

審査の結果の要旨

論文題目：Mixed Modeling Approach to Small Area Estimation

題目和訳：混合効果モデルによる小地域推定へのアプローチ

氏名：菅澤 翔之助

本論文は、小地域推定問題に対して混合効果モデル (Mixed Effects Models) を用いた手法の開発に取り組んでいる。混合効果モデルとは、線形回帰による母数効果と確率変数による変量効果によって構成されるモデルのことで、小地域推定の枠組みでは地域効果を変量効果とするモデルが用いられる。その代表的なモデルとして Fay-Herriot (FH) モデルと枝分かれ誤差回帰モデル (Nested Error Regression, NER) モデルがあり、前者は集計データを解析するために用いられ、後者は個票データが利用可能なときに使われている。両モデルはこの分野の基本モデルとして非常に広く用いられているが、分布の正規性や誤差項の同質性など強い仮定が要求されるので、実際のデータ解析の様々な場面でモデルやそのモデルから誘導される推測方法の妥当性が問題になる場合がある。そこで本論文では、既存の混合効果モデルを優越する柔軟なモデルと推測方法を提案し、それに付随する理論結果と数値解析・データ解析を行っている。具体的には以下のような章立てで構成される。

- 第1章 序
- 第2章 Fay-Herriot モデルにおける応答変数の変換
- 第3章 適用型変換をもつ混合効果モデル
- 第4章 混合効果モデルにおける条件付き平均 2 乗誤差
- 第5章 異質性をもつ枝分かれ誤差回帰モデル
- 第6章 平均と分散の同時縮小
- 第7章 変量効果の不確実性
- 第8章 不確実性をもつ経験ベイズ法

本論文で扱う小地域推定の背景、各章のモチベーションと概要を第1章で与えた上で、第2章から続く7つの章で研究成果をまとめている。

小地域推定の分野で従来使われてきた Fay-Herriot モデルと枝分かれ誤差回帰モデルの欠点の一つは、誤差項に正規性を仮定している点である。所得、支出、価格、生産量などの正值データを解析する場合には、データの分布は非対称であり正規性を仮定することは適切でない。代替策として対数変換が用いられることが多いが、必ずしも対数変換が妥当であるとは限らない。また Box-Cox 変換を用いて解析することも考えられるが、この変換の変換母数の最尤推定量は一致性を持たないことが知られており、理論構築は極めて困難である。第2章及び第3章は、この問題に対して理論的にも優れた混合効果モデルの提案を行っている。具体的には、Yang (2006) によって提案された双対巾変換 (Dual Power Transformation) を用いた変量効果モデルを提案している。これは従来のモデルや対数変換に基づいたモデルを含むとともに、データの情報から適切な変換を選択し柔軟に推定を行うことを可能にする。最大の問題であった変換母数の推定についても、最尤推定量の一致性を示すことができ、経験ベイズ推定量の導出とその平均 2 乗誤差 (MSE) の漸近 2 次近似を行っている。第2章では、集計データを解析するための Fay-Herriot モデルに双対巾変換を用いたモデルを扱い、推測手法の理論構築とともに数値実験を通して有用性を検証している。特に、対数変換を利用したモデルや変換を施さないモデルでは、変換が誤って特定されたケースにおいて推定精度が悪くなることを示している。一方、双対巾変換を用いた提案手法では、全てのケースで既存手法の性能に対して同等か優越することを確認している。第3章では、個票データを解析するための枝分かれ誤差回帰モデルに双対巾変換

を用いたモデルを扱い、同様な解析結果を与えている。ただし、Fay-Herriot モデルの場合とは異なり、有限母集団の枠組みで扱うことによって、母集団平均の推定と変換モデルとの関係を明確にし、経験ベイズ推定だけでなく経験ベイズ信頼区間の構成も行っている。

第 4 章では、小地域推定の経験ベイズ推定量の推定誤差を測る基準として従来の平均 2 乗誤差 (Mean Squared Error, MSE) の代わりに条件付き平均 2 乗誤差 (Conditional Mean Squared Error, CMSE) を用いて解析を行っている。条件付き平均 2 乗誤差は Booth and Hobert (1998) によって提案された基準で、推定したい小地域のデータを与えたときの条件付き平均 2 乗誤差を考えることになる。これは、実際の当該地域のデータが与えられたときに推定誤差がどの程度になるかを見積もることができる点で、通常の平均 2 乗誤差とは異なる推定誤差を調べることができる。Fay-Herriot モデルや枝分かれ誤差回帰モデルのように分布に正規性を仮定したモデルにおいて、条件付き平均 2 乗誤差と通常の平均 2 乗誤差とは漸近的に等しくなるが、非正規分布のもとでは必ずしも等しくないことが知られている。第 4 章では、分散が平均の 2 次関数で表される自然指数分布族と変量効果とを組み合わせた一般的な混合効果モデルを考え、条件付き平均 2 乗誤差の考察を行っている。この一般的なモデルは計数データや 2 値データを解析するためのモデルを含む広いクラスであり、条件付き平均 2 乗誤差が正規分布以外のケースでは平均 2 乗誤差と漸近的に異なることが示される。また、実際に条件付き平均 2 乗誤差を利用するためにその漸近不偏推定量を導出している。

第 5 章では、分散不均一なデータに対するモデルの開発を行なっている。枝分かれ誤差回帰モデルでは観測値の分散はすべての地域で等しいことを仮定しているが、Jiang and Nguyen (2012) は、その仮定が必ずしも成り立たないことを実データの例を通して示している。Jiang and Nguyen (2012) は分散不均一な枝分かれ誤差回帰モデルを提案しているが、そのモデルのパラメータ数は地域の数に応じて増大してしまうため安定な推定を行うことが困難である。そこで、第 5 章では、不均一分散は共変量で説明されるケースが多くあることに注目し、不均一分散が共変量の関数で与えられる構造を持つ枝分かれ誤差回帰モデルを提案している。このモデルに対してモーメント法による一致推定量を提案し、小地域パラメータに対する経験最良線形不偏予測量 (EBLUP: Empirical Best Linear Unbiased Predictor) を導出し、そのリスク評価のための MSE の漸近不偏推定量を導出している。数値実験および実データ解析によって、この提案手法と先行研究で与えられた様々な手法との比較を行い、その結果、提案手法が多くの場合に良いパフォーマンスを示すことを確認している。

第 6 章では、平均と分散の同時縮小モデルの開発を行なっている。Fay-Herriot モデルにおいて、各小地域のサンプリング分散を既知と仮定してモデルを推定するのが通常の方法であるが、実際に使用する場合は推定したサンプリング分散を用いることになる。しかし、サンプリング分散の推定値は標本数が小さいと不安定になり、それを既知として利用したモデルから得られる推定値も不安定なものになってしまう。この問題に対し、Maiti et al. (2014) はサンプリング分散もモデルから安定的に推定する同時縮小モデルを提案した。Maiti et al. (2014) では周辺尤度最大化によるパラメータ推定の方法を提案しているが、モデルの複雑さゆえにパラメータ推定値が安定しないという欠点を伴う。そこで同時縮小モデルの階層表現を利用して、マルコフ連鎖モンテカルロ法 (MCMC) による推定方法を提案している。導出された MCMC は簡単な分布からの乱数発生 of の繰り返しのみで実行できることがわかり、数値実験においても Maiti et al (2014) の手法を優越することを確認している。

第 7 章、第 8 章は変量効果の不確実性の問題を扱っている。地域によっては変量効果

が不必要なケースがあり、Datta and Mandal (2015)では、変量効果がある確率でモデルに含まれないことを許すモデルを Fay-Herriot モデルの枠組みで提案し、Fay-Herriot モデルに対して良い性能をもつことを示した。第 7 章は、このアイデアを枝分かれ誤差回帰モデルに適用し、MCMC による推定方法を提案している。さらに、有限母集団に対する推定問題への適用も行い、有限母集団内のサンプル数が多くなるほど提案手法が既存手法よりも優越する結果を数値実験から得ることができる。第 8 章は、第 4 章で扱った一般的なモデルの枠組みで Datta and Mandal (2015)のアイデアを適用することで、変量効果の不確実性の議論をカウントデータや 2 値データを含む枠組みまで広げることを考えている。そのために、小地域パラメータの事前分布として、Ghosh and Maiti (2004)で使われた共役事前分布とある 1 点分布の混合分布を用いることを提案している。この設定のもと、Expectation-Maximization (EM) アルゴリズムによるパラメータ推定法を提案し、小地域平均の推定量を導出している。そして、数値実験を通して提案手法が既存手法よりも良い性能をもつことを確認している。

論文の評価

市区町村や更に細かい地域の特性値を推定しようとする、データ数が少なくなるためその地域のデータだけで推定したのでは大きな推定誤差が生じてしまう。これを小地域推定問題という。この問題を解決するために全領域で共通な母数として埋め込まれる固定効果と地域毎に異なる変量効果とを組み入れた混合効果モデルが使われる。固定効果を共通母数に基づいてとることで全データを用いて安定した推定を与えることができ、地域効果を変量にとることで小地域毎の標本平均を全データに基づいた安定した推定値の方向へ縮小する効果が生み出されることになる。このようにして、混合効果モデルを用いることにより、そのモデルから導かれる経験ベイズ推定量もしくは経験最良線形不偏予測量は縮小効果により推定値の安定化と推定精度の改善を図ることができる。実際どの程度の改善がなされるのかについては、平均 2 乗誤差の漸近評価並びにその 2 次漸近不偏推定値を用いることにより評価することができる。ここでいう漸近理論は、地域の個数を大きくしていく意味での漸近理論であり、地域毎のデータ数は有界であることに注意する。

本論文では、家計調査や地価公示価格などのデータを解析する上で分布の正規性を仮定する従来の混合効果モデルが適切でないことを示し、そうした様々な問題を解決するモデルの工夫を提案し、推測を行う上での理論結果を示すとともに数値計算やデータ解析を通して有用性を確認している。第 2 章と第 3 章で扱った双対巾変換を用いた混合効果モデルについての研究は最初のものであり、変換母数の一致推定量を与え、それに基づいた経験ベイズ推定量の導出と平均 2 乗誤差に関する漸近理論を与えたことは、この分野への大きな貢献として評価できる。第 3 章で議論された条件付き平均 2 乗誤差と通常平均 2 乗誤差とが、正規分布のときには漸近的に同等であるのに対して、非正規分布においては漸近的に異なるという結果は、興味深い。第 5 章と第 6 章は地域毎の異質性を扱った章であり、分散の異質性を分散関数を用いて表現するモデルが第 5 章であり、分散を変量として扱うモデルが第 6 章である。いずれも、理論的な解析結果を詰め、数値計算のためのアルゴリズムの提示、実際のデータ解析による有用性の確認を行っており、独創性、有用性が評価される。第 7 章と第 8 章で扱った変量効果の不確実性の問題は最近活発に研究されているテーマである。常に世界の最新の研究動向に注目し、データ解析の視点から研究テーマの動機付けを明らかにして、モデルの開発、推測手法の漸近的性質の導出、数値実験による精査、データ解析による有用性の検証を行っており、そうした真摯な研究姿勢や優れた研究能力は高く評価される。また第 2 章から第 8 章までの各章はそれぞれ一流の国際雑誌に掲載されている。

以上，説明してきたように，本論文は，小地域推定を行うための混合効果モデルの開発に関して，菅澤翔之助氏自身の独創的で優れた研究成果をまとめたものである。その中で提案されたモデルとそのモデルから導かれる経験ベイズ推定量及び統計的推測のための理論結果は小地域推定などの応用分野において利用価値の高いものであり，この分野における貴重な貢献であると高く評価できる。従って，論文審査委員会は全員一致で，菅澤翔之助氏が博士（経済学）の学位を授与されるにふさわしいという結論に達した。

審査委員
久保川達也
大森裕浩
下津克己
加藤賢悟
入江薫