベイズ決定理論を利用して、例えば線形モデルに対する未観測値の予測分布が提案され、他の予測分布を優越するような予測分布のクラスの検討がされている。これらは、観測データと未観測データが独立の場合の検討である。一方で、評価が十分に行われていない従属的な関係にあるデータに対する予測分布の優越性の評価を得ることが目的となる。

3. 研究の方法

上述の研究背景および研究目的から、GLM に汎用的かつ絶対尺度のモデル評価基準が望まれ、カテゴリカルデータにおける権威である Zheng and Agresti によって回帰相関係数(RCC) が提案(Stat. Med., 2000)され、研究分担者である江島等によって ECC (Statist. Probab. Lett., 2007), ECD (Comput. Statist. Data Anal., 2010) 等のモデル評価基準が提案されてきたが、まだその理論的性質は十分に議論されていない。また研究代表者およびその共著者との先行研究で、GLM の一種であるポアソン回帰モデルの RCC の理論値の導出と、その推定量の提案に関する研究成果 (Comput. Statist. Data Anal., 2016)), 及び、その推定量の一致性と漸近正規性に関する研究成果 (Comm. Statist. Theory Methods, 2018) も得ていた。これらの成果を活かした研究を目指した解決法となる。より具体的には、RCC に適用した手法を、その他のモデル評価基準である ECC, ECD 等の理論的性質の解明することである。研究分担者は ECC および ECD を開発した研究者である。理論的な観点からのサポートを得て、まだ未解明な統計的な性質を得ることにある。

また、二つ目の研究課題に関しては、過去の研究代表者の時系列解析及び打ち切りデータに関する研究成果を元に研究を開始する。特に Bayes 型の予測分布の評価においては、The Bayesian Choice (C. Robert, 2007)における基本的なベイズ決定論に基き検討を始める。対象を打ち切りデータとした研究結果は、研究代表者が知る限り存在しえないため、未知なる検討となる。そのため、Komaki (Biometrika, 2001)や Aitchison(Biometrika, 1975)における先行研究を参考に研究を開始する。

また研究協力体制として、研究代表者は課題研究 1 年目に 1 年間、オーストラリアの Australian National University (ANU)に在外研究員として滞在していた。そのため、研究分担者に加え、在外研究先の大学での著名な研究者との連携を考えた。その際、研究テーマを共有し、複数の研究テーマに関して意見の交換や研究協力を得ていた。これらの研究テーマに精通している海外の一流の大学で研究している研究者の助力を得て、両課題とも研究を促進する。

4. 研究成果

まずは 1 つ目の課題である GLM の絶対評価の評価基準に関する研究においては、研究分担者である江島及び ANU の研究者等と共に、ECC および ECD の基礎統計量となる mpp (measure of predictive power) と呼ばれる統計量の基本的な性質を議論することができ、共著論文として研究成果 (ANZJS, 2020)をまとめることができた。またこの成果に関連した複数の結果を国内研究会や国際学会の場で発表をした。その研究発表の一部の成果を (1 時間を超える) 招待講演 (Fellows Research Meetings, 2019)として発表を行った。また ECC および ECD に関連する大きな成果として、研究分担者である江島は ECC や ECD に関連する洋書の書籍「Statistical Data Analysis and Entropy」の出版 (Springer, 2019)を行った。

ベイズ型の予測分布に関しては、滞在先の ANU やその他の関連研究者と共に複数の研究成果を達成した。具体的には、指数到着を仮定した Type II 打ち切りデータに関するベイズ型の予測分布の提案と、Kullback-Leibler divergence 損失を用いたリスクの下で、提案した予測分布の定量的な評価に関する成果を得た。その結果を複数の国内学会や国際学会で発表を行った。特に、滞在先の ANU にて講演の依頼を受け、1 時間を超える招待講演を行った。

上記は代表的な研究成果ではあるが、上記の研究テーマを推進する過程において発生した研究課題があり、関連する多くの研究成果を得た。

直接的な研究成果とは異なるが、研究代表者が 1 年目に在外研究員として ANU に滞在した経験及びそこで得た著名な研究者との密な関係が得られたことは副次的な成果として大きな収穫であった。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 3件/うちオープンアクセス 3件）

1. 著者名 Takeshi Kurosawa, Francis K.C. Hui, A.H. Welsh, Kousuke Shinmura, and Nobuoki Eshima	4. 巻 62(3)
2. 論文標題 On goodness-of-fit measures for Poisson regression models	5. 発行年 2020年
3. 雑誌名 Australian & New Zealand Journal of Statistics	6. 最初と最後の頁 340-366
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/anzs.12303	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Nobuoki Eshima, Claudio Giovanni Borroni, Minoru Tabata, and Takeshi Kurosawa	4. 巻 23(2)
2. 論文標題 An entropy-based tool to help the interpretation of common-factor spaces in factor analysis	5. 発行年 2021年
3. 雑誌名 Entropy	6. 最初と最後の頁 1-13
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/e23020140	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 N. Eshima, M. Tabata, C. G. Borroni	4. 巻 20
2. 論文標題 An Entropy-Based Approach for Measuring Factor Contributions in Factor Analysis Models	5. 発行年 2018年
3. 雑誌名 entropy	6. 最初と最後の頁 1-12
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/e20090634	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 F. Honda and T. Kurosawa	4. 巻 8
2. 論文標題 Bias reduction of a conditional maximum likelihood estimator for a Gaussian second-order moving average model	5. 発行年 2021年
3. 雑誌名 Modern Stochastics : Theory and Applications	6. 最初と最後の頁 435-463
掲載論文のDOI（デジタルオブジェクト識別子） 10.15559/21-VMSTA187	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計21件（うち招待講演 2件 / うち国際学会 8件）

1. 発表者名 尾関伸之, 黒沢健
2. 発表標題 Type II 打ち切りデータに対するベイズ予測分布の許容性
3. 学会等名 第34 回日本計算機統計学会大会(オンライン)
4. 発表年 2020年

1. 発表者名 尾関伸之, 黒沢健
2. 発表標題 Type II 打ち切りデータに対するベイズ予測分布の被覆確率及び未観測値の推定bias
3. 学会等名 2020 年度統計関連学会連 合大会(オンライン)
4. 発表年 2020年

1. 発表者名 尾関伸之, 黒沢健
2. 発表標題 Statistical properties of the best estimator for type II censored data
3. 学会等名 統計科学研究部門第4 回統計科学セミナー(オンライン)
4. 発表年 2021年

1. 発表者名 金澤峻太郎, 黒沢健
2. 発表標題 一般共変量を伴うポアソン回帰モデルの 回帰相関係数の推定量の漸近正規性
3. 学会等名 第33 回日本計算機統計学会シンポジウム
4. 発表年 2019年

1. 発表者名 Takeshi Kurosawa, Naoto Suzuki
2. 発表標題 A new Liu estimator under a linear constraint
3. 学会等名 International Biometric Society Australasian Region Conference (国際学会)
4. 発表年 2019年

1. 発表者名 江島伸興
2. 発表標題 エントロピーに基づく共通因子空間の分析
3. 学会等名 2020年度 統計関連連合大会
4. 発表年 2019年

1. 発表者名 T. Kurosawa, K. Shinmura, N. Eshima
2. 発表標題 Entropy correlation coefficient for the poisson regression model
3. 学会等名 Three-day atms work- shop Multi- and high-dimensional statistics - Copulas - Survival analysis - Model selection (国際学会)
4. 発表年 2018年

1. 発表者名 西航平, 黒沢健
2. 発表標題 順序統計量に対する事後予測分布の優越性評価
3. 学会等名 2018 年度統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 鈴木尚人, 黒沢健
2. 発表標題 制約条件下の新Liu 型推定量
3. 学会等名 計算機統計学会
4. 発表年 2018年

1. 発表者名 T. Kurosawa, K. Shinmura, F. Hui, A. Welsh, N. Eshima
2. 発表標題 Predictive power measures for the poisson regression model
3. 学会等名 Fellows Research Meetings (Goulburn 35) (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 T. Kurosawa, K. Nishi
2. 発表標題 Dominance of posterior predictive densities over plug-in densities for order statistics
3. 学会等名 Seminar 1 in 2019 at Research School of Finance, Actuarial Studies & Statistics (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 黒沢健, 井上明也, 岩下基
2. 発表標題 ロジスティック回帰モデルに対する 回帰相関係数とスパムメール判定への応用
3. 学会等名 日本・オペレーションズリサーチ学会
4. 発表年 2019年

1. 発表者名 T. Kurosawa, K. Shinmura, F. K. C. Hui, A. H. Welsh, N. Eshima
2. 発表標題 Estimators of Goodness-of-fit Measures for a Poisson Regression Model
3. 学会等名 ISI WSC 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 江島伸興
2. 発表標題 因子分析におけるエントロピーに基づく因子効果評価：試験データ分析への応用
3. 学会等名 日本行動計量学会
4. 発表年 2018年

1. 発表者名 小谷直己, 黒沢健
2. 発表標題 ロジスティック回帰モデルにおける分離問題とモデル評価尺度
3. 学会等名 応用統計学会2021年度年会
4. 発表年 2021年

1. 発表者名 Takeshi Kurosawa, Kouhei Nishi, Nobuyuki Ozeki
2. 発表標題 Dominance of posterior predictive densities over plug-in densities for order statistics
3. 学会等名 Australian and New Zealand Virtual Statistical Conference (国際学会)
4. 発表年 2021年

1. 発表者名 小谷直己, 黒沢健
2. 発表標題 分離データに対するロジスティック回帰のモデル尺度の推定法
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 Naoki Kotani and Takeshi Kurosawa
2. 発表標題 A goodness-of-fit measure for logistic regressions under separation
3. 学会等名 IASC-ARS2022 (国際学会)
4. 発表年 2022年

1. 発表者名 和田賢太郎, 黒沢健
2. 発表標題 EIVを伴うポアソン回帰モデルのnaive推定量
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 和田賢太郎, 黒沢健
2. 発表標題 EIVを伴うポアソン回帰モデルの修正Naive推定量の近似
3. 学会等名 第35回日本計算機統計学会シンポジウム
4. 発表年 2021年

1. 発表者名 Kentarou Wada, Takeshi Kurosawa
2. 発表標題 The naive estimators of a Poisson regression model with measurement errors
3. 学会等名 IASC-ARS2022 (国際学会)
4. 発表年 2022年

〔図書〕 計3件

1. 著者名 Nobuoki Eshima	4. 発行年 2019年
2. 出版社 Springer	5. 総ページ数 257
3. 書名 Statistical Data Analysis and Entropy	

1. 著者名 John M. Lachin、宮岡 悦良(監訳)、遠藤 輝(訳)、黒沢 健(訳)、下川 朝有(訳)、寒水 孝司(訳)	4. 発行年 2020年
2. 出版社 共立出版	5. 総ページ数 704
3. 書名 医薬データのための統計解析	

1. 著者名 Nobuoki Eshima	4. 発行年 2021年
2. 出版社 Springer	5. 総ページ数 190
3. 書名 An Introduction to Latent Class Analysis - Methods and Applications	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	江島 伸興 (Eshima Nobuoki) (20203630)	久留米大学・医学部・客員教授 (37104)	

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 協力者	アラン ウェルッシュ (Welsh Alan)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
オーストリア	The Australian National University			
イタリア	University of Milano Bicocca			