

平成 30 年 5 月 2 日現在

機関番号：34504

研究種目：研究活動スタート支援

研究期間：2016～2017

課題番号：16H06853

研究課題名(和文) 発展途上国における死因分布のベイズ推定

研究課題名(英文) Bayesian estimation of distributions of causes of death in developing countries

研究代表者

國濱 剛 (KUNIHAMA, Tsuyoshi)

関西学院大学・経済学部・講師

研究者番号：40779716

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、戸籍情報が十分ではない発展途上国に対して、口頭剖検データを使って死因分布を推定するための新たなベイズ統計手法の開発を行った。考案した統計手法に対して、Population Health Metrics Research Consortiumデータを用いて、地域全体における死因分布の推定精度の評価を行った。複数のシナリオの下、従来の統計モデルと比較を行った結果、提案したベイズ統計モデルの方がより正確に死因分布を推定できることがわかった。

研究成果の概要(英文)：This research project develops a new Bayesian method for estimation of population distributions of deaths by cause using verbal autopsy data in developing countries without full-coverage civil registration system. Using the Population Health Metrics Research Consortium data, we assess accuracy of the estimation of distributions of cause-of-death by the proposed method. Under various scenarios, we compare the proposed method to an existing one and observe the outperformance of the new Bayesian approach.

研究分野：社会科学

キーワード：ベイズ統計学 マルコフ連鎖モンテカルロ法 多変量データ 条件付き確率

1. 研究開始当初の背景

死因などの人口学情報は、各国の政策作成、実行、評価において非常に重要である。ところが世界全体では3分の1以下の死に対してのみ原因が特定されており、特に貧困に苦しむ発展途上国では保健システムが不完全であるため死因情報が非常に乏しい。このような地域では戸籍制度が十分に発達しておらず、国全体を網羅する人口動態統計もないため、国民の健康状態を把握した上で、最も適切な保健戦略や支援策を適用することが極めて困難である。そこで、大規模調査を行うことで、各地域における死因情報を多く集める必要があるが、その際に問題となるのは、死因の特定方法である。理想としては、医師などの医学知識のある専門家をすべての調査地域に派遣し、死体を検案することであるが、広範囲な調査地域をカバーするには莫大な費用と時間が必要となり現実的ではない。そこで、現実的な死因特定手法として、家族、保健ケア提供者、コミュニティの構成員に死亡状況の説明を求める口頭剖検に基づく事後的死亡調査が広く行われている。

2. 研究の目的

本研究の目的は、発展途上国における死因の分布を推定するための新たな統計手法を開発することである。国全体を網羅する戸籍情報や人口動態統計が十分ではない発展途上国では、多くの死が病院外で起こり、記録として残されないため、国民の死因分布の情報にすら大きな不確実性が残る。解決策として、近親者等からの事情聴取に基づき死因を推定する口頭剖検を用いて死因を統計的に予測することが広く行われている。ところが既存の統計手法はモデリングの簡略化のために、各死因に対して症状の条件付独立性を仮定している一方で、実際のデータは明らかにこの条件を満たしていない。そこで、条件付独立性を仮定しない柔軟な統計手法を新たに開発し、死因の予測精度を向上させる必要がある。

口頭剖検の多くは、様々な病気に関連する症状の有無などの2値変数で構成されており、質問項目は100以上に及ぶ。さらに、遺族がすべての質問に回答するとは限らないことに加え、予算の関係から調査によって質問数が異なるため、結果として多くの欠損値が存在することになる。また、症状と死因を結びつける情報は、死因が与えられた下での症状の有無の確率の形で与えられることが多いため、ベイズの定理に基づいて、この条件付確率から、症状が与えられた下での死因の確率を表現して予測を行う。一般に、高次元2値変数の統計モデリングは困難であるため、従来手法では死因が与えられた下での症状の条件付独立性を仮定しているものの、多くのデータはこの仮定を満たさない。既存の予測手法ではこのような点を考慮に入れておらず、拡張方法も明らかでない。そ

こで、精度の高い死因予測を可能にするために、口頭剖検の特徴を捉えた新たな統計手法の開発が必要であり、条件付独立性を仮定せず、症状間の従属性を考慮した柔軟な予測統計モデルを新たに考案することを目指した。

3. 研究の方法

死因の条件付での死亡状況・症状の新たな統計モデルとして、多変量正規分布に従う潜在変数に基づく多変量プロビットモデルを考案した。様々な分野で広く用いられているプロビットモデルでは、2値変数を正規分布に従う潜在変数によって表現することが可能であり、事後分布からの効率的なパラメータのサンプリング方法が提案されている。その拡張として、本研究における死亡状況・症状変数のモデリングのための多変量プロビットモデルへの応用を行った。具体的には、多変量正規分布に従う潜在変数を仮定し、各変数が正の値を取る区間で1を返す指示関数を用いて元の変数への変換を行う。さらに、マルコフ連鎖モンテカルロ法を用いることで、モデル内のパラメータと潜在変数の効率的な推定が可能となった。

注意すべき点として、高次元における共分散行列のモデリングが考えられる。通常のベイズ統計学では逆ウィッシュарт分布を事前分布として仮定するが、変数が高次元の場合にはサンプル数に対してパラメータ数が多く、推定結果が不安定になり、予測精度が悪くなることが考えられる。そこで、計算時間の短縮のためにも共分散行列の構造を低次元パラメータでスパースに表現することが必要となる。そこで考えたのが、因子モデリングの応用である。100以上の症状変数に対して、数個の因子を想定することで、共分散行列の有効なパラメータ数を低く抑えることが可能となった。本研究においては、死因によって死亡状況・症状の相関構造は異なると考えられるため、平均だけでなく、共分散も死因に依存するような多変量プロビットモデルの構築を行った。

4. 研究成果

本研究では、戸籍情報や人口動態統計が十分ではない発展途上国における死因分布を推定するための新たなベイズ統計手法の開発を行った。提案したアプローチでは、社会調査データにおいて頻繁に発生する欠損値に対して特別な補完処理を必要とすることなく、100以上もの症状データのモデル化を行い、口頭剖検による死因を統計的に予測することを可能とした。新たな枠組みでは、既存の統計手法において繰り返し仮定されてきた症状の条件付き独立性を取り除き、多変量プロビットモデリングを応用して症状間の複雑な相関関係を柔軟に表現している。さらに、推定すべきパラメータ数を削減するために、因子モデルを取り入れて計算時間の短縮化、推定結果の安定性を実現している。提

案したベイズ統計モデルに対しては、マルコフ連鎖モンテカルロ法を用いた効率的なパラメータ推定方法を開発した。また、多数ある症状変数の中で、死因と関わり合いが強そうなものがどれかを推定する方法の提案も行っている。

本研究で考案した統計手法に対して、Population Health Metrics Research Consortium による口頭剖検用死因データを用いて、地域全体における死因分布の推定精度の評価を行った。複数の現実的なシナリオの下、条件付き独立性を仮定する従来の統計モデルと比較を行った結果、提案したベイズ統計モデルの方がより正確に死因分布を推定できることがわかった。加えて、死因と質問項目との条件付き従属性の推定も行い、病歴などの一部の変数が死因との間に比較的強めの相関があることを示した。

また、死因が与えられた下での症状の条件付き独立性は、統計手法を用いた口頭剖検による死因分析の多くに仮定されているが、これが非常に制約的であることを、Population Health Metrics Research Consortium による口頭剖検用死因データを用いて明確に示した。提案したベイズ統計モデルでは因子モデルを用いてパラメータ数の削減を行っているが、分析前に事前決定する因子数に推定結果がどの程度依存するのかも Population Health Metrics Research Consortium データを用いて詳しく調べた。また、100 以上の症状変数に対して数個の因子を用いるだけで十分精度の高い死因分布推定が可能であることも明らかにした。

さらに、トレーニングデータにおける死因分布とテストデータにおける死因分布の形状のずれによって推定精度に違いが生じるのかどうか複数地域の死因データを用いて調べ、形状のずれが存在する場合は追加調査を行って対象地域のデータを集めることで分布推定を改善できることを示した。

これまでに得られた研究成果を論文としてまとめ、様々な国内・国際学会で発表を行った。研究論文はワーキングペーパーとして一般に公開し、応用統計学の国際雑誌に投稿を行った。また、他分野の専門家や実務家も利用できるように、提案した分析手法のためのプログラミングコードも公開している。

今後の展望として、口頭剖検調査データの特徴をさらに明示的に統計モデルに取り込んだ上で、口頭剖検による死因分析を行い、その予測精度の向上を目指す。1つの方向性は、空間情報を統計モデルに組み込むことである。伝染病や風土病などによる死者が多い場合、死因分布はその地域の空間・地理的な要因に大きく依存すると思われる。また、近くに位置する調査地点の死因分布は遠く離れた場所のものより似ていると考えるのが自然であるため、空間的に近い場所からのデータに重きを置いて死因分布の推定を行う方が効率的だと予想される。課題として挙げ

られることは、空間情報とデータの他の特徴をどのように同時に統計モデル化するかである。口頭剖検のための質問項目数は100以上に及ぶこともあり、遺族がすべての質問に回答するとは限らないことに加え、予算の都合から質問数が減ることもあるため、結果としてデータの中に多くの欠損値が存在することになる。そのため、多項ロジットモデルのような一般的な統計手法を用いるためには欠損値の補完が必要となるが、高次元データにおける欠損値の補完は容易ではなく、利用する補完方法に結果が大きく依存する可能性が高い。そこで、欠損値の補完を必要としないアプローチとして本研究で行ったように、死因が与えられた下での症状の条件付き確率をモデル化する方法があるが、高次元多変量データに空間情報を組み込むことは一般的に容易ではないため、複雑になり過ぎず計算可能なベイズ統計モデルの開発が必要となる。

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0件)

[学会発表](計4件)

1. Kunihama, T., Li, Z., Clark, S. and McCormick, T. "Bayesian estimation of distributions of causes of death with verbal autopsy surveys", 統計関連学会連合大会, 南山大学, 2017年9月
2. Kunihama, T., Li, Z., Clark, S. and McCormick, T. "Bayesian estimation of distributions of causes of death with verbal autopsy surveys", The Second ISBA-EAC Conference, Changchun, 2017年7月(招待講演)
3. Kunihama, T., Li, Z., Clark, S. and McCormick, T. "Bayesian estimation of distributions of causes of death with verbal autopsy surveys", The 1st International Conference on Econometrics and Statistics, Hong Kong, 2017年6月(招待講演)
4. Kunihama, T., Li, Z., Clark, S. and McCormick, T. "Bayesian estimation of distributions of causes of death with verbal autopsy surveys", 研究会「高次元・高頻度データのベイズ計量経済分析」東北学院大学, 2017年2月(招待講演)

[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：
出願年月日：
国内外の別：

取得状況（計0件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究代表者

國濱 剛 (KUNIHAMA, Tsuyoshi)
関西学院大学・経済学部・講師
研究者番号：40779716

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：

(4) 研究協力者

()