様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　１３　日現在

機関番号：　１２５０１
研究種目：　基盤研究(C)（一般）
研究期間：　2013～2016
課題番号：　２５３３００３４

研究課題名（和文）擬似尤度に基づく情報量基準の構築と過分散を持つ離散データの解析への応用

研究課題名（英文）Information criterions based on quasi-likelihood with application to over-dispersion problems

研究代表者

汪　金芳（Wang, Jinfang）

千葉大学・大学院理学研究科・教授

研究者番号：１０２７０４１４

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）：(1)確率分布の代わりに、平均と分散構造のみを仮定するセミパラメトリックモデルを提案し、それに基づく一般化線形モデルの拡張を行った。また拡張されたセミパラメトリック回帰モデルにおけるモデルの選択法を提案した。（２）定理証明支援系 Coq とその拡張である SSReflect を用いて，Wang (2010) で提案された代数系 Cain、とそれに基づく条件付き独立性の形式化を行った。（３）要約表基づく「セル回帰分析」の手法を開発した。また、セル回帰から得られる事後分布をとより詳細なデータを用いて、新しいベイズ予測の手法を提案した。

研究成果の概要（英文）：(1)We proposed some new model selection criterions for semi-parametric regression models. These models only use the assumptions on mean and variance functions instead of the full parametric assumptions as used in traditional generalized linear model. (2)We formalized the theory of cain (2010) using the interactive theorem-prover Coq/SSReflect. As a consequence, we formalized the theory on conditional independence based on cain. (3)We proposed a new Bayesian prediction method by combining some independent source data with the target detailed data. For the summary source data in the form of tables, we proposed cell regression methods based on integrated predictive probabilities.

研究分野：　数理統計学

キーワード：　一般化線形モデル　conditional independence　cain　quasi-likelihood　causal inference　Bayesian inference　coq/SSReflect　cell regression

１．研究開始当初の背景

統計学の歴史は回帰分析の歴史といっても過言ではない．しかし，計算的観点や数学的な解析の難しさから，20 世紀の 70 年代の前半までに回帰分析における誤差の分布について，主に正規分布という不自然な仮定が置かれている．一般化線型モデル（GLM）は，正規分布の仮定を大幅に緩和し，誤差分布を，正規分布やガンマ分布などの連続型分布を始め，二項分布やポアソン分布などの離散型分布まで，全ての指数型分布族に属する確率分布を解析の対象とすることが可能となり，GLM は統計学の発展史上で最も重要な進展の 1 つとして挙げられている．

しかし，社会科学や自然科学における多くのデータが，GLM で規定されている指数分布族に従わないことも多い．その重要な例として，多くの 2 値データや計数データは，従来用いられてきた二項分布やポアソン分布に従わないことがよく知られている．例えば，トリヴァース＝ウィラード仮説が主張するように，息子を出産しやすい母親と，娘を出産しやすい母親が別れており，男の子が生まれる人数は決して二項分布には従わない．したがって，性比データを解析するときには，二項分布に基づく通常の logistic 回帰分析では，誤った結論を導いてしまう恐れがある．それは母集団の不均一性により，過分散という現象が起きているからである．

このような現実的な要請から，GLM の枠組みを拡張する研究が多くなされてきた．その代表的なものが，確率分布を仮定せずに，平均と分散の関係のみを指定し，擬似尤度を構築する方法である．過分散データを解析するための擬似尤度の有効性について多くの研究がなされている．しかし，これらの論文は，いずれも回帰パラメータの推定や検定などに関する研究であり，適切なモデルを選択する方法については議論されていない．本研究の代表者はこれまでに，解析力学の手法を導入し，任意の平均と分散の仮定に基づく擬似スコアから出発して，擬似尤度関数の構築を中心に，多くの研究成果が得られている．本研究の主な目的は，赤池氏が用いた情報理論的なアプローチをセミ・パラメトリック推測の場合に拡張し，擬似スコアに基づく情報量規準の開発を目指す．

２．研究の目的

一般化線型モデルは，それまでに回帰分析において欠かせなかった正規分布の仮定を緩和し，統計学の発展史上で最も重要な進展の 1 つである．しかし，現実における多くのデータは，例えば，過分散（overdispersion）などの現象が起き，GLM で規定される指数型分布族に従わないことも多い．そのため，これまでに GLM の枠組みを拡張する研究が多くなされてきた．その１つが，確率分布の仮定をせずに，平均と分散の関係のみを指定することにより，擬似尤度と呼ばれるものに基づいて，推定や検定などを行う方法である．しかし，確率分布を仮定しないため，統計モデルを選択するための方法の構築が格段に難しくなり，擬似尤度に基づくモデル選択法に関する研究はこれまでに殆どなされていない．本研究の目的は，平均と分散に関する情報のみに基づく，セミパラメトリック・情報量基準を構築し，特に過分散を持つ離散データの解析への応用を目指す．

３．研究の方法

（１） 擬似尤度に基づく情報量基準の構築においては、理論的研究とシミュレーションを用いた。理論的手法としては、擬似尤度に基づくセミパラメトリックモデルの枠組みで、モデル選択の問題を凸解析の問題として定式化した。

（２） 条件付き独立性の形式化の研究においては、主に定理証明支援系 Coq とその拡張である SSReflect を用いた。

（３） ベイズ的予測問題の研究においては、セル解析の新しい解析法を提案した。

４．研究成果

(1) 複雑なデータ解析において伝統的なパラメトリックモデルの指定が難しい場合が多い。過分散が伴う計数データの解析がこのような典型例である。本研究では、確率分布の代わりに、平均と分散構造のみを仮定するセミパラメトリックモデルを提案し、それに基づく一般化線形モデルの拡張を行った。本研究の主な貢献は、このように拡張されたセミパラメトリック回帰モデルにおけるモデルの選択法の提案である。セミパラメトリックモデルの枠組みでは、確率分布を仮定しないため、Kullback-Leibler 情報量を直接計算できない。本研究では、モデル選択の問題をある種の変分問題として解決した。シミュレーションとデータ解析を通して、提案するモデル選択法の有効性を確認した。

(2) 画像診断法では被験者の病気の有無をしばしば複数の読影者による判断が行われる。本研究では複数の読影者から得られたクラスターデータに基づいて、2 つの画像診断法に対する非劣性検定を提案した。コンピュータシミュレーションなどを行い、名目上の有意水準に概ね達していることを確認し、また急性くも膜下出血患者に対

して実施した動脈瘤診断法から得られたデータに本提案手法を適用し、その有効性を確認した。

(3) 病気の有無に関する画像診断はしばしば複数の読影者により行われる。これまで、コンセンサスや多数決によって、1 人の読影者に帰着させ、診断法の感度と特異度の推定を行われている。本研究では、複数の読影者による評価データを統合するための多次元変量効果モデルを提案し、それに基づいた感度と特異度の推定法や、感度と特異度の同時信頼区間の構築法を提案した。シミュレーションや、アルツハイマー・データへの適用を通して、提案手法の有効性を確認した。

(4) 定理証明支援系 Coq とその拡張である SSReflect を用いた，Wang(2010) で提案された代数系 Cain、とそれに基づく条件付き独立性の形式化を行った。

(5) 連続変量の離散化を行い、頻度として表に纏めることがしばしばある。このような疎表に基づく「セル回帰分析」の手法を開発した。また、セル回帰から得られる事後分布をとより詳細なデータを用いて、新しいベイズ予測の手法を提案した。提案された手法を糖尿病予測の問題に適用し、その有効性を確認した。

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 5 件）

1. Saeki, H., Tango, T. and Wang, J. (2017). Statistical Inference for Non-inferiority of Difference in Proportions of Clustered Matched-pair Data from Multiple Raters, Journal of Biopharmaceutical Statistics, 27(1): 70-83. 査読有り, DOI: 10.1080/10543406.2016.1148709

2. R. Yamaguchi, K. Kin, S. Shimoyama, M. Hagiwara, M. Yamamoto and J. Wang (2017), Formalization of the Conditional Independence using Coq/SSReflect, Technical Reports of Mathematical Sciences, Chiba University, Volume 29 (1), 1-40. 査読なし http://www.math.s.chiba-u.ac.jp/report/files/17001.pdf

3. Saeki, H., Tango, T. and Wang, J. (2016), Estimating the diagnostic accuracy from multiple raters based on a bivariate random effects model, 計量生物学，37 巻 1 号，p.23-44.査読有り https://www.jstage.jst.go.jp/article/jjb/37/1/37_23/_article/-char/ja/

4. 汪 金芳（2014），RSS/JSS ——英国王立統計学会との共同認定，「統計学ガイダンス」，日本統計学会・数学セミナー編集部編. 82—85, 日本評論社. 査読なし https://www.nippyo.co.jp/shop/book/6582.html

5. Jinfang Wang (2013). Statistical disclosure control using the epsilon-uncertainty intervals and the grouped likelihood method. 経済系, Vol. 258, 37 --47. 査読有り http://ci.nii.ac.jp/search?q=AN00302437

〔学会発表〕（計 26 件）

1. Wang, J. (2017), Cell Regression and Reference Priors,科研費シンポジウム「統計的モデリングと計算アルゴリズムの数理と展開」, 2017 年 2 月 18 日(土)～19 日，名古屋大学・情報科学研究科（愛知県・名古屋市）

2. Jinfang Wang, Minoru Arai, Shigeru Kobayashi, Masami Sumi, Shigetoshi Hosaka (2017). On posterior predictive probabilities of diabetes based on comprehensive medical examination data, 2017 年度日本計量生物学会年会，中央大学後楽園キャンパス，2017 年 3 月 16 日（東京都・文京区）

3. 佐伯浩之 ・丹後 俊郎・汪金芳 (2016). 複数の評価者による対応のあるクラスターデータの割合の差の非劣性に関する統計的推論, 統計関連学会連合大会, 2016 年 9 月 5 日, 金沢大学（石川県・金沢市）

4. 汪金芳 (2016). Cell regression, 科学研究費基盤研究(S) 「計算代数統計による統計と関連数学領域の革新」による研究集会,「数理統計ひこね 2016」滋賀大学彦根キャンパス本部棟 3 階大会議室, 2016 年 12 月 2 日 ～ 12 月 3 日（滋

賀県・彦根市）

5. Jinfang Wang (2016). Bayesian prediction based on profile-reference data, COMPSTAT 2016, The 22nd International Conference on Computational Statistics, Auditorium/Congress Palace Principe Felipe, 23-26 August 2016, Oviedo（Spain）

6. Jinfang Wang (2015). Big Math Data, Invited talk at The Department of Statistics, Ewha Women's University, 2015.5.8, Seoul（Korea）

7. Jinfang Wang (2015). Big Math Data: possibilities and challenges, Invited talk at The, Dept of Mathematical Sciences, KAIST, 2015.8.5, Deajeon（Korea）

8. Jinfang Wang (2015). Formalization of probabilistic conditional independence using Coq/SSReflect. SIAM Conference on Applied Algebraic Geometry, August 6, CAMP/NIMS, Deajeon（Korea）

9. Hiroyuki Saeki, Toshiro Tango and Jinfang Wang (2015), Nonparametric confidence intervals for sensitivity and specificity from multiple raters, 2015.7.31, 60th ISI World Statistics Congress, Rio de Janeiro（Brazil）

10. Jinfang Wang (2015), Cain algebra for probabilistic conditional independence, Workshop on Formalization of Applied Mathematical Systems, Joint Conference of Chiba University and the University of Hawai`i, September 25 to October 2, 2015 at the University of Hawai`i, Manoa (US)

11. Jinfang Wang (2015), Formalization of cain using Coq/SSReflect, Workshop on Formalization of Applied Mathematical Systems, Joint Conference of Chiba University and the University of Hawai`i, September 25 to October 2, 2015 at the University of Hawai`i, Manoa (US)

12. **佐伯浩之**, **丹後俊郎**, 汪金芳 (2014). **変量効果モデルを用いた複数の読影者による画像診断法の精度の推定**, **京都大学数理解析研究所** RIMS **共同研究**「Asymptotic Statistics and Its Related Topics」, 2014 **年** 3 **月** 3 **日**--5 **日**, **京都大学（京都府・京都市）**

13. **湯毅平**, 汪金芳 (2014). **擬似尤度に基づくモデル選択法と過分散データへの応用**, **京都大学数理解析研究所** RIMS **共同研究**「Asymptotic Statistics and Its Related Topics」, 2014 **年** 3 **月** 3 **日**--5 **日**, **京都大学（京都府・京都市）**

14. **汪金芳**, **萩原学**, **山本光晴** (2014). Interactive theorem proving of probabilistic conditional independence relations using Coq/SSReflect, 2014 **年統計関連連合大会**, 2014 **年** 9 **月** 14 **日**, **東京大学（東京都・文京区）**

15. **佐伯浩之**, **丹後俊郎**, 汪金芳 (2014). **変量効果モデルを用いた複数の読影者による画像診断法の精度の推定**, **京都大学数理解析研究所** RIMS **共同研究**「Asymptotic Statistics and Its Related Topics」**数理解析研究所講究録** 1910, 1-19, 2014 **年** 3 **月** 3 **日**-5 **日**, **京都大学（京都府・京都市）**

16. **湯毅平**, 汪金芳 (2014). **擬似尤度に基づくモデル選択法と過分散データへの応用**, **京都大学数理解析研究所** RIMS **共同研究**「Asymptotic Statistics and Its Related Topics」, **数理解析研究所講究録** 1910, 20-28, 2014 **年** 3 **月** 3 **日**-5 **日**, **京都大学（京都府・京都市）**

17. **汪金芳・萩原学・山本光晴** (2014). Formalization of statistical conditional indenpendence relations using Coq/SSReflect, **科学研究費によるシンポジウム「多様な分野における統計科学の教育・理論・応用の新展開」**, 2014.10.25, **新潟大学（新潟県・新潟市）**

18. **萩原学**, **久我健一**, **松田茂樹**, **桜井貴文**, 汪 金芳,**山本光晴** (2014), Big Math Data: meeting the challenges of analyzing mathematical sciences, **科学研究費によるシンポジウム**「Workshop on Statistical Methods for Large Complex Data」, 2014.11.11, **筑波大学（茨城県・つくば市）**

19. J. Wang (2014) (Invited Lecture), Model Selection Based on Quasi-likelihood with Application to Overdispersed Data, May 2, 2014, Ewha Womans University, Seoul (Korea)

20. H. Saeki, T. Tango and J. Wang (2014). Estimating the accuracy of diagnostic imaging based on multiple raters using random effects model. 27th International Biometric Conference, July, 6-11, 2014, Florence (Italy).

21. J. Wang (2014), Big Math Data: embracing the challenges from mathematical sciences, Kyoto International Conference on Modern Statistics in the 21st Century, November 17, 2014, Kyoto International Conference Center（京都府・京都市）

22. Wang, J. and Y. Tang (2013). Model Selection for semiparametric Bayesian models with application to overdispersion, 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong.

23. 佐伯浩之，丹後俊郎，汪金芳 (2013). 複数の読影者による対応のあるクラスターデータの割合の差の信頼区間，2013 年度統計関連学会連合大会，2013 年 9 月 8 日 ～ 11 日，大阪大学（大阪府・吹田市）

24. 佐伯浩之，丹後俊郎，汪金芳 (2013). 変量効果モデルを用いた複数の読影者による画像診断法の精度の推定，科学研究費シンポジウム「一般化線形モデルの最新の展開とその周辺」，2013 年 11 月 8 日−10 日，千葉大学（千葉県・千葉市）

25. 湯毅平，汪金芳 (2013). Information criterion based on quasi-likelihood with application to over-dispersed data, 科学研究費シンポジウム「一般化線形モデルの最新の展開とその周辺」，2013 年 11 月 8 日−10 日，千葉大学（千葉県・千葉市）

〔図書〕（計 1 件）

汪 金芳 (2016)，一般化線形モデル （統計解析スタンダード），朝倉書店（212 頁）

〔産業財産権〕

○出願状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等
http://www.math.s.chiba-u.ac.jp/~wang/

６．研究組織

(1)研究代表者
汪 金芳(Jinfang Wang)
千葉大学・大学院理学研究科・教授
研究者番号:10270414

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：

(4)研究協力者
（　　）