

平成 30 年 6 月 13 日現在

機関番号：10101

研究種目：基盤研究(C) (一般)

研究期間：2013～2017

課題番号：25330031

研究課題名(和文) コピュラを用いた多変量生存時間データの判別分析法とその医学データへの応用

研究課題名(英文) Discriminant analysis of multivariate survival data based on copulas

研究代表者

鈴川 晶夫 (Suzukawa, Akio)

北海道大学・公共政策学連携研究部・教授

研究者番号：00277287

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)：コピュラ(多次元確率分布とその周辺分布を結びつける関数)を用いて多変量生存時間分布をモデル化を行った。このモデルに基づいて、多変量生存時間データの判別分析法を開発した。特に、応用範囲を医学データに絞って、医学研究における実際の多変量生存時間データに対して応用可能な判別分析法を開発した。また、多変量生存時間データ解析においては、データの打ち切りと非多変量正規性が重要な問題となる。打ち切りパターンに対してパラメトリックモデルを構築し、また、コピュラを用いて多変量生存時間分布のモデル化を行うことによって、これらの問題に柔軟に対応可能な判別分析法を開発した。

研究成果の概要(英文)：Multivariate survival time distribution was modeled using copula (a function that connects multidimensional probability distribution and its marginal distributions). Based on this model, we developed a discriminant analysis method of multivariate survival time data. In particular, focusing on the application range to medical data, we developed discriminant analysis methods applicable to actual multivariate survival time data in medical research. In data analysis of multivariate survival time, censoring of data and non-multivariate normality are important problems. We proposed a parametric model for censoring patterns. Based on copula modeling of multivariate survival distributions, we developed flexible discriminant analysis methods.

研究分野：多変量解析

キーワード：生存時間データ 判別分析 コピュラ 打ち切りデータ

1. 研究開始当初の背景

90年代以降、古典的な単変量生存時間解析法(カプラン・マイヤー法やコックス回帰法等)を多変量生存時間データの解析法へ拡張する研究が行われてきた。2000年頃から多変量生存時間データの解析を扱う洋書も複数出版されてきた(Hougaard (2000), Aalen *et al.* (2008), Hanagal (2011)等)。本研究は、分析目的と応用対象を判別と医学データに絞った多変量生存時間解析法研究と位置づけられる。

Sklar (1959)により導入されたコピュラ(copula)は、多次元確率分布とその周辺確率分布を結びつける関数である。様々なコピュラによって多次元確率変数間の従属性を柔軟にモデル化することが可能であり、コピュラは非多変量正規性のもとでの多変量解析研究に有用である。近年、ファイナンスの分野では、複数のリスクの結合分布モデルとしてコピュラが用いられている (Cherubini *et al.* (2004)等)。

研究代表者は、単変量生存時間分布の推測法について研究した(Suzukawa (2002, 2004, 2010)等)。特に、生存時間データの最大の特徴であるデータの打ち切り(censoring)に対処可能な手法の開発を行った。医学データへの応用研究(Sato *et al.* (2008)等)において、多変量生存時間への拡張の必要性を強く感じた。例えば、同じ親から生まれた複数のラットの腫瘍形成時間は多変量生存時間である。

透析患者のカテーテル感染時間データ(McGilchrist and Aisbett (1991))は、打ち切りのある2変量データであり、透析患者は病気のタイプによって4つの群に分類される。このような状況において、ある生存時間データがどの群に属するか不明であるとき、その属する群を判別する分析(判別分析)は応用上重要であるにも関わらず、ほとんど

研究が行われていない。そのような経緯から、観測打ち切りを含む多変量生存時間データに対する判別分析法を開発すべきであるとの認識を強くもつに至った。

2. 研究の目的

本研究は、コピュラ(多次元確率分布とその周辺分布を結びつける関数)を用いた多次元確率分布のモデル化に基づいて、多変量生存時間データの判別分析法を開発し、実際の医学データへの応用における有用性について検討することを目的とする。

開発手法に関して、理論的考察やシミュレーション実験や医学データへの応用を行い、その有効性と問題点を明らかにする。一般的な多変量データと比較して、多変量生存時間データの解析で考慮すべき重要な問題は、データの打ち切りと非多変量正規性の問題である。これらの問題に対処可能であり、また実際の医学データに応用可能な判別分析法を開発することが本研究の目的である。

多変量生存時間分布に対するパラメトリックモデルとセミパラメトリックモデルを整理・構築し、各モデルの特徴を明らかにする。打ち切りを含む多変量生存時間データに基づく未知パラメータ推定法を整理・開発し、判別問題への適用の観点からその性質を明らかにする。ベイズ判別ルール等に基づいて判別手法を具体的に定式化する。定式化された判別手法について、その誤判別率の漸近評価公式を導出し、判別手法の性能を理論的に明らかにする。シミュレーション実験に基づいて判別手法の誤判別率を数値的に評価し、判別手法の性能や特徴を明らかにする。Freireich (1963)の白血病寛解時間データ(白血病患者の寛解時間がメルカプトプリン水和物治療とプラセボのマッチドペア(2変量)で観測され、各患者のペアは白

血病の状態により2群に分類される)等に対して、開発した判別手法を適用し、モデルの妥当性等の観点からその有用性や問題点を明らかにする。

3. 研究の方法

医学研究における実際の多変量生存時間データに対して応用可能な判別分析法を開発することが本研究の目的である。そのために、次の5つの項目を遂行する。多変量生存時間データの解析に関する先行研究について網羅的に整理する。適用範囲が広く、データへの適合性の評価可能な多変量生存時間分布モデルを構築し、データの打切りに対処可能な推定法を開発する。打切りのある多変量生存時間データに対する判別分析法を定式化する。漸近理論やシミュレーション実験に基づいて判別分析法の性能(誤判別率や頑健性)を評価する。実際の医学データに対して判別分析法を適用し、統計学的見地と医学的見地の両面から、その有用性と問題点について検証する。

4. 研究成果

コンピュータを用いた多変量生存時間分布モデルの定式化を行った。本研究に関連する先行研究の整理：打切りのある多変量生存時間データの解析法に関して、Hanley and Parnes (1983), Dabrowska(1988)に始まる多変量打切りデータのノンパラメトリック解析研究と、Clayton (1978), Oakes(1982)に始まる Frailty Modelに基づくパラメトリック解析法について、最新の研究成果まで網羅的に整理した。コンピュータ理論については、(1998)等の文献を参考として、最新の成果までを整理した。また、判別分析法の理論と応用に関するレビューを行った。コンピュータを用いた多変量生存時間分布モデルとその推定法の定式化

として、多変量生存時間データ解析のための統計モデルを、コンピュータを用いて構築した。また、構築されたモデルにおける未知パラメータの推定法を定式化した。

打切りのある多変量生存時間データに対する判別分析法の定式化を行った。多変量生存時間分布モデルの構築とその推測法の開発を行った。主に数理的扱いの容易さ等の技術的観点からモデル構築と推定法の定式化を行った。また、打切りパターンのモデル化に基づく生存時間分布推定法を多変量生存時間分布の推定問題へ拡張した。またその推定法の効率などについて理論的に検討した。2. 打切りのある多変量生存時間データに対する判別分析法の定式化多変量生存時間分布のセミパラメトリックモデルのもとで、ベイズ判別ルールに基づいて判別関数を導出し、その性質についての研究を遂行した。

打切りを含む多変量生存時間データに基づく未知パラメータ推定法を整理・開発し、判別問題への適用の観点からその性質を調べた。ベイズ判別ルール等に基づいて判別手法を具体的に定式化した。シミュレーション実験に基づいて判別手法の誤判別率を数値的に評価し、判別手法の性能や特徴を明らかにした。また、定式化された判別手法について、その誤判別率の漸近評価公式を導出し、判別手法の性質を理論的に検討した。

また、2変量アリキメデスコピュラを用いることにより独立な2変量正規分布に変換し、その変換された変量に基づく判別手法を定式化した。その判別手法の誤判別確率などについて、理論的・数値的検討を行った。その結果、2変量母集団分布にアルキメデスコピュラを仮定できる場合には、この判別手法がベイズ判別手法とほぼ同等の判別性能をもつことが明らかとなった。この手法は、周辺確率分布に対するモデル化を必要とせず、ベイズ判別手法よりも、実際のデータに対して柔軟に適用可能である。

多変量生存時間データの分析において、共有フレイルティモデルが用いられることが多い。共有フレイルティモデルは、アルキメデスコピュラに基づく多変量確率分布モデルである。本研究において、共有フレイルティモデルを一般化した多変量確率分布モデルを定式化した。このモデルは、複数の変量の従属性に関する交換可能性を仮定せず、統計モデルとして柔軟なモデルである。また、このモデルに基づく乱数生成アルゴリズムなどを定式化した。アルキメデスコピュラに基づく判別分析法を開発した。アルキメデスコピュラにより接合される二変量確率分布が独立一様分布に分解されることが知られている。本研究において、この独立分解定理を3変量以上の多変量アルキメデスコピュラに一般化する定理を示した。その定理に基づいて多変量独立正規化スコアを定義した。開発した判別手法は、その多変量独立正規化スコアを用いて判別を行う手法である。誤判別確率に関する理論的考察とシミュレーション実験の結果から、従来の判別手法（線形判別、ロジスティック判別等）に比べて、開発手法が優れた性質をもつことが示された。

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2件)

- Koshiro Yonenaga and Akio Suzukawa. (2018). Product of inverse Wishart and normal distributions. Proceedings of Korea and Hokkaido University 5th Workshop in Statistics, 5-10. (査読無).
- Zhaojun Jian and Akio Suzukawa. (2015). SUPPORT VECTOR MACHINE BASED ON DATA DEPTH. Economic Journal of Hokkaido University, Discussion

Paper Series A. No.282, 1-15. (査読無).

〔学会発表〕(計20件)

- 方宏強, 鈴木晶夫. クラスタサポートベクトルマシンの性質. 2016年度統計関連学会連合大会. 2016年9月. 金沢大学.
- 鈴木晶夫. 一般化共有フレイルティモデル. 2016年度統計関連学会連合大会. 2016年9月. 金沢大学.
- 安田佑喜, 鈴木晶夫. 2変量アルキメデスコピュラに基づく判別分析. 2016年度統計関連学会連合大会. 2016年9月. 金沢大学.
- Yuki Yasuda, Akio Suzukawa. Bivariate discriminant analysis based on Gumbel copula. Hokkaido University and Korea University 4th Workshop in Statistics. 2016年6月. Korea University.
- Akio Suzukawa. Semiparametric models for lifetime distribution. Hokkaido University and Korea University 4th Workshop in Statistics. 2016年6月. Korea University.
- 鈴木晶夫. Multivariate survival analysis based on shared frailty models. 2016年日本数学会年会(招待講演). 2016年3月. 筑波大学.
- Yuki Yasuda, Akio Suzukawa. Discriminant analysis based on vine copulas. The 9th Conference of the Asian Regional Section of the IASC. 2015年12月. National University of Singapore.
- Akio Suzukawa. Generalized Archimedean copulas. The 9th Conference of the Asian Regional Section of the IASC. 2015年12月. National University of Singapore.
- 安田佑喜, 鈴木晶夫. ヴァインコピュラ

に基づく判別分析 .2015年度統計関連連
合大会 . 2015年9月 . 岡山大学 .

- 鈴川晶夫 . 生存時間分布に対する共有フ
レイルティモデルの一般化 . 2015年度統
計関連連合大会 . 2015年9月 . 岡山大学
- Shogo Omote, Akio Suzukawa. On the
nonlinear projection trick in the
kernel methods. Korea University and
Hokkaido University 3rd Workshop in
Statistics. 2015年2月 . 北海道大学 .
- Ryo Yawata, Akio Suzukawa. Reversible
Jump Markov Chain Monte Carlo Method.
Korea University and Hokkaido
University 3rd Workshop in Statistics.
2015年2月 . 北海道大学 .
- Akio Suzukawa. A generalization of the
shared frailty models for
multivariate lifetime distributions.
Korea University and Hokkaido
University 3rd Workshop in Statistics.
2015年2月 . 北海道大学 .
- Zhaojun Jian, Akio Suzukawa . Data
Depth and Support Vector Machine .
Korea University and Hokkaido
University 3rd Workshop in Statistics .
2015年2月 . 北海道大学 .
- 簡肇均, 鈴川晶夫 . Depth-SVM: Support
Vector Machine Based on Data Depth .
日本計算機統計学会第28回シンポジウ
ム . 2014年11月15日 . 沖縄科学技術大学
院大学 .
- 八幡涼, 鈴川晶夫 . 多変量打切りデータ
に基づく混合正規分布のベイズ推定 . 日
本計算機統計学会第28回シンポジウム .
2014年11月15日 . 沖縄科学技術大学院大
学 .
- 鈴川晶夫 . 短期的流行効果をもつ一般化
比例オッズ比モデル . 統計関連学会連合
大会2014 . 2014年09月15日 . 東京大学 .
- 鈴川晶夫 . 模倣効果をもつ生存時間分布

モデル . 科研費シンポジウム「時間事象
データ解析に関する理論と方法論およ
びその応用」 . 2013年12月21日 . 弘前大
学(青森県, 弘前市) .

- 鈴川晶夫 . 生存時間分布に対する流行
性モデルについて . 2013年09月10日 .
統計関連学会連合大会 . 大阪大学(大阪
府吹田市) .
- Akio Suzukawa . A survival model with
epidemic effects . HOKKAIDO
UNIVERSITY-KOREA UNIVERSITY SECOND
JOINTWORKSHOP IN STATISTICS . 2013年6
月 . Korea University .

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

6 . 研究組織

(1)研究代表者

鈴川 晶夫 (SUZUKAWA, Akio)
北海道大学・公共政策大学院連携研究部・教
授

研究者番号 : 00277287