

令和 6 年 6 月 7 日現在

機関番号：32682

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K01421

研究課題名（和文）所得・貧困に関する時系列および時空間データのベイズ解析

研究課題名（英文）Bayesian time series and spatio-temporal Bayesian analysis of income and poverty

研究代表者

小林 弦矢（Genya, Kobayashi）

明治大学・商学部・専任教授

研究者番号：00725103

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究では、単年度・国全体のレベルではなく、複数時点および複数地域における所得のグループデータの分析のための時系列および時空間ベイズ解析の方法と、グループデータ等に関連するベイズモデリング方法に関する研究を行った。具体的には、グループデータに対する時空間混合モデルや線形混合モデル、多変量カウントグループデータに対する因子モデル、カウントデータに対する混合予測モデルなどを取り扱った。本研究で得られた成果は国際学術誌への掲載・投稿および国際学会において報告された。

研究成果の学術的意義や社会的意義

日本において所得データはグループデータの形式で公表されることが多いのに対し、時空間モデルなどに基づく高度な分析方法はグループデータに対しては確立されていないため、本研究による新しい種々のベイズモデリングの方法は所得データなどのグループデータに対して極めて有用な分析手法となる。特に、提案手法は空間補間や将来予測にも使用することができるため、任意の時点・地域において任意の所得・貧困指標を得ることができ、政策立案等のための情報提供において大きく貢献することが期待できる。またこれらから派生した他の提案手法も諸科学分野において頻繁に利用されることが期待される。

研究成果の概要（英文）：In this project, we have studied the spatio-temporal Bayesian modelling of grouped income data for multiple time points and municipalities and Bayesian modelling for other types of grouped data analyses. More specifically, we have considered spatio-temporal mixture and linear mixed models for grouped data, a factor model for multivariate grouped count data, and a mixture Bayesian predictive model for count data. The results obtained from our studies are reported at international conferences. The papers are submitted to or published on international journals.

研究分野：ベイズ統計学

キーワード：ベイズ統計学 時空間モデル グループデータ

## 1. 研究開始当初の背景

日本において所得の格差や貧困に関する問題は特に大きな注目を集めてきており、学術的にも政策的にも重要な課題として近年取り上げられてきている。格差・貧困に関する課題を効果的に解決するためには、対象となる人口を一様に扱うのではなく、人口のどのような層や地域においてどの程度格差や貧困の度合いが深刻であるかについて把握した上で、それぞれの対象に対して必要な対策を講じる必要がある。格差・貧困の程度を把握するためにはすべての家計に対して調査を行うことが現実的に不可能であるため、標本調査によって得られたデータに基づいて各種指標を推計する必要がある。所得の指標を表す指標として、各所得階級に属する家計を数える度数分布(グループデータ)、平均所得、所得の中央値などが従来より用いられてきた。公的データにおいて個票レベルで所得データを公表すると個人が特定される可能性があるということもあるため、グループデータの形に集計され公表される。

所得に関するデータから上記の指標を推計するには大まかに2つのアプローチがあると考えられる。一つ目は所得分布に基づいたアプローチで、所得データに対して何らかの所得分布を仮定しその所得分布のパラメータを推定することで、所得分布から誘導される指標に対しての推定値を得るということである。もう一つのアプローチは小地域推定に基づくアプローチである。小地域推定はデータに含まれる標本数が少ない複数の地域について関心のある指標を安定的に推定することを目指す統計学的手法である。小地域推定では線形混合モデルが広く用いられ、使用するデータが所得中央値など関心のある数量についての集計値(直接的な推定量)が個票データかによって、それぞれ地域レベルモデルとユニットレベルモデルに分かれる。

本研究開始時点において上の二つのアプローチには以下の問題点が挙げられる。所得分布によるアプローチは、単一時点の単一の国や地域の所得分布とそれに伴うジニ係数などの推定のみに従来焦点を当ててきた。単一の国を推定対象としては各地域における所得の状況を把握することができない。また粒度の小さい地域ごとに所得分布を別個に推定するにあっても、標本数が小さいという理由から安定的に分布を推定することが不可能である。所得データがグループデータの場合にはデータに含まれる情報が限られるためこの問題は特に深刻である。所得に関するデータの蓄積はなされている一方で、所得分布の文脈においては複数地域・複数時点における分布推定に関する研究の蓄積がなされていない。小地域推定のアプローチにおいては、グループデータに対して適用可能な統計モデルに関する研究が進んでおらず、連続な結果変数に対する線形混合モデルが利用されている。さらに両アプローチにおいては基本的にある一時点におけるデータしか取り扱うことができず、パネルデータのような時系列的・空間的な構造も持つデータの分析には対応していない。

## 2. 研究の目的

前節の学術的と政策的要請に基づき、本研究は中央値などの所得に関する指標、ロレンツ曲線やジニ係数などの格差に関する指標を安定的に推定するために、以下の項目についてベイズ統計学に基づく分析手法の開発を行っていく。

### 1. 時系列データに基づいたロレンツ曲線の安定的なベイズ推定

本研究ではまず時系列データからロレンツ曲線を安定的に推定する方法を考案する。時系列構造により時点間でデータの情報を借り合うことで、ロレンツ曲線およびそれから誘導されるジニ係数のより安定的な推定と将来の予測を行うことが可能になる。

### 2. 全国市区町村の所得・貧困指標に関する時空間ベイズ推計

ここでは地域単位でのパネルデータが利用可能であるという前提のもと、グループデータから所得・貧困指標を安定的に推定する方法の開発を行い、前節に挙げた所得分布と小地域推定の両方の側面からアプローチを行っていく。

所得分布の側面では、時間の経過とともに空間上で所得分布自体や所得などに関する指標がどのように変化・推移していくのかを明らかにするために、所得分布のパラメータが時空間上で変化していくような潜在的な構造を階層的に取り入れた状態空間モデルを構築する。

小地域推定の側面においては、まず単一時点における線形混合モデルをグループデータの枠組みで利用できるようにモデルの定式化を行う。具体的には対数正規分布の超母集団のもとでの有限母集団を考えることで、グループデータの各所得階級の生起確率の定式化を行うことで尤度関数を構築する。

### 3. 上記研究から派生する種々のデータ分析手法の開発

日本では所得データはグループデータで公表されることが多いが、グループデータは特に社会

科学領域において頻繁に出現する。これは所得データのように個人情報のある程度マスキングするのに加えて、回答者の負荷を和らげるなどといった理由により厳密な値ではなくグループデータとしてデータが収集されることがある。ベイズ統計学の枠組みでグループデータに対する新たな分析手法の開発も行う。

### 3. 研究の方法

それぞれにおいて、例えば時点間や地域間の情報を効果的に借り合うことのために潜在変数を導入し、状態空間モデルや階層モデルに基づいたモデルの構築を行う。各階層において時系列構造や空間相関などの構造を導入する。提案するモデルは解析的に解けないことが考えられ、シミュレーションに基づく数値計算手法の開発も併せて行う。具体的にはマルコフ連鎖モンテカルロ法やモンテカルロ EM アルゴリズムなどを採用し、アルゴリズムの導出を行う。新しく開発・提案する統計モデルの有効性を数値実験および実データの応用にもとづいて学術的および応用的な観点の両方において評価する。

### 4. 研究成果

2021 年度は、住宅・土地統計調査における所得のグループデータから日本全国の市区町村の所得分布の推定・予測をするために、柔軟な時空間混合モデルを考案し、モデルの推定方法の開発と実データへの適用を行った。ここで扱うグループデータは、各世帯の所得は観測されておらず、各市区町村において予め設定された各所得階級に属する世帯数のみが複数年度に渡って観測されているものとする。各時点・市区町村における所得分布を柔軟にモデリングするために、潜在的な家計所得が対数正規分布の混合モデルによってモデリングを行い、時空間効果をその混合割合に導入した。これにより、柔軟性を保ちつつ、パラメータや潜在変数の数を比較的少なくすることが可能となった。さらに、混合モデルにおいて必要なコンポネント数を決定する方法も既存の混合モデルにおける方法に従って検討した。提案するモデルに対し、Polya-Gamma データ拡大に基づくマルコフ連鎖モンテカルロ法を開発した。日本の実データへ適用し、既存モデルでは不可能であった、各市区町村の所得分布の空間補間および将来予測を行えるということのデモンストレーションを行った。ここまでの研究結果は、国際学会 XV World Conference of the Spatial Econometrics Associationなどで報告を行った。また家計調査から得られる複数時点での累積所得割合のデータに対し、パラメトリックなロレンツ曲線を仮定し、そのパラメータが状態空間モデルによって時変するという構造を考慮することで、複数時点間でデータの情報を仮り合い、安定的にロレンツ曲線の推定ができるようにした。また尤度関数をディリクレ分布ではなく、一般化モーメント法に基づく損失関数を基に一般化ベイズの枠組みでモデルの推定を試みた。

2022 年度は、引き続きグループデータに基づく市区町村レベルの所得・貧困指標に関する推測の問題に関する研究に取りかかった。2021 年度では住宅土地統計調査の集計結果に対する柔軟な時空間混合モデルを開発したのに対し、2022 年度は同調査の個票データを用いることを想定して小地域推定の考え方に基づいた統計的分析手法の開発を行った。具体的には、グループデータを構成する観測されない潜在的な世帯収入に対してベイズ線形混合モデルを考え、尤度関数の構築において住宅土地統計調査における各世帯が票差に選ばれる確率をもとにサーベイウェイトを明示的に取り入れた。数値実験の結果から、標本が包含確率に基づいて構成される場合、提案手法のほうがサーベイウェイトを考慮しない手法よりもパフォーマンスが優れていることがわかった。またここまでのグループデータおよび時空間モデルに関する研究の副産物として、多変量のカウントグループデータに対する因子モデルを考案し、ベイズ統計学の枠組みでマルコフ連鎖モンテカルロ法に基づく推定方法を開発した。さらに個別の予測モデルを統合するベイズ予測統合において、統合関数を混合分布によって構成することで、複数の時系列に対してパラメータ数を節約しつつも個別の予測モデルの予測精度を向上させるモデルを考案した。これらの提案手法に関する研究結果は国際学会および国内学会・国内研究集会などで報告された。

前年度で行った本研究課題から派生したグループデータおよび時空間モデルに関する研究に連続し、2023 年度では以下の研究を行った。前年度の因子モデルについて、パラメータ拡張の導入方法を簡便化し、因子コンポネントの識別をパラメータ拡張ではなくマルコフ連鎖モンテカルロ法の事後処理にまわすことで、識別性の改善を実現することができた。さらにこれらの研究から派生し、本年度では、新たに単変量カウントデータの条件付き分布の推定とそれに基づく分位点回帰モデルの一般化ベイズ推定の方法を提案した。所得分布などのグループデータに対する分布推定の方法を考え方を参考に、カウントデータの背後に連続な潜在変数を導入し、Pitman-Yor 過程に基づくカーネル混合によって被説明変数と説明変数の同時分布を推定し、それから被説明変数の条件付き分布を推定することを考えた。この条件付き分布から条件付き分位点を計算し、あるその分位点と回帰関数との間の損失を最小化するように回帰関数のパラメータを求めることでカウントデータの分位点回帰を行うことを提案した。数値実験及び実データへの適用から、既存のジッタリングや連続化離散分布に基づく方法よりも提案手法がうまく機能することが確認された。

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 Sugasawa Shonosuke, Kobayashi Genya	4. 巻 174
2. 論文標題 Robust fitting of mixture models using weighted complete estimating equations	5. 発行年 2022年
3. 雑誌名 Computational Statistics & Data Analysis	6. 最初と最後の頁 107526 ~ 107526
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2022.107526	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 Kawakubo Yuki, Kobayashi Genya	4. 巻 184
2. 論文標題 Small area estimation of general finite-population parameters based on grouped data	5. 発行年 2023年
3. 雑誌名 Computational Statistics & Data Analysis	6. 最初と最後の頁 107741 ~ 107741
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2023.107741	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5件（うち招待講演 5件 / うち国際学会 1件）

1. 発表者名 Genya Kobayashi
2. 発表標題 Spatio-temporal smoothing, interpolation and prediction of income distributions based on grouped data
3. 学会等名 5th International Conference on Econometrics and Statistics (招待講演)
4. 発表年 2022年

1. 発表者名 Genya Kobayashi
2. 発表標題 Bayesian factor zero-inflated Poisson model for multiple collapsed count data
3. 学会等名 2022年統計関連学会連合大会 (招待講演)
4. 発表年 2022年

1. 発表者名 Genya Kobayashi
2. 発表標題 Clustering and predicting time series count data via mixture of Bayesian predictive synthesis
3. 学会等名 科研費シンポジウム「ベイズ統計学の最近の展開」(招待講演)
4. 発表年 2023年

1. 発表者名 Genya Kobayashi
2. 発表標題 Spatio-temporal smoothing, interpolation and prediction of income distributions based on grouped data
3. 学会等名 XV World Conference of the Spatial Econometrics Association (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Genya Kobayashi
2. 発表標題 Spatio-temporal smoothing, interpolation and prediction of income distributions based on grouped data
3. 学会等名 CSIS DAYS 2021 (招待講演)
4. 発表年 2021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			
韓国	Korea University			