

令和元年6月7日現在

機関番号：62603

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16014

研究課題名(和文)レアイベントデータに対する統計的推測と臨床研究への応用

研究課題名(英文)Statistical inference for rare event data and its application in clinical researches

研究代表者

長島 健悟(Nagashima, Kengo)

統計数理研究所・医療健康データ科学研究センター・特任准教授

研究者番号：20510712

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：レアイベントデータにおけるCox回帰モデルの推測手法として用いられている、Heinze & Schemperの方法の理論的な特性について検討し、キュムラントにおける条件が成り立つことの証明、回帰パラメータの推定量の一致性の証明、理論的に妥当なAIC型およびBIC型のモデル選択規準の提案を行った。本研究の成果については、学術論文誌へ掲載され、また学会発表を行い、研究概要を解説したウェブページを公開済みである。また、生存時間を評価指標としたときのサンプルサイズ設計手法について検討し、頻用されている既存法の計算式に誤りがあることを示し、妥当な計算式を提案した。この結果は学術論文誌へ投稿中である。

研究成果の学術的意義や社会的意義

本研究では、特にレアイベントデータにおけるCox回帰モデルにおける、各種統計量の漸近的性質等を明らかにし、新たな手法を提案した。これらの結果は同様の条件下でのさらなる応用・手法開発に繋がると考えられ、本研究の学術的意義は高いと考えられる。また、実際の臨床研究データへの適用や統計ソフトウェアの公開など、手法を実際に利用する際に活用できる参考情報も提供した。将来的には、臨床医学分野での共同研究を通じ、既存の方法では妥当な評価ができないケースにおける応用を目指す。臨床研究の成果として新たな知見が得られれば、実社会に貢献することが可能であり、意義深い結果をもたらすことが期待される。

研究成果の概要(英文)：Information criteria for Firth's penalized partial likelihood approach in Cox regression models were proposed and theoretical justification of the Heinze and Schemper's estimator was provided (Nagashima & Sato, Statistics in Medicine 2017, DOI: 10.1002/sim.7368). An AIC-type information criterion based on the risk function and BIC-type criterion were evaluated, which performed well under rare event data. The proposed AIC-type criterion was applied to prospective observational study data. It proved that the conditions of bias reduction for the cumulants in the Heinze & Schemper's estimator holds. Moreover, sample size calculations for the Kaplan-Meier estimator in single-arm survival studies were proposed, and it was shown that the existing method uses an inappropriate approximation that can influence the accuracy especially under rare event data. This result was submitted to a peer-review journal.

研究分野：生物統計学

キーワード：生存時間解析 情報量規準 サンプルサイズ設計 モデル選択規準 レアイベントデータ

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

レアイベントデータとは、ある着目したイベント（例えば死亡）の発生頻度が極端に小さくなる構造を持つデータであり、様々な統計学的問題が生じうる。具体的には、Cox 回帰モデルを用いた生存時間解析において monotone likelihood と呼ばれる問題が生じ、回帰パラメータの推定値とその標準誤差が無限大に発散し、妥当な推測を行う事が不可能となることがよく知られている [1, 3]。例えば、二種類の治療法の生存率を比較する臨床試験において、一方の治療効果が非常に大きくイベントがほとんど発生しない場合には、治療効果の評価が困難になるという問題が生じる。Heinze & Schemper [3] はレアイベントデータの統計解析手法として、Cox 回帰モデルに Firth のバイアス修正法 [2] を適用した罰則付き部分尤度法、罰則付き部分尤度比検定、および罰則付きプロファイル部分尤度信頼区間を提案し、monotone likelihood の状況においても推測が可能であることを示した。しかし、これらの手法は経験的に提案され、理論的な評価がなされていなかった。そのため、この条件下でのモデル選択規準も研究されておらず、罰則付き部分尤度関数に基づいた経験的な情報量規準が用いられていた。また、レアイベントを考慮したサンプルサイズ設計の方法も議論されておらず、方法論の整備が遅れている状況であった。さらに、Cox 回帰モデルにおいては、層ごとに異なるベースラインハザード関数を仮定した層別モデルがあるものの、レアイベントデータのもとでの推測手法は今までに研究されていなかった。

2. 研究の目的

生存時間解析におけるレアイベントデータの問題は、実際の臨床研究においてもしばしば起こる問題であるにもかかわらず、経験的に提案された既存法が適用され続けてきた。しかし、予備的な研究から、実際に妥当でない情報量規準が用いられていることが明らかになった。この情報量規準が用いると、モデルの適切さを無視し、候補モデル集合の中でパラメータ数が最大のモデルを高い確率で選択することがわかった。したがって、理論的な評価が不十分である現状を解決する必要があると考えられた。このような背景から、本研究ではレアイベントデータに対する理論的に妥当な統計手法を開発し、臨床研究へ応用することを目的とした。

具体的には、(1) 罰則付き部分尤度に基づく仮説検定と区間推定法の改良、(2) レアイベントデータにおける Cox 回帰モデルの妥当なモデル選択規準の開発、(3) レアイベントを考慮したサンプルサイズ設計の方法、(4) レアイベントデータにおける層別 Cox 回帰モデルの統計的推測手法の開発を行うこととした。

3. 研究の方法

本研究では、まず既存法である罰則付き部分尤度法の理論的特性を検討する。既存法は通常の Cox 回帰モデルの拡張であるため、Andersen & Gill [4] の結果を活用する。この結果から (1) の罰則付き部分尤度に基づく仮説検定と区間推定法の改良について検討を行う。

(2) レアイベントデータにおける Cox 回帰モデルの妥当なモデル選択規準の開発として、予備的な研究で明らかとなった妥当でない情報量規準の理論的な問題点を改善し、妥当な情報量規準の提案する。さらに、他のモデル選択規準への拡張についても検討を行う。

(3) レアイベントを考慮したサンプルサイズ設計の方法については、本研究では単群の生存時間解析の場合を検討する。予備的な検討では、現在非常に広く用いられている既存法がある大標本近似に基づいており、この近似の方法に問題がある事がわかっている。レアイベントの状況では小標本の状況に近くなるため、近似精度が非常に悪くなる可能性がある。本研究では、別の近似手法を用いることで、性能を改善することを示す。

層別 Cox 回帰モデルにおける仮説検定と信頼区間については、通常の Cox 回帰モデルと同様の議論が成り立つ事が知られている [5]。したがって、(4) レアイベントデータにおける層別 Cox 回帰モデルの統計的推測手法については、仮説検定と信頼区間については前述の (1) の結果が応用できる見込みである。層別 Cox 回帰モデルにおけるモデル選択規準は未解決の問題である。セミパラメトリックモデルのモデル選択規準については既存の研究があり、その結果を応用するアプローチにより新たな手法を開発する。

4. 研究成果

(1) および (2) に関する成果として、罰則付き部分尤度法の理論的特性を検討した結果、主に 3 つの結果を得た。1 つめは Andersen & Gill が与えた正則条件のもとで、Heinze & Schemper の方法による罰則付き部分尤度法について、回帰パラメータ推定量のバイアス減少のためのキムラントに関する条件を満たすことを示した。2 つめは同じ正則条件のもとで、上記の推定量に一致性があることを証明した。3 つめは、予備的な研究で明らかとなった妥当でない情報量規準の理論的な問題点を示し、その問題点を修正する方法を提案し、妥当な AIC 型および BIC 型の情報量規準を導出した。提案した AIC 型の情報量規準を実際の臨床データに適用し、臨床的にも説明性の高いモデル選択が可能であることを示した。これらの結果は査読付論文として出版済みである（雑誌論文）。提案した方法を適用可能なスクリプトを作成し、論文の補足資料として公開した（雑誌論文の補足資料）。論文はオープンアクセスとなっており、誰でも閲覧できる。また、研究代表者のホームページ上で提案した方法を日本語で解説した資料を公開した（その他）。

(3)に関する成果としては、現在非常に広く用いられている既存法[6]の問題点を明確にし、理論的な問題点を修正するサンプルサイズ計算式を提案した。また、大規模なシミュレーションの結果から、提案した方法は幅広い条件下で妥当なサンプルサイズを与えることを示した。一方、既存法は過小なサンプルサイズを与える傾向があった。この結果は学術雑誌に投稿中であり、現在査読中である。提案したサンプルサイズ計算式を適用可能なソフトウェアを開発し、誰でも自由に利用できるように公開した(その他)。アウトリーチ活動として、一般向けのイベントにおいて研究内容の紹介を行った(その他)。

(4)については現在未解決の問題になっている層別 Cox 回帰モデルにおけるモデル選択規準について検討を進めており、セミパラメトリックモデルのモデル選択規準を層別 Cox 回帰モデルの場合に応用する方法を検討し、シミュレーションを実施している。より具体的な成果が出た段階で、結果をまとめて学術雑誌に投稿する予定である。

<引用文献>

- [1] Bryson MC, Johnson ME. The incidence of monotone likelihood in the Cox model. *Technometrics* 1981; **23**: 381-383.
- [2] Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993; **80**: 27-38.
- [3] Heinze G, Schemper M. A solution to the problem of monotone likelihood in Cox regression. *Biometrics* 2001; **57**: 114-119.
- [4] Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* 1982; **10**: 1100-1120.
- [5] Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag, 1993.
- [6] Cancer Research and Biostatistics. One arm survival power/sample size calculator, 2002. Accessed: April 10, 2019.
<https://stattools.crab.org/Calculators/oneNonParametricSurvivalColored.htm>

5. 主な発表論文等

〔雑誌論文〕(計1件)

Nagashima K, Sato Y. Information criteria for Firth's penalized partial likelihood approach in Cox regression models. *Statistics in Medicine* 2017; **36**(21): 3422-3436. DOI: 10.1002/sim.7368. (査読有)
<https://doi.org/10.1002/sim.7368>

〔学会発表〕(計1件)

長島健悟. Information criteria for Firth's penalized partial likelihood approach in Cox regression models. 統計数理研究所リスク解析戦略研究センター 第9回 生物統計ネットワークシンポジウム, 2018.
<https://nshi.jp/research/files/netsympo180319.pdf> (発表資料)

〔図書〕(計0件)

なし

〔産業財産権〕

出願状況(計0件)

なし

取得状況(計0件)

なし

〔その他〕

ホームページ

Sample size calculation for one sample non-parametric survival.
<https://nshi.jp/en/js/onesurvyr/>
<https://nshi.jp/contents/js/onesurvyr/>
Firth の方法を用いた場合のモデル選択規準について。
<https://nshi.jp/contents/stat/firthic/>

アウトリーチ活動

長島健悟. Kaplan-Meier 推定量に基づく一標本検定のサンプルサイズ設計. 統計数理研究所オープンハウス, 2019. (ポスター発表)

6. 研究組織

(1) 研究分担者

なし

(2) 研究協力者

なし

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。