

## 科学研究費助成事業 研究成果報告書

平成 26 年 5 月 28 日現在

機関番号：17501

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700339

研究課題名(和文)母集団特性値推定のための事前・補助情報を組み込む新しい層別推定量の開発と応用

研究課題名(英文)Development of new stratified estimators that incorporate prior and auxiliary information for estimating population characteristics and its applications

研究代表者

大山 哲司(Ohyama, Tetsuji)

大分大学・医学部・助教

研究者番号：60574085

交付決定額(研究期間全体)：(直接経費) 1,600,000円、(間接経費) 480,000円

研究成果の概要(和文)：層別無作為抽出を利用した標本調査により母集団特性値を推定することは、様々な研究分野で行われている。その際、補助変数を利用することで通常の不偏推定量より精度の高い推定量が得られ、比推定量や回帰推定量、それらを包含するキャリブレーション推定量などがある。一方で標本調査は定期的に行われるなどして、母集団特性値に関する事前情報が得られている場合がある。本研究では従来のように補助変数を利用するだけでなく、事前情報も利用することでさらに推定精度の高い推定量を開発した。

研究成果の概要(英文)：In various research areas, sample surveys based on the stratified random sampling are often conducted to estimate population characteristics. We can obtain more effective estimators, such as ratio, regression, and calibration estimators, than usual unbiased estimators by incorporating auxiliary variables. On the other hand, it is often the case that we have prior information on population characteristics since sample surveys are often regularly conducted. In this research, we developed more effective estimators by using not only auxiliary variables but also prior information.

研究分野：情報学

科研費の分科・細目：統計科学

キーワード：キャリブレーション 層別無作為抽出 事前情報 補助変数

### 1. 研究開始当初の背景

医学研究をはじめとして社会調査や世論調査などでは、ある母集団における興味ある変数の母平均や母分散などの母集団特性値を推定するために、母集団から単純無作為抽出を行う標本調査がよく用いられる。母集団がいくつかの異質な部分集団から成り立っていることが事前に分かっている場合には、その部分集団ごとに無作為抽出を行う層別無作為抽出法を用いることで、母集団特性値の推定精度を単純無作為抽出法よりも向上させることができる。そのため、層別無作為抽出法を利用した調査は様々な分野で行われている。また、事前に層別できない場合であっても、事後的に層を構成することによって推定精度を向上させることができるため、標本調査における層別推定量の利用は非常に有用である。

しかし近年では、訪問調査のような従来型の調査方法は回収率が著しく低下し、調査環境が悪化していることが指摘されている。回収率が低下すると標本サイズが小さくなるために推定量の分散が大きくなるほか、推定におけるバイアスが生じる可能性がある。このような問題に対処するには、当然ながら事前に調査方法を慎重に計画することが最も重要であるが、それでも無回答を完全に避けることはできない。データを得た後の事後的な対処法として、無回答標本に対して値を代入して補完する方法、調査不能の理由を明確にしてこれを補助情報として利用する方法などが研究されている。後者のように補助情報を利用する方法としては、例えば層別比推定量や層別回帰推定量があり、これらをさらに発展させた推定量も研究されている。また、事前に母集団を層別することができなかった場合でも事後的に標本を層化し、推定精度を高める事後層化推定量や、2つ以上の補助変数で事後的に層を構成するレイキング比推定量などもある。さらには、これらをより一般化したキャリブレーション推定量が提案されており、調査における無回答に伴うバイアスの低減、推定精度向上のために用いられている。

一方、調査研究は定期的に行われてデータが蓄積されていたり、類似の研究が過去に行われていたりして、母集団特性値についての事前情報が得られている場合が多い。母集団特性値を推定する際には、これを有効に利用することで推定精度の向上が期待できる。その方法としてベイズ流の推定方法が考えられるものの、事前情報に対して分布を仮定する必要がある。そこで層別無作為抽出の場合において、分布を仮定することなく母集団特性値の事前情報を組み込んだ推定量を本研究の前に開発した。事前情報を組み込んだ推

定量の平均2乗誤差と補助変数を用いない不偏推定量の平均2乗誤差を理論的に比較するとともに、シミュレーションによる比較も行って推定精度が向上することを示した。しかしながらこの推定量では、事前情報を用いるのみで補助変数を同時に利用することは考えていなかった。

### 2. 研究の目的

従来から用いられている母集団特性値の層別不偏推定量や、補助変数を組み込んだ比推定量、回帰推定量、それらを包括的に含むキャリブレーション推定量などに対し、さらに過去の調査や別の調査等で得られている母集団特性値の事前情報を組み込むことを考える。特にベイズ流の推定のように特定の分布の仮定を置くことなく事前情報を組み込むことを考え、従来の推定量より精度の高い推定量を開発する。

また、実際の調査に対して応用する場合に考えられる問題として、

- (1) 事前情報の質をどのように見極め、事前情報を用いた方が良い場合と逆に使わない方が良い場合をどのように判断するか
- (2) あるいは近年の調査で大きな問題となっている無回答例が存在する場合にどのように対応するか

の2つを数理的に検討するとともに、シミュレーションによる検討も行う。

### 3. 研究の方法

補助情報を利用する層別推定量は、大きく次の2種類に分けられる。

- (1) 比推定量や回帰推定量を包括的に含むキャリブレーション推定量。
- (2) 未知パラメータを用いていったん推定量の形を仮定しておき、平均2乗誤差を最小にするように未知パラメータを与え推定量を構成する。

(1)は補助変数の情報を利用して各層の重みを調整した推定量であり、比推定量や回帰推定量の他にも事後層化推定量やレイキング比推定量などが含まれる。(2)の形式の推定量は古くから、また近年でも多く研究されている。本研究でははじめにそれらのレビューを行った。

同時に、すでに提案した各層の不偏推定量に対して事前情報を組み込んだ層別推定量に対して、キャリブレーションの考え方により補助変数を組み込む推定量の開発を行った。理論的な考察として、開発した推定量の平均2乗誤差を導出し、従来から用いられている不偏推定量、結合型回帰推定量と事前情報のみを用いた推定量との精度の比較を行った。また、母集団の分布に正規分布を仮定した場合のシミュレーションを行い、従来の推定量との比較検討を行った。

また、平均2乗誤差を最小とする推定量の開発とその理論的評価を行った。また、開発

した推定量の有用性を数値実験により調べた。

#### 4. 研究成果

層別無作為抽出法において興味のある変数についての各層での事前情報が得られている場合を想定し、各層の不偏推定量と事前情報の加重平均を考え、それらを各層の重みを用いて加重平均した推定量を考え、この推定量自体は事前情報を組み込んだ層別推定量として以前提案したものであるが、これに対してさらに補助情報を組み込んでより精度の高い推定量を構築するため、キャリブレーション推定量で用いられている方法を利用した。ここでは、各層の補助変数の真値は用いず母集団での補助変数の真値のみを用い、新しい推定量を考案した。この推定量は、ある条件下では従来の結合型回帰推定量や事前情報のみを組み込んだ推定量より平均2乗誤差を小さくすることができ、精度の高い推定量であることが示された。ただしいくつかの未知パラメータが含まれているため、実際のデータに適用する際にはその未知パラメータをデータから推定する必要がある。その場合でも従来の推定量より精度の高い推定量であるかをシミュレーションにより調べた。シミュレーションでは母集団の分布に正規分布を仮定し、事前情報と真値の差、興味のある変数や補助変数のバラツキを変えた検討を行った。その結果、提案した推定量はほとんどの場合で通常の不偏推定量を大きく上回る精度をもち、また結合型回帰推定量や事前情報のみを利用した推定量と比べても多くの場合で精度が高いことが確認できた。特に、各層の事前情報が真値に近いときほど、興味のある変数と補助変数との相関が強いときほど、また興味のある変数の変動係数が小さいときほど、提案した推定量は他の推定量と比べて精度が高いことが明らかになった。

上記の研究は母集団における補助変数の真値が既知である場合の考察であったが、各層における真値も既知である場合の考察も行った。各層の不偏推定量と事前情報の加重平均に対して、個別型回帰推定量と同様の回帰構造で補助変数を組み込んだ形を仮定し、その平均2乗誤差を最小にするように回帰係数を定めた。このようにして得られた推定量は、事前情報のみを用いた推定量や個別型回帰推定量よりも精度の高い推定量であることが理論的に示された。シミュレーションでも有用な推定量であることが確認された。

また、近年提案された指数比型推定量に対して事前情報を組み込んだ推定量を、単純無作為抽出の場合と層別無作為抽出の場合で考察した。指数比型推定量は、不偏推定量やその他様々な推定量よりも、ある条件下では有効な推定量であることが示されており、それに対してさらに事前情報を組み込むことで、より精度の高い推定量が得られた。この

ことはシミュレーションにより数値的にも確かめることができた。

補助変数と同時に事前情報も組み込む推定量の開発はできたが、無回答例が存在する場合の対応については十分な考察はできなかった。この点を実際のデータをもとに検討することが今後の課題である。

#### 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 1 件)

[1] Tetsuji Ohyama

Prior value incorporated calibration estimator in stratified random sampling, *Statistics and Probability Letters*, 83, 2013, 46-51, 査読あり, doi:10.1016/j.spl.2012.08.023

[学会発表](計 2 件)

[1] 大山哲司, 江島伸興

処置前後での連続な bounded outcome score の変化に対する 2 変量ベータ分布を用いた解析, 日本計算機統計学会第 27 回シンポジウム, 2013 年 11 月 15 日, 市民会館崇城大学ホール

[2] 大山哲司

層別無作為抽出法において事前情報と補助変数を組み込んだ母集団特性値の推定, 統計関連学会連合大会, 2012 年 9 月 10 日, 北海道大学

[図書](計 件)

[産業財産権]

出願状況(計 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 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

大山 哲司 (OHYAMA TETSUJI)

大分大学・医学部・助教

研究者番号：60574085

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )

研究者番号：