

令和元年6月19日現在

機関番号：32660

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16021

研究課題名(和文) 層別化された分割表に対する有限混合モデルを用いた検定の構築と発展

研究課題名(英文) Construction and application of tests for stratified contingency tables using finite mixture models

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) 2,300,000円

研究成果の概要(和文)：本研究では、層別サンプリングではなく、母集団から通常のランダムサンプリングを行うことで得られる 2×2 の分割表に対し、どのようにして適切な検定法を構築するか、という問題を扱った。主要な成果としては、有限混合モデルを用いることで、層別化された 2×2 の分割表に対する新たな検定法の提案を行った。またそこで得られた結果を応用することで、生存時間解析、時系列データの解析への拡張、そして木構造モデル構築のための新たな手法の提案も行った。

研究成果の学術的意義や社会的意義

本研究で提案した手法は、ランダムサンプリングを前提として 2×2 の分割表を用いる解析法の提案であり、実際にそのようにデータが集められることは、生物統計を始めとする様々な分野で多く存在するため、有意義であると考えられる。サンプリング状況により、本研究で提案した手法の方が従来法よりも、より少ない標本で、より正しい結果を返す場合が多くあり、提案法を用いることで、例えば新薬等の開発において、より少ないコストでスピーディーな開発につながる事が期待される。

研究成果の概要(英文)：In this research, we dealt with the problem how to construct an appropriate test for 2×2 contingency tables obtained by random sampling. As the main result, we proposed a new test method based on the finite mixture model. We also proposed survival and time series analyses based on this main result.

研究分野：統計学

キーワード：分割表 有限混合モデル 生存時間解析 木構造モデル

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

1. 研究開始当初の背景

本研究では 2×2 の分割表を用いた検定を扱った。これは特に生物統計の分野で最も一般的かつ基本的な検定の一つである。関心のある二値の応答（陽性・陰性）を持つ独立な二群（群1・群2）に対し得られる標本は、 2×2 の分割表の形で与えられる。よく用いられる一つの例として、新薬群と対象薬群の患者に対し、得られる応答が薬の効果の有無である場合が挙げられる。この例において、 2×2 の分割表の4つのセルはそれぞれ、（新薬群かつ効果あり）、（対象薬群かつ効果あり）、（新薬群かつ効果なし）、（対象薬群かつ効果なし）に対応し、各患者はその中の1つのセルに含まれる事になる。ここで興味の対象は「この二群間で効果（応答の真の発現率）に差は有るのか？」であり、一般的には、非制限型尤度によるコクラン検定、もしくは制限型尤度によるマンテル・ヘンツェル検定が用いられてきた。

一方母集団内に、応答に関わるカテゴリカルな共変量（もしくは分類された連続な共変量）が存在する場合が存在する。先の新薬の例を挙げると、例えば男性と女性で薬の効果が異なる場合、もしくは10代の患者、20代の患者、... と分類していった場合に、各年代で薬の効果が異なる場合等が考えられる。二群間でこのような共変量の標本分布が異なる場合に、それを無視して一つの 2×2 の分割表を用いて検定を行うと、バイアスを含んだ結果が得られることになる。そこでこのような場合は、共変量の値によって層別化した 2×2 の分割表を用いる検定が一般的に行われる。これは仮定として、標本を集める前に応答に関わる共変量が予測されているとする。そのような共変量の値ごとに標本を集めることで（層別サンプリング）、複数の層別化された 2×2 の分割表を構築することができる。ここで興味の対象は「一か所以上の層において、二群間の効果に差は有るのか？」となり、層調整済みコクラン検定、もしくは層調整済みマンテル・ヘンツェル検定が広く用いられてきた。

2. 研究の目的

層別サンプリングによる検定手法は理にかなっているように思われるが、以下の疑問が挙げられる：層別サンプリングではなく母集団から通常のランダムサンプリングを行い、そしてそれに対する適切な検定法が存在する場合、層別サンプリングによる検定と比べどちらが優れているのか？ 言い換えると、層別サンプリングを行うということは、母集団内における共変量の各層の分布（比率）を無視していることになるので、層別ではない通常のランダムサンプリングを行い、その分布を考慮した検定を行う方が、より正しい結果（帰無仮説下では名目水準に近い、もしくは対立仮説下では検出力が高い）を与えてくれる可能性が考えられる。そこでこの疑問に着目し、通常のランダムサンプリングに対する有限混合モデルを用いた層別化された 2×2 の分割表に対する検定法の提案を行うこと、またその結果を利用した発展的手法を構築することを目的とした。

より具体的な目的としては「母集団内における真の層比率が既知の場合における有限混合モデルを用いた検定統計量の構築」、「層別化された 2×2 の分割表を用いる生存データへの検定手法への拡張」、「層別化された 2×2 の分割表を用いる検定を利用した生存木の構築」を行うことを目標とした。

3. 研究の方法

上述した3点の目的それぞれについて、提案手法の数式構築、その性質の考察、従来法との比較、実データへの適用と考察を順に行っていった。これら項目を達成するためのアプローチとしては主に、文献調査、理論的（数式による）検討、シミュレーションによる検討、そして学会発表を通して得られる第三者からの意見・疑問の検討を行った。また検定統計量の数式的導出が困難であったとき、そして生存データへの検討において層比率の推定が困難であったときは、対処としてそれぞれ、ブートストラップ法、時間区間を用いるアプローチを用いた。

4. 研究成果

主要な研究成果としては3つ挙げられる。一つ目は、「母集団内における真の層比率が既知の場合における有限混合モデルを用いた検定統計量の構築」である。先行研究では真の層比率が未知の場合における理論構築を行っていたが、実際のデータ解析において、母集団内の興味がある共変量の分布は既知の場合も多く存在し、真の層比率が既知の場合における通常のランダムサンプリングによる検定法の提案を行った。

二つ目は「層別化された 2×2 の分割表を用いる生存データへの検定手法への拡張」である。医療データの解析で広く用いられている検定にログランク検定があるが、その一つの解釈としては、比較する二群間で注目するイベントが発生した時点における 2×2 の分割表を考え、それを全てのイベント発生時点について考慮する事で二群間の生存関数に差が有るのかを検定している。これは全てのイベント発生時点を層として考えていることに等しく、生存データは層別サンプリングで集められるわけではないので（イベント発生時点は共変量のように予め予測することは出来ないため）、検定統計量の分散を低く見積もっている可能性が示唆されていた。そこで生存データに対し、有限混合モデルを用いることで分散を正しく見積もった、ログランク検定に代わる新たな検定法の構築を行った。

三つ目は上述の結果を用いて「層別化された 2×2 の分割表を用いる検定を利用した生存木の構築」を行った。生存木とは木構造のモデルによってデータ（共変量空間）を分割し、生存

関数が有意に異なる群へと分類する手法であり，そのモデル構築における最も重要な基準は，データの分割における評価関数の与え方である．生存木構築において一般的に用いられている基準の一つにログランク検定を用いる手法があるが，その基準を提案手法に変えることで，新たな木構造構築の手法を提案した．

5．主な発表論文等

〔雑誌論文〕(計3件)

Asanao Shimokawa, Etsuo Miyaoka, Application of the bootstrap method for change points analysis in generalized linear models, Japanese Journal of Statistics and Data Science, 査読有, 1, 2018, 413-433.

Asanao Shimokawa, Etsuo Miyaoka, On stratified adjusted tests by binomial trials, International Journal of Biostatistics, 査読有, 13(1), 2017.

Takuma Kurosawa, Asanao Shimokawa, Etsuo Miyaoka, A Note on Transition Models for Binary 2 x 2 Cross-Over Data, 査読有, 3(5), 2017.

〔学会発表〕(計21件)

Asanao Shimokawa, Etsuo Miyaoka, Construction of a survival tree based on prediction accuracy, 29th IBC, 2018.

Asanao Shimokawa, Etsuo Miyaoka, On survival tree under the dependency between failure and censoring, 2018 JSM, 2018.

Asanao Shimokawa, Etsuo Miyaoka, Statistical construction of a decision tree, ISCB and ASC 2018, 2018.

Asanao Shimokawa, Etsuo Miyaoka, A comparative study on the measures for the prediction accuracy of a survival model, 11th CMStat, 2018.

下川 朝有, 宮岡 悦良, 依存打ち切りを考慮した生存木構築について, 2018年度統計関連学会連合大会, 2018.

下川 朝有, 宮岡 悦良, C-index を利用した生存木の構築について, 日本計算機統計学会第32回シンポジウム, 2018.

Asanao Shimokawa, Etsuo Miyaoka, On change points analysis based on resampling methods, 61st ISI, 2017.

Asanao Shimokawa, Etsuo Miyaoka, Application of finite mixture models for survival analysis with competing risks, JSM 2017, 2017.

Asanao Shimokawa, Masataka Kuroda, Etsuo Miyaoka, Symbolic data analysis using classification tree, 10th CMStat, 2017.

Takuma Kurosawa, Asanao Shimokawa, Etsuo Miyaoka, Application of transition models for higher-order crossover designs with binary responses, 10th CMStat, 2017.

下川 朝有, 宮岡 悦良, 競合リスクのある生存解析について, 日本計算機統計学会第31回大会, 2017.

下川 朝有, 宮岡 悦良, 有限混合モデルを用いた競合リスクの解析について, 2017年度統計関連学会連合大会, 2017.

下川 朝有, 黒田 正孝, 宮岡 悦良, シンボリックデータに基づく分類木の構築について, 日本計算機統計学会第31回シンポジウム, 2017.

Asanao Shimokawa, Takuma Kurosawa, Etsuo Miyaoka, On the stratified test of cross tables based on finite mixture models, 28th IBC, 2016.

Asanao Shimokawa, Takuma Kurosawa, Etsuo Miyaoka, Application of ensemble methods for detection of changepoints in generalized linear models, COMPSTAT 2016, 2016.

Takuma Kurosawa, Asanao Shimokawa, Etsuo Miyaoka, Dynamic binary choice panel data model with fixed effect based on transition model, COMPSTAT 2016, 2016.

下川 朝有, 黒澤 匠雅, 宮岡 悦良, 層別化された分割表を用いた近似検定について, 日本計算機統計学会第30回大会, 2016.

黒澤 匠雅, 下川 朝有, 宮岡 悦良, 日パネルデータに対する推移モデルについて, 日本計算機統計学会第30回大会, 2016.

下川 朝有, 黒澤 匠雅, 宮岡 悦良, 機械学習を用いた一般化線形モデルにおけるチェンジポイントの検出について, 2016年度統計関連学会連合大会, 2016.

黒澤 匠雅, 下川 朝有, 宮岡 悦良, 二値クロスオーバーデータに対する推移モデルについて, 2016年度統計関連学会連合大会, 2016.

21 下川 朝有, 黒澤 匠雅, 宮岡 悦良, ブートストラップ法を用いたチェンジポイントの検出について, 日本計算機統計学会第30回シンポジウム, 2016.

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

6 . 研究組織

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：

ローマ字氏名：

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。