# 層別2段抽出法におけるブートストラップ法を用いた推定量の分散の推定についての一考察

新潟大学教育学部　伏木 忠義

**概 要**

標本調査において推定量の分散を推定することは基本的な問題の一つである．推定量が単純なものである場合は，分散の推定はそれほど難しいものではないが，複雑な推定量を用いる場合にはブートストラップ法を用いることが候補の一つとなる．本研究では，社会調査でよく利用される確率比例抽出—単純無作為抽出による層別 2 段抽出法を仮定し，ブートストラップ法を用いた推定量の分散の推定を扱う．

## 1　はじめに

標本調査において推定量の分散を推定することは基本的な問題の一つである．推定量が単純なものである場合には，分散に対する推定量は理論的にも扱えるものだが，複雑な推定量を用いる場合には理論的に扱うことが難しい場合もある．たとえば、標本調査において母集団平均や母集団総計を推定する場合の推定量の分散の推定量は陽に与えられている (たとえば，土屋 (2009) などを参照) が，複雑な量を推定する場合は推定量の分散の推定量が与えられていないこともある．そのような場合にはブートストラップ法を用いることが一つの選択肢となる．

一方，有限母集団に対して通常のブートストラップ法を適用すると偏りのある推定量の分散の推定量が得られることが知られている．有限母集団におけるブートストラップ法に関しては，Shao & Tu (1995) や Wolter (2007) などに全般的な説明がある．日本語で読める文献としては，標本誤差推計研究会 (1998) などが存在する．

社会調査における標本調査のデザインとして層別 2 段抽出法が利用されることがしばしばある．特に 1 段目で確率比例抽出を行い，2 段目で非復元単純無作為抽出を行うとある場合には自己加重標本となるため，このようなデザインが用いられることがしばしばある．本研究ではこのような設定において推定量の分散を推定するためにブートストラップ法を行う方法について考える．また，未回収がある場合のブートストラップ法に関して記述する．

本論文の構成は以下の通りである．第 2 節では，確率比例抽出—単純無作為抽出の 2 段抽出法における母集団割合に対するブートストラップ法を記述する．第 3 節では，層別 2 段抽出法における母集団割合に対するブートストラップ法を記述する．第 4 節では層別 2 段抽出法における複雑な推定量の分散の推定を行うブートストラップ法を，第 5 節ではさらに調査不能がある場合のブートストラップ法について記述する．第 2 節，第 3 節は既存の文献に記述された方法をまとめたものであり，第 4 節，第 5 節の結果は第 2 節，第 3 節の結果をもとにいくらか一般の状況でブートストラップ法を適用する場合に関して若干の考察を行った結果を記している．

# 2　PPR-SI におけるブートストラップ法

標本を $S$，母集団を $U$ とし，その要素数を $n$，$N$ とする．変数 $y$ の母集団総計 $\tau_y$ あるいは母集団平均 $\mu_y = \tau_y/N$ が目的の知りたい量とする．

ここでは 2 段抽出法を扱う．

1 段目では $M$ 個の PSU を $U_1, U_2, \ldots, U_M$ から確率比例による復元抽出を行う．ここで，$U = U_1 \cup U_2 \cup \cdots \cup U_M$ であり，$I = \{1, 2, \ldots, M\}$ とする．ただし，$U_a$ の要素数は $N_a$ であり，$\sum_{a \in I} N_a = N$ とする．

1 段目で選ばれた PSU を $a_1, a_2, \ldots, a_m$ とする．2 段目では，1 段目で選ばれた $k$ 番目の PSU $U_{a_k}$ から無作為非復元抽出で要素を抽出したものを $S_k$ で表す．ただし，$S_k$ の要素数は $n_{a_k}$ とする．標本全体は $S$ とし，標本サイズ $n$ は $n = \sum_{j=1}^{m} n_{a_j}$ となる．要素 $i$ の包含確率を $\pi_i$ とすると，$i$ が PSU $U_a$ に属する場合は，$\pi_i = \frac{N_a}{N}\frac{n_a}{N_a} = \frac{n_a}{N}$ で与えられる．

## 2.1　1 段目で 1 個の PSU のみが抽出される場合

まず，2 段抽出の 1 段目で 1 つの PSU $U_a$ のみが抽出される場合を考えよう．変数 $y$ の母集団平均 $\mu_y$ の HT 推定量は，

$$\hat{\mu}^{(1)}_{y,\text{PPR-SI}} = \frac{1}{N}\sum_{i \in S}\frac{y_i}{\pi_i} = \frac{1}{N}\sum_{i \in S_1}\frac{y_i}{\frac{n_a}{N}} = \frac{1}{n_a}\sum_{i \in S_1} y_i = \bar{y}_1$$

で与えられる．ただし，$\bar{y}_1$ は $S_1$ における $y$ の平均である．

このとき，$\hat{\mu}^{(1)}_{y,\text{PPR-SI}}$ の期待値は，

$$\mathrm{E}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}] = \mathrm{E}[\,\mathrm{E}[\,\hat{\mu}^{(1)}_{y,\text{PPR-SI}}|U_a]\,] = \mathrm{E}[\mu_{y,a}] = \sum_{a \in I} p_a \mu_{y,a} = \sum_{a \in I}\frac{N_a}{N}\frac{1}{N_a}\sum_{i \in U_a} y_i = \frac{1}{N}\sum_{a \in I}\sum_{i \in U_a} y_i = \frac{1}{N}\sum_{i \in U} y_i = \mu_y$$

で，その分散は，

$$\begin{aligned}
\mathrm{V}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}] &= \mathrm{V}[\,\mathrm{E}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}|U_a\,]\,] + \mathrm{E}[\,\mathrm{V}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}|U_a\,]\,] \\
&= \mathrm{V}[\mu_{y,a}] + \mathrm{E}\left[\frac{1}{n_a(N_a - 1)}\left(1 - \frac{n_a}{N_a}\right)\sum_{i \in U_a}(y_i - \mu_{y,a})^2\right] \\
&= \sum_{a \in I} p_a(\mu_{y,a} - \mu_y)^2 + \sum_{a \in I} p_a \frac{1}{n_a(N_a - 1)}\left(1 - \frac{n_a}{N_a}\right)\sum_{i \in U_a}(y_i - \mu_{y,a})^2
\end{aligned}$$

で与えられる．ただし，$\mu_{y,a} = \sum_{i \in U_a} y_i/N_a$，$p_a = N_a/N$ である．

## 2.2　1 段目で $m$ 個の PSU が抽出される場合

1 段目で抽出された PSU を $a_1, a_2, \ldots, a_m$ とする．変数 $y$ の母集団平均 $\mu_y$ の推定量は，

$$\hat{\mu}_{y,\text{PPR-SI}} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{N}\sum_{i \in S_j}\frac{y_i}{\frac{n_{a_j}}{N}} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n_{a_j}}\sum_{i \in S_j} y_i = \frac{1}{m}\sum_{j=1}^{m}\bar{y}_j = \bar{\bar{y}}$$

で与えられる．ここで，$\bar{\bar{y}} = \frac{1}{m}\sum_{j=1}^{m}\bar{y}_j$ である (必ずしも $y$ に関する標本平均とはならないことに注意する)．$\hat{\mu}_{y,\text{PPR-SI}}$ は，$m$ 個の独立な $\hat{\mu}^{(1)}_{y,\text{PPR-SI}}$ の平均と考えられるので，その期待値は

$$\mathrm{E}[\hat{\mu}_{y,\text{PPR-SI}}] = \frac{1}{m}\sum_{j=1}^{m}\mathrm{E}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}] = \mathrm{E}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}] = \mu_y$$

であり，分散は

$$\begin{aligned}\mathrm{V}[\hat{\mu}_{y,\text{PPR-SI}}] &= \frac{1}{m^2}\sum_{j=1}^{m}\mathrm{V}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}] = \frac{1}{m}\mathrm{V}[\hat{\mu}^{(1)}_{y,\text{PPR-SI}}] \\ &= \frac{1}{m}\left\{\sum_{a\in I}p_a(\mu_{y,a}-\mu_y)^2 + \sum_{a\in I}p_a\frac{1}{n_a(N_a-1)}\left(1-\frac{n_a}{N_a}\right)\sum_{i\in U_a}(y_i-\mu_{y,a})^2\right\} \qquad (1)\end{aligned}$$

で与えられる．

分散 $\mathrm{V}[\hat{\mu}_{y,\text{PPR-SI}}]$ の不偏推定量は

$$\hat{\mathrm{V}}[\hat{\mu}_{y,\text{PPR-SI}}] = \frac{1}{m(m-1)}\sum_{j=1}^{m}(\bar{y}_j-\bar{\bar{y}})^2$$

で与えられる．ここで，

$$\begin{aligned}\hat{\mathrm{V}}[\hat{\mu}_{y,\text{PPR-SI}}] &= \frac{1}{m(m-1)}\sum_{j=1}^{m}(\bar{y}_j-\mu_y)^2 - \frac{1}{m-1}(\bar{\bar{y}}-\mu_y)^2 \\ &= \frac{1}{m(m-1)}\sum_{j=1}^{m}(\bar{y}_j-\mu_{y,a_j})^2 + \frac{2}{m(m-1)}\sum_{j=1}^{m}(\bar{y}_j-\mu_{y,a_j})(\mu_{y,a_j}-\mu_y) + \frac{1}{m(m-1)}\sum_{j=1}^{m}(\mu_{y,a_j}-\mu_y)^2 \\ &\quad -\frac{1}{m-1}(\bar{\bar{y}}-\mu_y)^2\end{aligned}$$

より，

$$\begin{aligned}&\mathrm{E}[\hat{\mathrm{V}}[\hat{\mu}_{y,\text{PPR-SI}}]|\boldsymbol{a}=(a_1,a_2,\ldots,a_m)] \\ =\ & \frac{1}{m(m-1)}\sum_{j=1}^{m}\mathrm{E}[(\bar{y}_j-\mu_{y,a_j})^2|\boldsymbol{a}] + \frac{2}{m(m-1)}\sum_{j=1}^{m}(\mu_{y,a_j}-\mu_y)\mathrm{E}[\bar{y}_j-\mu_{y,a_j}|\boldsymbol{a}] + \frac{1}{m(m-1)}\sum_{j=1}^{m}(\mu_{y,a_j}-\mu_y)^2 \\ & -\frac{1}{m-1}\mathrm{E}[(\bar{\bar{y}}-\mu_y)^2|\boldsymbol{a}] \\ =\ & \frac{1}{m(m-1)}\sum_{j=1}^{m}\frac{1}{(N_{a_j}-1)n_{a_j}}\left(1-\frac{n_{a_j}}{N_{a_j}}\right)\sum_{i\in U_{a_j}}(y_i-\mu_{y,a_j})^2 + \frac{1}{m(m-1)}\sum_{j=1}^{m}(\mu_{y,a_j}-\mu_y)^2 - \frac{1}{m-1}\mathrm{E}[(\bar{\bar{y}}-\mu_y)^2|\boldsymbol{a}]\end{aligned}$$

となるが，(1) に注意してこの期待値をとると，

$$\begin{aligned}&\mathrm{E}[\hat{\mathrm{V}}[\hat{\mu}_{y,\text{PPR-SI}}]] = \mathrm{E}[\mathrm{E}[\hat{\mathrm{V}}[\hat{\mu}_{y,\text{PPR-SI}}]|\boldsymbol{a}]] \\ =\ & \frac{1}{m-1}\left\{\sum_{a\in I_a}p_a\frac{1}{(N_a-1)n_a}\left(1-\frac{n_a}{N_a}\right)\sum_{i\in U_a}(y_i-\mu_{y,a})^2 + \sum_{a\in I}p_a(\mu_{y,a}-\mu_y)^2\right\} - \frac{1}{m-1}\mathrm{V}[\hat{\mu}_{y,\text{PPR-SI}}] \\ =\ & \frac{m}{m-1}\mathrm{V}[\hat{\mu}_{y,\text{PPR-SI}}] - \frac{1}{m-1}\mathrm{V}[\hat{\mu}_{y,\text{PPR-SI}}] = \mathrm{V}[\hat{\mu}_{y,\text{PPR-SI}}]\end{aligned}$$

が得られ，$\hat{\mathrm{V}}[\hat{\mu}_{y,\text{PPR-SI}}]$ の不偏性が確かめられる．

### 2.3　$\mathrm{V}[\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}]$ を推定するためのブートストラップ法

次のようなブートストラップ法を考える．

Step 1.　$b = 1, 2, \ldots, B$ で以下を繰り返す．
・$(m-1)$ 個の部分標本 $S_1^{(b)}, S_2^{(b)}, \ldots, S_{m-1}^{(b)}$ を標本における $m$ 個の部分標本 $S_1, S_2, \ldots, S_m$ から復元抽出する．
・$\bar{y}_j^{(b)}$ を $S_j^{(b)}$ の平均として，$\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}^{(b)} = \frac{1}{m-1}\sum_{j=1}^{m-1}\bar{y}_j^{(b)}$ とする．
Step 2.　$\mathrm{V}[\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}]$ を

$$\hat{v}_{y,\mathrm{PPR\text{-}SI}}^{(B)} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}^{(b)} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}})^2$$

で推定．

このとき，$\mathrm{E_B}[\hat{v}_{y,\mathrm{PPR\text{-}SI}}^{(B)}] = \hat{\mathrm{V}}[\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}]$ が成り立つ．ただし，$\mathrm{E_B}$ はブートストラップ標本に関する期待値を意味する．このことは，以下のように確認される．

$$\mathrm{E_B}[\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}^{(b)}] = \mathrm{E_B}[\bar{y}_{j^{(b)}}] = \frac{1}{m}\sum_{j=1}^{m}\bar{y}_j = \bar{\bar{y}} = \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}$$

に注意すると，

$$\begin{aligned}
\mathrm{E_B}[\hat{v}_{y,\mathrm{PPR\text{-}SI}}^{(B)}] &= \mathrm{E_B}[(\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}^{(b)} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}})^2] \\
&= \mathrm{E_B}\left[\left(\frac{1}{m-1}\sum_{j=1}^{m-1}\bar{y}_{j^{(b)}} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}\right)^2\right] \\
&= \mathrm{E_B}\left[\left\{\frac{1}{m-1}\sum_{j=1}^{m-1}\left(\bar{y}_{j^{(b)}} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}\right)\right\}^2\right] \\
&= \frac{1}{(m-1)^2}\sum_{i,j=1}^{m-1}\mathrm{E}_B\left[\left(\bar{y}_{i^{(b)}} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}\right)\left(\bar{y}_{j^{(b)}} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}\right)\right] \\
&= \frac{1}{(m-1)^2}\sum_{j=1}^{m-1}\mathrm{E}_B\left[\left(\bar{y}_{j^{(b)}} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}\right)^2\right] \\
&= \frac{1}{(m-1)}\mathrm{E}_B\left[\left(\bar{y}_{j^{(b)}} - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}\right)^2\right] \\
&= \frac{1}{m(m-1)}\sum_{j=1}^{m}\left(\bar{y}_j - \hat{\mu}_{y,\mathrm{PPR\text{-}SI}}\right)^2 \\
&= \hat{\mathrm{V}}[\hat{\mu}_{y,\mathrm{PPR\text{-}SI}}]
\end{aligned}$$

が得られる．

## 3　STPPR-SI におけるブートストラップ法

PSU が $L$ 個の層にわかれているとしよう．第 $l$ 層に含まれる PSU の添え字集合を $I_l$ とし，$a \in I_l$ に対して $U_{l,a}$ が対応する PSU となる．また，それぞれの層における要素数を $N_1, N_2, \ldots, N_L$ とし ($\sum_{l=1}^{L} N_l = N$)，$U_{l,a}$ の要素数を $N_{l,a}$ と表す ($\sum_{a \in I_l} N_{l,a} = N_l$)．また，$W_l = N_l/N$，$W_{l,a} = N_{l.a}/N_l$ とする．

$I_l$ から $U_{l,a_{l,1}}, U_{l,a_{l,2}}, \ldots, U_{l,a_{l,m_l}}$ の $m_l$ 個の PSU を確率復元抽出し，抽出された PSU からは非復元単純無作為抽出を行う．得られた標本を $S_{l,1}, S_{l,2}, \ldots, S_{l,m_l}$ とする．また，$U_{l,a}$ からの標本サイズを $n_{l,a}$ とする．

このとき，$\mu_y$ の推定量として

$$\hat{\mu}_{y,\text{STPPR-SI}} = \sum_{l=1}^{L} W_l\,\hat{\mu}_{y,l,\text{PPR-SI}} = \sum_{l=1}^{L}\frac{W_l}{m_l}\sum_{j=1}^{m_l}\bar{y}_{l,j}$$

を考える．ただし，$\hat{\mu}_{y,l,\text{PPR-SI}} = \bar{\bar{y}}_l = \sum_{j=1}^{m_l}\bar{y}_{l,j}/m_l$ は第 $l$ 層において，$\mu_{y,l} = \sum_{i\in U_l} y_i/N_l$ を 2 節の方法で推定したものと考えることができる．ここで，$\bar{y}_{l,j} = \frac{1}{n_{l,a_{l,j}}}\sum_{i\in S_{l,a_{l,j}}} y_i$ とした．2 節の結果から $\hat{\mu}_{y,l,\text{PPR-SI}} = \mu_{y,l}$ となるので，

$$\mathrm{E}[\hat{\mu}_{y,\text{STPPR-SI}}] = \sum_{l=1}^{L} W_l \mathrm{E}[\hat{\mu}_{y,l,\text{PPR-SI}}] = \frac{1}{N}\sum_{l=1}^{L} N_l\mu_{y,l} = \frac{1}{N}\sum_{i\in U} y_i = \mu_y$$

が成り立つ．また，

$$\begin{aligned}\mathrm{V}[\hat{\mu}_{y,\text{STPPR-SI}}] &= \sum_{l=1}^{L} W_l^2 \mathrm{V}[\hat{\mu}_{y,l,\text{PPR-SI}}]\\ &= \sum_{l=1}^{L}\frac{W_l^2}{m_l}\left\{\sum_{a\in I_l} W_{l,a}(\mu_{y,l,a}-\mu_{y,l})^2 + \sum_{a\in I_l} W_{l,a}\frac{1}{n_{l,a}(N_{l,a}-1)}\left(1-\frac{n_{l,a}}{N_{l,a}}\right)\sum_{i\in U_{l,a}}(y_i-\mu_{y,l,a})^2\right\}\end{aligned}$$

となる．

$\mathrm{V}[\hat{\mu}_{y,\text{STPPR-SI}}]$ の推定量として

$$\hat{\mathrm{V}}[\hat{\mu}_{y,\text{STPPR-SI}}] = \sum_{l=1}^{L}\frac{W_l^2}{m_l(m_l-1)}\sum_{j=1}^{m_l}(\bar{y}_{l,j}-\bar{\bar{y}}_l)^2$$

を考える．これが $\mathrm{V}[\hat{\mu}_{y,\text{STPPR-SI}}]$ の不偏推定量となることは第 2 節と同様の計算を行うことで確認できる．

このとき次のようなブートストラップ法を考える．

Step 1.　$b = 1, 2, \ldots, B$ で以下を繰り返す．

・$l = 1, 2, \ldots, L$ それぞれにおいて $(m_l - 1)$ 個の部分標本 $S_{l,1}^{(b)}, S_{l,2}^{(b)}, \ldots, S_{l,(m_l-1)}^{(b)}$ を $m_l$ 個の部分標本 $S_{l,1}, S_{l,2}, \ldots, S_{l,m_l}$ から復元抽出する．

・$l = 1, 2, \ldots, L$ それぞれにおいて $\hat{\mu}_{y,l,\text{PPR-SI}}^{(b)} = \frac{1}{m_l-1}\sum_{j=1}^{m_l-1}\bar{y}_{l,j}^{(b)}$ とする．ただし，$\bar{y}_{l,j}^{(b)}$ を $S_{l,j}^{(b)}$ の平均としている．

・$\hat{\mu}_{y,\text{STPPR-SI}}^{(b)} = \sum_{l=1}^{L} W_l\,\hat{\mu}_{y,l,\text{PPR-SI}}^{(b)}$ とする．

Step 2.　$\mathrm{V}[\hat{\mu}_{y,\text{STPPR-SI}}]$ を

$$\hat{v}_{y,\text{STPPR-SI}}^{(B)} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\mu}_{y,\text{STPPR-SI}}^{(b)} - \hat{\mu}_{y,\text{STPPR-SI}})^2$$

で推定．

このとき，$\mathrm{E}_{\mathrm{B}}[\hat{v}_{y,\text{STPPR-SI}}^{(B)}] = \hat{\mathrm{V}}[\hat{\mu}_{y,\text{STPPR-SI}}]$ が成り立つことを以下のように確認できる．まず，

$$\begin{aligned}\mathrm{E}_{\mathrm{B}}[\hat{\mu}_{y,\text{STPPR-SI}}^{(b)}] &= \mathrm{E}_{\mathrm{B}}\left[\sum_{l=1}^{L} W_l\,\hat{\mu}_{y,l,\text{PPR-SI}}^{(b)}\right] = \sum_{l=1}^{L} W_l\,\mathrm{E}_{\mathrm{B}}\left[\bar{y}_{l,j}^{(b)}\right]\\ &= \sum_{l=1}^{L} W_l\,\mathrm{E}_{\mathrm{B}}\left[\bar{y}_{l,j}^{(b)}\right] = \sum_{l=1}^{L}\frac{W_l}{m_l}\sum_{j=1}^{m_l}\bar{y}_{l,j} = \hat{\mu}_{y,\text{STPPR-SI}}\end{aligned}$$

が成り立つ．これから，

$$
\begin{aligned}
\mathrm{E_B}[\hat{v}^{(B)}_{y,\text{STPPR-SI}}] &= \mathrm{E_B}[(\hat{\mu}^{(b)}_{y,\text{STPPR-SI}} - \hat{\mu}_{y,\text{STPPR-SI}})^2] = \mathrm{E_B}\left[\left(\sum_{l=1}^{L} \frac{W_l}{m_l-1}\sum_{j=1}^{m_l-1}\bar{y}^{(b)}_{l,j} - \sum_{l=1}^{L} W_l \bar{\bar{y}}_l\right)^2\right] \\
&= \mathrm{E_B}\left[\left\{\sum_{l=1}^{L} \frac{W_l}{m_l-1}\sum_{j=1}^{m_l-1}\left(\bar{y}^{(b)}_{l,j} - \bar{\bar{y}}_l\right)\right\}^2\right] = \mathrm{E_B}\left[\sum_{l=1}^{L} \frac{W_l^2}{(m_l-1)^2}\sum_{j=1}^{m_l-1}\left(\bar{y}^{(b)}_{l,j} - \bar{\bar{y}}_l\right)^2\right] \\
&= \sum_{l=1}^{L} \frac{W_l^2}{(m_l-1)}\mathrm{E_B}\left[\left(\bar{y}^{(b)}_{l,j} - \bar{\bar{y}}_l\right)^2\right] = \sum_{l=1}^{L} \frac{W_l^2}{(m_l-1)}\frac{1}{m_l}\sum_{j=1}^{m_l}\left(\bar{y}_{l,j} - \bar{\bar{y}}_l\right)^2
\end{aligned}
$$

が成り立つ．

(注意) この段階では，

Step 1.　$b = 1, 2, \ldots, B$ で以下を繰り返す．
・$l = 1, 2, \ldots, L$ それぞれにおいて $(m_l - 1)$ 個の部分標本 $S^{(b)}_{l,1}, S^{(b)}_{l,2}, \ldots, S^{(b)}_{l,(m_l-1)}$ を $m_l$ 個の部分標本 $S_{l,1}, S_{l,2}, \ldots, S_{l,m_l}$ から復元抽出する．
・$l = 1, 2, \ldots, L$ それぞれにおいて $\hat{\mu}^{(b)}_{y,l,\text{PPR-SI}} = \frac{1}{m_l-1}\sum_{j=1}^{m_l-1}\bar{y}^{(b)}_{l,j}$ とする．ただし，$\bar{y}^{(b)}_{l,j}$ を $S^{(b)}_{l,j}$ の平均としている．
Step 2.　$\mathrm{V}[\hat{\mu}_{y,\text{STPPR-SI}}]$ を

$$
\hat{v}^{(B)}_{y,\text{STPPR-SI}} = \sum_{l=1}^{L} \frac{W_l^2}{B}\sum_{b=1}^{B}(\hat{\mu}^{(b)}_{y,l,\text{PPR-SI}} - \hat{\mu}_{y,l,\text{PPR-SI}})^2
$$

で推定．

というブートストラップ法を用いても良い．

## 4　推定量が $\hat{\theta} = g(\boldsymbol{\mu}_{\boldsymbol{x},\text{STPPR-SI}})$ と表される場合

本節では，STPPR-SI を仮定し，推定したい量 $\theta$ が $\theta = g(\boldsymbol{\mu}_{\boldsymbol{x}})$ と表され，その推定量が $\hat{\theta} = g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}})$ と表される場合に $\mathrm{V}(\hat{\theta})$ を推定するためのブートストラップ法を扱う．

まず，

$$
\begin{aligned}
\hat{\theta} &= g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}}) \\
&\approx g(\boldsymbol{\mu}_{\boldsymbol{x}}) + \nabla g(\boldsymbol{\mu}_{\boldsymbol{x}})^T(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}} - \boldsymbol{\mu}_{\boldsymbol{x}})
\end{aligned}
$$

より，

$$
\mathrm{V}[\hat{\theta}] \approx \nabla g(\boldsymbol{\mu}_{\boldsymbol{x}})^T \mathrm{V}[\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}}]\nabla g(\boldsymbol{\mu}_{\boldsymbol{x}})
$$

が得られる．

$\boldsymbol{\mu}_{\boldsymbol{x}} = (\mu_1, \mu_2, \ldots, \mu_p)$ とすると，$\mathrm{V}[\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}}]$ は $p \times p$ 行列だが，例えばその $(1, 2)$ 要素 $\mathrm{E}[(\hat{\mu}_{1,\text{STPPR-SI}} - \mu_1)(\hat{\mu}_{2,\text{STPPR-SI}} - \mu_2)]$ は

$$
\begin{aligned}
&\mathrm{E}[(\hat{\mu}_{1,\text{STPPR-SI}} - \mu_1)(\hat{\mu}_{2,\text{STPPR-SI}} - \mu_2)] \\
&= \sum_{l=1}^{L}\frac{W_l^2}{m_l}\sum_{a\in I_l}\frac{N_{l,a}}{N_l}\left\{(\mu_{1,l,a} - \mu_{1,l})(\mu_{2,l,a} - \mu_{2,l}) + \frac{1}{n_{l,a}(N_{l,a}-1)}\left(1 - \frac{n_{l,a}}{N_{l,a}}\right)\sum_{i\in U_{l,a}}(x_{1,i} - \mu_{1,l,a})(x_{2,i} - \mu_{2,l,a})\right\}
\end{aligned}
$$

で与えられ，

$$\hat{\mathrm{E}}[(\hat{\mu}_{1,\mathrm{STPPR\text{-}SI}}-\mu_1)(\hat{\mu}_{2,\mathrm{STPPR\text{-}SI}}-\mu_2)]=\sum_{l=1}^{L}\frac{W_l^2}{m_l(m_l-1)}\sum_{j=1}^{m_l}(\bar{x}_{1,l,j}-\bar{\bar{x}}_{1,l})(\bar{x}_{2,l,j}-\bar{\bar{x}}_{2,l})$$

は $\mathrm{E}[(\hat{\mu}_{1,\mathrm{STPPR\text{-}SI}}-\mu_1)(\hat{\mu}_{2,\mathrm{STPPR\text{-}SI}}-\mu_2)]$ の不偏推定量となる．よって，$(i,j)$ 要素が $\hat{\mathrm{E}}[(\hat{\mu}_{i,\mathrm{STPPR\text{-}SI}}-\mu_i)(\hat{\mu}_{j,\mathrm{STPPR\text{-}SI}}-\mu_j)]$ である $p\times p$ 行列を

$$\hat{\mathrm{V}}[\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}}]$$

とすると，

$$\hat{\mathrm{V}}[\hat{\theta}]=\nabla g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})^T\hat{\mathrm{V}}[\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}}]\nabla g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})$$

によって $\mathrm{V}[\hat{\theta}]$ を推定することができる．しかし，この方法では $\nabla g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})$ などの計算を行う必要がある．

そこで，本節では次のようなブートストラップ法を考える．

Step 1.　$b=1,2,\ldots,B$ で以下を繰り返す．
・$l=1,2,\ldots,L$ それぞれにおいて $(m_l-1)$ 個の部分標本 $S^{(b)}_{l,1},S^{(b)}_{l,2},\ldots,S^{(b)}_{l,(m_l-1)}$ を $m_l$ 個の部分標本 $S_{l,1},S_{l,2},\ldots,S_{l,m_l}$ から復元抽出する．
・上のようにして構成したブートストラップ標本から $\hat{\boldsymbol{\mu}}^{(b)}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}}$ を求める．
・$\hat{\theta}^{(b)}=g(\hat{\boldsymbol{\mu}}^{(b)}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})$ を求める．
Step 2.　$\mathrm{V}[\hat{\theta}]$ を

$$\hat{v}^{(B)}_{\hat{\theta}}=\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{(b)}-\hat{\theta}\right)^2$$

で推定．

このとき，

$$\hat{\theta}^{(b)}-\hat{\theta}\approx\nabla g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})^T(\hat{\boldsymbol{\mu}}^{(b)}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}}-\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})$$

より，

$$\hat{v}^{(B)}_{\hat{\theta}}\approx\nabla g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})^T\left\{\frac{1}{B}\sum_{b=1}^{B}(\hat{\boldsymbol{\mu}}^{(b)}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}}-\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})(\hat{\boldsymbol{\mu}}^{(b)}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}}-\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})^T\right\}\nabla g(\hat{\boldsymbol{\mu}}_{\boldsymbol{x},\mathrm{STPPR\text{-}SI}})$$

が成り立つ．

$$\hat{v}^{(B)}_{x_1,x_2,\mathrm{STPPR\text{-}SI}}=\frac{1}{B}\sum_{b=1}^{B}(\hat{\mu}^{(b)}_{x_1,\mathrm{STPPR\text{-}SI}}-\hat{\mu}_{x_1,\mathrm{STPPR\text{-}SI}})(\hat{\mu}^{(b)}_{x_2,\mathrm{STPPR\text{-}SI}}-\hat{\mu}_{x_2,\mathrm{STPPR\text{-}SI}})$$

とすると，

$$\mathrm{E}_{\mathrm{B}}[\hat{v}^{(B)}_{x_1,x_2,\mathrm{STPPR\text{-}SI}}]=\sum_{l=1}^{L}\frac{W_l^2}{m_l(m_l-1)}\sum_{j=1}^{m_l}(\bar{x}_{1,l,j}-\bar{\bar{x}}_{1,l})(\bar{x}_{2,l,j}-\bar{\bar{x}}_{2,l})$$

となり $\hat{v}^{(B)}_{\hat{\theta}}$ を用いることで $\hat{\mathrm{V}}[\hat{\theta}]$ の近似値を得ることができる．

(注意) $\hat{v}^{(B)}_{\hat{\theta}}$ は $\mathrm{V}[\hat{\theta}]$ の不偏推定量とは必ずしもなっていない．

## 5　未回収がある場合の分散

近年の社会調査においては未回収が発生することが一般的であり，そのような場合における推定方法が研究されている (たとえば，Särndal & Ludström (2005)，Bethlehem et al.(2011))．GREG 推定量はそのような場合における推定方法のひとつである．

$\mu_y$ の GREG 推定量は，

$$\hat{\mu}_{y,\text{GREG}} = \hat{\mu}_{y,\text{STPPR-SI}} + (\boldsymbol{\mu}_{\boldsymbol{x}} - \hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}})^T \left(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}\boldsymbol{x}^T,\text{STPPR-SI}}\right)^{-1} \hat{\boldsymbol{\mu}}_{y\boldsymbol{x},\text{STPPR-SI}}$$

で与えられる．ここで，$\hat{\boldsymbol{\mu}}_{\boldsymbol{x}\boldsymbol{x}^T,\text{STPPR-SI}} = \frac{1}{N}\sum_{i\in S} w_i \boldsymbol{x}_i \boldsymbol{x}_i^T$ であり，$\hat{\boldsymbol{\mu}}_{y\boldsymbol{x},\text{STPPR-SI}} = \frac{1}{N}\sum_{i\in S} w_i y_i \boldsymbol{x}_i$ である．ただし，$w_i$ は個人 $i$ に対する重みである．

よって，$\boldsymbol{\mu}_{\boldsymbol{x}}$ が既知のもとで

$$\hat{\mu}_{y,\text{GREG}} = g(\hat{\mu}_{y,\text{STPPR-SI}}, \hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}}, \hat{\boldsymbol{\mu}}_{y\boldsymbol{x},\text{STPPR-SI}}, \hat{\boldsymbol{\mu}}_{\boldsymbol{x}\boldsymbol{x}^T,\text{STPPR-SI}})$$

と書くことができる．$(\hat{\mu}_{y,\text{STPPR-SI}}, \hat{\boldsymbol{\mu}}_{\boldsymbol{x},\text{STPPR-SI}}, \hat{\boldsymbol{\mu}}_{y\boldsymbol{x},\text{STPPR-SI}}, \hat{\boldsymbol{\mu}}_{\boldsymbol{x}\boldsymbol{x}^T,\text{STPPR-SI}})$ の収束先で展開を行うことで前節と同様の議論を行うことができ，$\hat{\mu}_{y,\text{GREG}}$ の分散をブートストラップ法で推定することができる．

## 参考文献

Bethlehem, J., Cobben, F. & Schouten, B. (2011). *Handbook of nonresponse in household surveys*. Wiley.

標本誤差推計研究会 (1998)．標本誤差の推計方法，統計情報研究開発センター．

土屋隆裕 (2009)．標本調査法，朝倉書店．

Särndal, C. E. & Ludström, S. (2005). *Estimation in surveys with nonresponse*. Wiley.

Shao, J. & Tu, D. (1995). *The jacknife and bootstrap*. Springer.

Wolter, K. M. (2007). *Introduction to variance estimation. 2nd ed*. Springer.