

平成30年6月19日現在

機関番号：62603

研究種目：研究活動スタート支援

研究期間：2016～2017

課題番号：16H07406

研究課題名（和文）混合効果モデルにおける変量効果の特定化と小地域推定への応用

研究課題名（英文）Specification of random effects distributions in mixed-effects models with applications to small area estimation

研究代表者

菅澤 翔之助（Sugasawa, Shonosuke）

統計数理研究所・リスク解析戦略研究センター・特任研究員

研究者番号：50782380

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）：混合効果モデルを用いた小地域推定法において、新たな統計モデルの開発とそれに伴う理論解析を行った。特に、地域差を表現する変量効果の分布の仮定が必ずしも現実のデータと整合的でないことに注目し、それらを解消する有効な手法の開発を目指した。本研究では、データ変換を用いた手法の開発、不均一分散を扱うモデルの提案、変量効果の選択を可能にする有効な手法の開発、リンク関数を用いたモデルの提案、を行った。

研究成果の概要（英文）：We developed new statistical models and theories for small area estimation based on mixed-effects models. Especially, we focused on the fact that the distributional assumption of random effects is not necessarily consistent with real data, and aimed to develop effective methods to solve the problem. In this research project, we developed small area models with data transformation, heteroscedastic variances, selection of random effects and link functions.

研究分野：経済統計

キーワード：小地域推定 混合効果モデル 階層ベイズ 標本調査

1. 研究開始当初の背景

地域別の母数(平均など)に関する推定値は、地域別の経済状況の把握や地域振興に関する政策立案の場面で、近年需要が高まっている。一方で、標本調査において十分な標本数を確保することは一般的に困難であり、十分なサンプルが得られない状況で単純に集計した推定値は分散が大きく(精度が低く)なってしまうことが知られている。一般に、このような問題は小地域推定問題と呼ばれている。

この問題に対して、地域ごとの標本のみから推定値を算出するのではなく、有益な外部情報や、統計モデルを利用して推定精度を高める方法論に関する研究が現在盛んに行われている。そのような手法は、小地域推定法として知られており、代表的なモデルあるFay-Herriot (FH) モデルをはじめ、混合効果モデルが幅広く利用されている。混合効果モデルでは、地域差を表現する変量効果に対して何らかの分布(正規分布など)を仮定するのが一般的である。一方で、実際のデータに対してこのような分布の仮定が満たされないケースも多く、その場合には、混合効果モデルによって算出される小地域母数の推定値の精度が著しく低下してしまう可能性がある。小地域推定法を幅広い問題で運用するためには、このような問題点を解決することが必要不可欠であり、有効な手法の開発が必要とされている。

2. 研究の目的

本研究では、小地域推定法における変量効果の分布の仮定と現実のデータの乖離に注目し、それを埋めるための有効な統計手法の開発を行う。

3. 研究の方法

本研究は、小地域推定への応用を想定した混合効果モデルに関連する統計手法の開発

が研究対象であり、新たな方法論の開発と数理的な論証、新たに開発した方法論の実装と発信(関連する統計家が開発した手法を利用できるよう、開発した手法を実装したプログラムを作成し、公開する)、シミュレーション実験による既存手法との比較、実際の標本調査への適用による実証的な評価などの過程を経て、課題解決のための方法論の開発・評価を行う。

4. 研究成果

(1) データ変換を用いた小地域推定モデルの開発

標本調査で頻繁に現れる正值データに対して、データ変換を用いたアプローチについての研究を行なった。従来の方法論では、データの特性に関わらず対数変換が広く用いられており、対数変換が妥当でないケースでは算出される推定値の信用性が損なわれてしまう危険性があった。本研究では、対数変換を含むパラメトリックな変換を導入し、既存手法の拡張を行った。地域母数の推定量の導出およびブートストラップ法による漸近的リスク評価法を導出し、理論的な性質を示すと同時に、数値実験・実データへの適用を通して提案手法の有用性を示した。

(2) 不均一分散を考慮した小地域モデルの提案

集計データに対するFHモデルでは、標本から計算した分散を既知の値としてモデルに取り込んで推定値を導出する。これは標本分散の不確実性を無視することになり、推定量のリスク評価などの観点で好ましくないことが指摘されてきた。本研究では、標本分散と地域ごとに異なる(不均一な)真の分散をモデリングする階層構造を新たに導入する形で、FHモデルを拡張することを行った。マルコフ連鎖モンテカルロ法を利用したベイズ推測の方法を開発し、標本分散の不確実性

を取り込んだ小地域推定を可能にした。

一方で、個別データが得られる場合において、地域ごとに分散が異なっている事例も報告されている。このような個別データにおける不均一分散の問題を取り扱うために、分散関数を導入した混合効果モデルを新たに提案した。分布に正規性を仮定せず、モーメント法による推定方法を開発し、その漸近的性質、地域母数推定量の漸近的リスク評価指標を導出した。

(3) 変量効果の選択問題に対するアプローチ

FH モデルをはじめとする小地域モデルにおける変量効果の選択問題についての研究を行った。従来のモデルでは、すべての地域に変量効果を導入している一方で、変量効果を除くことで推定精度が向上するケースがあることが指摘されてきた。本研究では、このような変量効果の選択問題に対して、確率的に変量効果が存在する構造を導入し、既存モデルの拡張を提案した。カウントなど非正規性のデータも考慮した広いクラスの集計データを想定し、自然指数型分布族をもとにした新しい階層モデル(小地域モデル)を開発した。また、個別データのケースにおいても、変量効果の不確実性を導入したモデルを提案し、マルコフ連鎖モンテカルロ法を用いたベイズ流の推定方法を開発した。

(4) リンク関数を用いた小地域モデルの提案

変量効果の正規性の仮定を解消する研究を行った。本研究では、リンク関数を用いて変量効果をモデル化することにより、正規性の仮定が妥当でないケースにおいても利用可能なモデルを開発した。推定値の算出および推定値のリスク評価のために、モンテカルロ法を駆使したアルゴリズムを開発した。

(5) 関連する成果

上記の研究成果と関連して、データ変換を利用した有限母集団の母数を予測する方法論の開発、ダイバージェンスの導入によるFHモデルのロバスト化、地域などによってグループ分けされたデータに対する一般的なモデリング手法の開発なども行った。また、論文として発表はされていないが、本研究プロジェクトに関連するテーマとして小地域推定法の発展に貢献することが期待される。

5. 主な発表論文等

[雑誌論文](計 6 件)

(1) Sugasawa, S., Kubokawa, T. and Rao, J. N. K. (2018). Small area estimation via unmatched sampling and linking models. TEST 27, 407-427. (査読有り)

(2) Sugasawa, S. and Kubokawa, T. (2017). Bayesian estimators in uncertain nested error regression models. Journal of Multivariate Analysis 153, 52-63. (査読有り)

(3) Sugasawa, S., Tamae, H. and Kubokawa, T. (2017). Bayesian estimators for small area models shrinking both means and variances. Scandinavian Journal of Statistics 44, 150-167. (査読有り)

(4) Sugasawa, S. and Kubokawa, T. (2017). Transforming response values in small area prediction. Computational Statistics & Data Analysis 114, 47-60. (査読有り)

(5) Sugasawa, S. and Kubokawa, T. (2017). Heteroscedastic nested error regression models with variance functions. Statistica Sinica, 27, 1101-1123. (査読有り)

(6) Sugasawa, S., Kubokawa, T. and Ogasawara, K. (2017). Empirical uncertain Bayes methods in area-level models. Scandinavian Journal of Statistics 44,

684-706. (査読有り)

[学会発表] (計 7 件)

(1) Sugasawa, S. (2018 年 3 月). Latent mixture modeling for clustered data. ISM Symposium on Environmental Statistics, The Institute of Statistical Mathematics.

(2) Sugasawa, S. (2017 年 7 月). Robust empirical Bayes small area estimation with density power divergence. Small Area Estimation 2017, University Pierre et Marie Curie.

(3) Sugasawa, S. (2017 年 6 月). Robust empirical Bayes small area estimation with density power divergence. 26th South Taiwan Statistical Conference, National Taipei University.

(4) 菅澤翔之助, 久保川達也, 小笠原浩太 (2017 年 5 月). Spike-Slab 型事前分布を用いた地域データ解析. 日本計算機統計学会第 31 回大会, 東京理科大学森戸記念館.

(5) 菅澤翔之助 (2017 年 3 月). 局所尤度を用いた disease mapping. 第 7 回生物統計ネットワークシンポジウム, アクロス福岡.

(6) Sugasawa, S. (2017 年 2 月). Flexibly transformed empirical best prediction in finite population. ISI-ISM-ISSAS Joint Conference 2017, Indian Statistical Institute.

(7) 菅澤翔之助, 小林弦矢, 川久保友超 (2016 年 9 月). 潜在混合密度回帰モデルによるグループデータ解析. 2016 年度統計関連学会連合大会, 金沢大学角間キャンパス.

6 . 研究組織

(1) 研究代表者

菅澤 翔之助 (Shonosuke Sugasawa)

統計数理研究所・リスク解析戦略研究センター・特任研究員

研究者番号 : 50782380