

令和 5 年 6 月 13 日現在

機関番号：32660

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19762

研究課題名（和文）生存時間と打ち切り時間の依存性を考慮した解析法の構築

研究課題名（英文）Construction of analytical method for dependent censoring

研究代表者

下川 朝有（Shimokawa, Asanao）

東京理科大学・理学部第二部数学科・講師

研究者番号：80756297

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究では、生存時間と打ち切り時間の依存関係を考慮したモデル構築法の提案を目的に研究を行った。主要な成果としては、コピュラ関数に含まれる変数間の依存度合いを表すパラメータの統計的推測法、コピュラ関数を用いたノンパラメトリックなモデル構築法、変数間の依存を考慮した共変量を含むモデルの構築法について、幾つかの手法を提案し、その性質について理論的、またシミュレーションを用いて検証した。

研究成果の学術的意義や社会的意義

生存時間解析における一般的な手法は、生存時間と打ち切り時間の独立性を仮定しており、その依存性に関してモデルを検討する主要な手法は感度分析に限られている。しかしながら実際のデータ解析を行う状況によっては、この独立性の仮定自体が疑わしい場合が多く存在する。本研究ではセミパラメトリック及びノンパラメトリックな枠組みにおける、依存打ち切りを前提としたモデル構築法を提案した。このようなモデル構築法は他に提案されておらず、独立性が仮定できない打ち切りが発生することが多い医療データ解析等の分野において特に有意義であると考えられる。

研究成果の概要（英文）：In this research, the purpose was to propose a model construction method that takes into account the dependence between survival time and censoring time. As the main result, we proposed several methods to estimate the parameters expressing the degree of dependence between variables in copula functions. Moreover, we proposed the model construction methods using copula functions and covariates.

研究分野：統計学

キーワード：生存時間解析 依存打ち切り 木構造 ランダムフォレスト

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

打ち切りデータを利用して生存時間の解析を行う場合、イベント発生までの時間(生存時間)を表す確率変数と、打ち切り発生までの時間(打ち切り時間)を表す確率変数の間に何かしらの仮定を置く必要がある。何故ならこれは、各被験者(各標本)は生存時間もしくは打ち切り時間の片方が観測された時点で他方の時間は観測されないため、標本からこれら確率変数間の関係性を特定することは不可能である、という事実に起因する。そこで広く用いられる仮定として、これらの独立性の仮定があげられる。すなわちがんの再発の例で述べると、着目する再発までの時間(生存時間)と、何らかの理由で患者が追跡不可能になるまでの時間(打ち切り時間)は互いに依存せず、独立であるという仮定である。この仮定の下で多くの生存時間解析の手法が提案されており、代表的なものをあげると Kaplan-Meier 推定や Cox 回帰等があげられる。

あるイベントの生存時間に着目する研究において、もしその研究期間が事前に定まっており、全ての被験者がその期間内の最終時間まで観測されるとするならば、生存時間と打ち切り時間は独立であると考えることが出来る。しかしながら多くの医療データには、その他様々な理由による打ち切りが含まれる。それらを考慮した一意なモデルの構築は、医療研究を始めとする多くの生存時間を扱う研究において重要な課題であると考えられる。

コンピュータを用いてこれらの依存性を表す研究は主に、パラメトリックなモデルを前提とした感度分析による研究において行われている。しかし現実には一意なモデルの確立は多くの場面において広く望まれる。また一般的な生存時間データの共変量を含む解析においては、Cox 回帰が広く使われているが、医療データの解析においてはその解釈の容易性、また交互作用を自然に考慮可能であるため木構造モデルが扱われることも多く、そのようなモデルの構築は有意義であると考えられた。

### 2. 研究の目的

本研究では、「生存時間と打ち切り時間の依存性を考慮したモデル構築法の確立とその有意性の検証」を目的とした。生存時間を扱う解析の多くは、生存時間と打ち切り時間の独立性を仮定するが、この仮定が現実のデータにおいて成り立つかは疑わしい場合が多く存在する。そのような場合の解析手法としては、変数間の依存性をコンピュータ等の関数を用いてモデル化する手法が存在するが、推定・検定を含むパラメータの統計的推測法、また共変量を含む場合のモデル化において、幾つかの課題を残している。

そこで本研究では「コンピュータ関数に含まれる変数間の依存度合いを表すパラメータの統計的推測法」、「コンピュータ関数を用いたノンパラメトリックなモデル構築法」、そして「変数間の依存性を考慮した共変量を含むモデルの構築法」について、数式を用いた理論的考察、シミュレーションによる比較検証、また実データ解析を通じた有用性の検討によって明らかにしていくことを目的とした。

### 3. 研究の方法

「コンピュータ関数に含まれる変数間の依存度合いを表すパラメータの統計的推測法についての検討」を目標に、生存時間と打ち切り時間を表す確率変数の従う分布にワイブル分布を仮定した下で、Clayton コピュラ及び Frank コピュラを仮定したモデルについて、最尤法にもとづく推定量の導出の検討を行った。続いて、得られる推定量の有限標本下における性質について、予定していた通りシミュレーションを用いた検討を行った。加えて、ブートストラップ法を用いた信頼区間の構築、及び検定についても考察し、シミュレーションによる比較検討も行った。

「コンピュータ関数を用いたノンパラメトリックなモデル構築法」については、生存時間及び打ち切り時間を表す確率変数が従う分布のパラメトリックな仮定を緩めるため、反復法を用いたモデル構築法について検討を行った。具体的に、変数間の依存性を表すパラメータの初期値を与えたもと、生存時間と打ち切り時間を表す分布の Kaplan-Meier 法によるノンパラメトリックな推定を行い、続いて、スプラインによるこれらの分布の補完を行った後、依存性を表すパラメータの推定を行うという、反復法について提案を行った。この手法に関して、シミュレーションによる性能の検証を行った。

「変数間の依存性を考慮した共変量を含むモデルの構築法についての検討」については、まず最初に比例ハザードモデルに基づくモデルについて検討を行った。このモデルについては、部分尤度を用いる手法に代わる新たな推定法が必要なが分かっていたため、従来のモデルよりも強い制限をおいた下で推定する手法について検討した。

木構造モデルに関しては既に応募者等により、生存時間及び打ち切り時間の分布が指数分布に従い、Clayton コピュラもしくは Frank コピュラを用いた場合の理論的構築と、シミュレーションによる性能検証を終えていたが、本研究ではこれをさらに発展させ、より一般化された分布及びコピュラについて研究を行った。具体的には、木構造モデルを構築する際の分割規準の検討、及び過学習を避けるための枝刈りによるモデル縮小法の検討を行い、シミュレーションを通して提案した手法の性質について比較、検討を行った。

加えて、上述した1つ目の目的に対するシミュレーション結果から、コンピュータに含まれる生存時間と打ち切り時間の依存度を表すパラメータの収束は非常に遅いことが分かっており、これ

に対処するため、木構造内に含まれるそのパラメータに対する適切な制限を与える方法についても検討した。具体的には、各共変量空間の分割に対する生存モデルについて、共通のコピュラ関数及び依存度パラメータを与えた場合、また線形関数等で制限を与えたパラメータの使用について、理論的またシミュレーションによる比較検討を行った。

最後に、ここまでの研究で得られたモデル構築手法を利用し、より予測に特化したモデル構築法を提案するため、アンサンブル学習(ランダムフォレスト)によるモデル構築法についても提案し、同様に検討を行った。

#### 4. 研究成果

「コピュラ関数に含まれる変数間の依存度合いを表すパラメータの統計的推測法についての検討」については、指数分布を仮定した下、Clayton コピュラを仮定したモデルについて、最尤推定量の有限標本下における性質がある程度得られた。またより一般化した結果として、生存時間と打ち切り時間を表す確率変数の従う分布としてワイブル分布、コピュラ関数として Frank コピュラを用いた場合に対しても、同様に結果が得られた。

また理論的な考察として、中心極限定理に基づく理論構築を行い、その漸近的な性質を得た。加えて、近似信頼区間および、ブートストラップ法を用いた信頼区間の構築についての性質及び、検定法についても同様に結果を得た。

「コピュラ関数を用いたノンパラメトリックなモデル構築法」については、変数間の依存度合いを表すパラメータと各分布の推定を交互に行う提案法について、シミュレーションによる性能比較を行い、提案法は一定条件下では有効に推定可能であるとの結果を得た。一方で、推定がうまく収束しない状況も存在することが分かり、今後の課題となった。

「変数間の依存を考慮した共変量を含むモデルの構築法についての検討」については、比例ハザードモデルを前提とした手法については、幾つかの制限をおいた下では推測が可能であることが得られた。一方で、実用において与える制限の妥当性については疑問が残る結果となり、今後の課題として残る結果となった。

木構造に基づく手法については、コピュラに含まれるパラメータの収束が非常に遅く、幾つかの制限について考察したが、シミュレーションを通した比較から、生存木の作成における適切な制限を与えることは難しく、予測精度を保つことは難しいことが得られた。特に、枝刈りによる過学習抑制ステップの手法については、課題を多く残す結果となった。

一方で、アンサンブル学習(ランダムフォレスト)を用いた手法においては、制限を与えたとしてもその精度をより高く保てることが得られた。またこの研究の過程で、教師無し学習に対するアンサンブル学習の幾つかの手法を提案することができ、特に異常値検知において有効であるという結果が得られた。

これらの結果を通して、生存時間と打ち切り時間の依存関係を考慮したモデル構築法の幾つかの提案と、それらの性質が得られた。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 錦織 護直、下川 朝有、宮岡 悦良	4. 巻 33
2. 論文標題 アンサンブル法に基づく異常値検知について	5. 発行年 2021年
3. 雑誌名 計算機統計学	6. 最初と最後の頁 77~90
掲載論文のDOI (デジタルオブジェクト識別子) 10.20551/jscswabun.33.2_77	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 -

1. 著者名 Asanao Shimokawa, Etsuo Miyaoka	4. 巻 31
2. 論文標題 Construction of a Survival Tree for Dependent Censoring	5. 発行年 2021年
3. 雑誌名 Journal of Biopharmaceutical Statistics	6. 最初と最後の頁 63-78
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10543406.2020.1792478	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 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 4件）

1. 発表者名 Asanao Shimokawa, Etsuo Miyaoka
2. 発表標題 On survival trees with competing risks
3. 学会等名 CMStatistics 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Asanao Shimokawa, Etsuo Miyaoka
2. 発表標題 Construction of random forest for competing risks
3. 学会等名 31st International Biometric conference (IBC 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Asanao Shimokawa, Etsuo Miyaoka
2. 発表標題 An approach for constructing decision tree using density estimates
3. 学会等名 43rd Annual Conference of the International Society for Clinical Biostatistics (ISCB 43) (国際学会)
4. 発表年 2022年

1. 発表者名 Asanao Shimokawa, Etsuo Miyaoka
2. 発表標題 Construction of random forest based on parametric bootstrapping
3. 学会等名 CMStatistics 2022 (国際学会) (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔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

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