Liet. mat. rink. LMD darbai, **50**, 2009, 281–286

Estimation of a total for rotated sample design using auxiliary information

Viktoras CHADYŠAS
Vilnius Gediminas Technical University, Faculty of Fundamental Sciences
Saulėtekio 11, LT-10223 Vilnius, Lithuania
e-mail: viktoras.chadysas@fm.vgtu.lt

Abstract. In this paper we focus on constructions of the total estimator for rotated sampling design. Successive sampling procedure using multi-phase sampling design have been developed. The composite ratio type estimator of the total using auxiliary information and its approximate variance is constructed under simple random sampling design on each phase.

Keywords: total estimator, sample rotation, multi-phase sampling, simple random sampling.

Introduction

Some surveys are continuous surveys using repeated sampling from the finite populations. The theory of successive sampling (sampling on two or more occasions, double sampling or two-phase sampling) has been studied by many authors [2,6]. The purpose has been to realize gains in precision of estimators for periodic surveys.

For example labour force surveys in official statistics are conducted quarterly, a household is interviewed in two subsequent quarters, allowed to rest for the next two, and returns to the sample for another two. Despite information from the previous surveys is available it is not used at the estimation stage at statistics Lithuania yet. The estimation procedure can be combined with multi-phase sampling. The previous-phase sample data can be used as auxiliary information for estimation of the population total in order to reduce variance of the estimator. The effectiveness of ratio estimator in the case of one-phase sampling design and highly correlated study and auxiliary variables is described in [5]. The composite ratio type estimator of the total, of the labour force survey in the case of the simple random sampling with rotation have been considered by Chadyšas and Krapavickaitė [1], but only for a two-occasion sampling scheme.

In the following section, we consider more complex sampling rotation scheme (we are using data for all quarters), than two-occasion sampling scheme (we are using data for two quarters). The composite ratio type estimator of the population total with use of auxiliary information using rotated sampling design is proposing under simple random sampling design on each phase in Section 2.

1. Sampling rotation scheme

Now we briefly, step by step, describe the sampling rotation scheme. Suppose we have a finite population $\mathcal{U} = \{1, \ldots, N\}$ with values y_i, $i = 1, \ldots, N$.

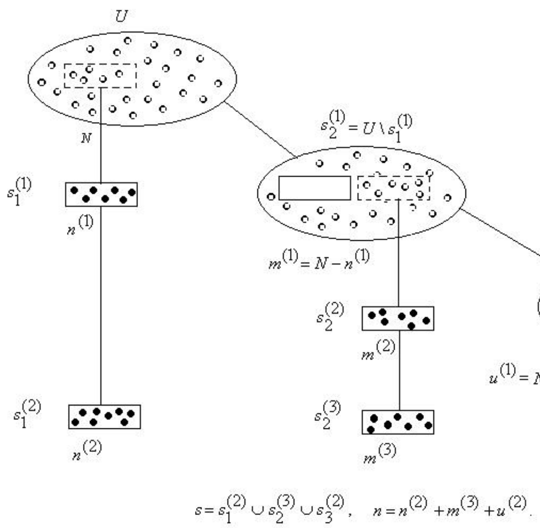

Figure 1. Sampling rotation scheme.

- *Step* 1. Firstly, the sample $s_1^{(1)}$ of size $n^{(1)}$ is drawn from the population $\mathcal{U}$ according to a simple random sampling design.
- *Step* 2. From the set $s_2^{(1)} = \mathcal{U} \setminus s_1^{(1)}$ sample $s_2^{(2)}$ of size $m^{(2)}$ is drawn in accordance with a simple random sampling design.
- *Step* 3a. The sample $s_1^{(2)}$ of size $n^{(2)}$ is drawn from $s_1^{(1)}$ to a simple random sampling design.
- *Step* 3b. The sample $s_2^{(3)}$ of size $m^{(3)}$ is drawn from $s_2^{(2)}$ according to a simple random sampling design.
- *Step* 3c. Finally sample $s_3^{(2)}$ of size $u^{(2)}$ is drawn from the set $s_3^{(1)} = \mathcal{U} \setminus (s_1^{(1)} \cup s_2^{(2)})$ accordance with a simple random sampling design.

Sampling rotation scheme is shown in Fig. 1. It is seen in presented scheme that the whole sample s consists of the union of three samples: $s_1^{(2)}$, $s_2^{(3)}$ and $s_3^{(2)}$. Firstly, we will construct three separate estimators of the total using data of samples $s_1^{(2)}$, $s_2^{(3)}$ and $s_3^{(2)}$ respectively. Secondly we will propose composite ratio type estimator of the total using that sample rotation scheme.

2. Estimator of total

In sample survey, auxiliary information can be used at the estimation stage to increase the accuracy of estimators. Using proposed sampling rotation scheme we can construct estimators of total with values obtained by observing the elements in the previous phases as auxiliary information.

Case 1. The sample $s_1^{(2)}$ is obtained by a two-phase sampling:

$$\mathcal{U} \longrightarrow s_1^{(1)} \longrightarrow s_1^{(2)}.$$

In this two-phase sample design, on the first phase the the simple random sample $s_1^{(1)}$ is drawn from the population $\mathcal{U}$ and on the second-phase a matched simple random sample $s_1^{(2)}$ is drawn from $s_1^{(1)}$. The values of the study variable on the first-phase can be used us auxiliary information. Let us denote the study variable on the first-phase by x with values x_i, $i \in s_1^{(1)}$, and the same variable on the second-phase by y with the values y_i, $i \in s_1^{(2)}$.

Using the first-phase simple random sample $s_1^{(1)}$ and the second-phase simple random sample $s_1^{(2)}$ we form a well known ratio estimator of the total

$$\hat{t}_{1y}^{(2)r} = \hat{t}_{1x}^{(1)} \frac{\hat{t}_{1y}^{(2)}}{\hat{t}_{1x}^{(2)}} = \hat{t}_{1x}^{(1)} \hat{r}^{(2)}, \tag{1}$$

where

$$\hat{t}_{1x}^{(1)} = \frac{N}{n^{(1)}} \sum_{i \in s_1^{(1)}} x_i, \quad \hat{t}_{1y}^{(2)} = \frac{N}{n^{(2)}} \sum_{i \in s_1^{(2)}} y_i,$$

$$\hat{t}_{1x}^{(2)} = \frac{N}{n^{(2)}} \sum_{i \in s_1^{(2)}} x_i, \quad \hat{r}^{(2)} = \frac{\sum_{i \in s_1^{(2)}} x_i}{\sum_{i \in s_1^{(2)}} y_i}.$$

Case 2. The sample $s_2^{(3)}$ is considered as a three-phase sample:

$$\mathcal{U} \longrightarrow s_2^{(1)} = \mathcal{U} \setminus s_1^{(1)} \longrightarrow s_2^{(2)} \longrightarrow s_2^{(3)}.$$

To extend the results to three-phase estimation, assume a third-phase simple random sample $s_2^{(3)}$ of size $m^{(3)}$ is selected from a second-phase sample $s_2^{(2)}$ of size $m^{(2)}$, which is itself a simple random sample of a first-phase sample $s_2^{(1)}$ of size $m^{(1)}$.

For three-phase sampling we can also form an estimator of the population total using values of the study variable on the second-phase as auxiliary information. Let us denote the study variable on the second-phase by x with values x_i, $i \in s_2^{(2)}$, and the same variable on the third-phase by y with the values y_i, $i \in s_2^{(3)}$. Using the second-phase sample $s_2^{(2)}$ and the third-phase sample $s_2^{(3)}$ we can form ratio estimator of the total

$$\hat{t}_{2y}^{(3)r} = \hat{t}_{2x}^{(2)} \frac{\hat{t}_{2y}^{(3)}}{\hat{t}_{2x}^{(3)}} = \hat{t}_{2x}^{(2)} \hat{r}^{(3)}, \tag{2}$$

where

$$\hat{t}_{2x}^{(2)} = \frac{N}{m^{(2)}} \sum_{i \in s_2^{(2)}} x_i, \quad \hat{t}_{2y}^{(3)} = \frac{N}{m^{(3)}} \sum_{i \in s_2^{(3)}} y_i,$$

$$\hat{t}_{2x}^{(3)} = \frac{N}{m^{(3)}} \sum_{i \in s_2^{(3)}} x_i, \quad \hat{r}^{(3)} = \frac{\sum_{i \in s_2^{(3)}} x_i}{\sum_{i \in s_2^{(3)}} y_i}.$$

Case 3. The sample $s_3^{(2)}$ is considered as a two-phase sample:

$$\mathcal{U} \longrightarrow s_3^{(1)} = \mathcal{U} \setminus (s_1^{(1)} \cup s_2^{(2)}) \longrightarrow s_3^{(2)}.$$

In this case we have not auxiliary information for elements in the first-phase sample $s_3^{(1)}$. In two-phase sampling, the population total $t = \sum_{i=1}^{N} y_i$ is estimated unbiasedly by the estimator

$$\hat{t}_{3y}^{(2)} = \frac{N}{u^{(2)}} \sum_{i \in s_3^{(2)}} y_i. \tag{3}$$

By linear combination of estimators (1), (2) and (3) we obtain a new composite ratio type estimator of total of study variable y

$$\hat{t}_y = \alpha \hat{t}_{1y}^{(2)r} + \beta \hat{t}_{2y}^{(3)r} + \gamma \hat{t}_{3y}^{(2)}, \tag{4}$$

where α, β and γ are constants ($0 < \alpha < 1, 0 < \beta < 1, 0 < \gamma < 1$), satisfying $\alpha + \beta + \gamma = 1$.

PROPOSITION 1. *An approximate variance of the composite estimator* (4) *is expressed by*

$$\begin{aligned} \mathrm{AV}(\hat{t}_y) = {} & \alpha^2 \mathrm{AV}\big(\hat{t}_{1y}^{(2)r}\big) + \beta^2 \mathrm{AV}\big(\hat{t}_{2y}^{(3)r}\big) + \gamma^2 \mathrm{V}\big(\hat{t}_{3y}^{(2)}\big) \\ & + 2\alpha\beta \operatorname{cov}\big(\hat{t}_{1y}^{(2)r}, \hat{t}_{2y}^{(3)r}\big) + 2\alpha\gamma \operatorname{cov}\big(\hat{t}_{1y}^{(2)r}, \hat{t}_{3y}^{(2)}\big) + 2\beta\gamma \operatorname{cov}\big(\hat{t}_{2y}^{(3)r}, \hat{t}_{3y}^{(2)}\big). \end{aligned} \tag{5}$$

Here we are using notations:

1. *The approximate variance of the ratio estimator of the total* $\hat{t}_{1y}^{(2)r}$ (1) *in the two-phase sampling design* (see, e.g., [4]) *is*

$$\mathrm{AV}(\hat{t}_{1y}^{(2)r}) = N^2 \left(1 - \frac{n^{(1)}}{N}\right) \frac{s_{1y}^2}{n^{(1)}} + N^{(2)} \left(1 - \frac{n^{(2)}}{n^{(1)}}\right) \frac{s_{1r}^2}{n^{(2)}}, \tag{6}$$

$$s_{1y}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \mu_y)^2, \quad \mu_y = \frac{1}{N} \sum_{i=1}^{N} y_i,$$

$$s_{1r}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - rx_i)^2, \quad r = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i}.$$

2. *The expression for variance of estimator of the population total $t = \sum_{i=1}^{N} y_i$ in the three-phase sample design to a certain sampling design is given in* [3]. *In the case of simple random sampling on each of the phase, an approximate variance of the ratio estimator of the total $\hat{t}_{2y}^{(3)r}$* (2) *in the three-phase sampling design is expressed:*

$$\mathrm{AV}(\hat{t}_{2y}^{(3)r}) = N^2\left(1 - \frac{m^{(2)}}{N}\right)\frac{s_{2y}^2}{m^{(2)}} + N^2\left(1 - \frac{m^{(3)}}{m^{(2)}}\right)\frac{s_{2r}^2}{m^{(3)}}, \tag{7}$$

where $s_{2y}^2 = s_{1y}^2$ and $s_{2r}^2 = s_{1r}^2$ are given in (6).

3. *In two-phase sampling, the variance of the population total $\hat{t}_{3y}^{(2)}$* (3) (*see, e.g.,* [5]) *is*

$$\mathrm{V}(\hat{t}_{3y}^{(2)}) = N^2\left(1 - \frac{u^{(2)}}{N}\right)\frac{s_{3y}^2}{u^{(2)}}, \tag{8}$$

where $s_{3y}^2 = s_{1y}^2$ are given in (6).

Remark 1. We use variance estimator $\widehat{\mathrm{V}}(\hat{t}_y)$ of the composite estimator (4), replacing s_{1y}^2 and s_{1r}^2 in the $\mathrm{AV}(\hat{t}_{1y}^{(2)r})$ of (5) by the estimates below

$$\hat{s}_{1y}^2 = \frac{1}{n^{(2)}-1}\sum_{i\in s_1^{(2)}}\left(y_i - \overline{y}^{(2)}\right)^2, \quad \overline{y}^{(2)} = \frac{1}{n^{(2)}}\sum_{i\in s_1^{(2)}} y_i,$$

and

$$\hat{s}_{1r}^2 = \frac{1}{n^{(2)}-1}\sum_{i\in s_1^{(2)}}\left(y_i - \hat{r}^{(2)}x_i\right)^2,$$

where $\hat{r}^{(2)}$ are given in (1).

In the $\mathrm{AV}(\hat{t}_{2y}^{(3)r})$, s_{2y}^2 and s_{2r}^2 is replaced respectively by

$$\hat{s}_{2y}^2 = \frac{1}{m^{(3)}-1}\sum_{i\in s_2^{(3)}}(y_i - \overline{y}^{(3)})^2, \quad \overline{y}^{(3)} = \frac{1}{m^{(3)}}\sum_{i\in s_2^{(3)}} y_i,$$

$$\hat{s}_{2r}^2 = \frac{1}{m^{(3)}-1}\sum_{i\in s_2^{(3)}}\left(y_i - \hat{r}^{(3)}x_i\right)^2,$$

where $\hat{r}^{(3)}$ are given in (2).

Finally, in the $\mathrm{V}(\hat{t}_{3y}^{(2)})$, s_{3y}^2 is replaced by

$$\hat{s}_{3y}^2 = \frac{1}{u^{(2)} - 1} \sum_{i \in s_3^{(2)}} (y_i - \overline{y})^2, \quad \overline{y} = \frac{1}{u^{(2)}} \sum_{i \in s_3^{(2)}} y_i .$$

3. Conclusions

This paper addresses a practical problem related to the estimation of total in multi-stage sampling with simple random sampling design in each of the phases. Proposed scheme can be used for official surveys. We therefore propose composite ratio type estimator of the total with values obtained by observing the elements in the previous phases as auxiliary information and its approximate variance in the case of the simple random sampling with rotation.

References

1. V. Chadyšas, D. Krapavickaitė. Dirbančių gyventojų skaičiaus vertinimas esant imties rotacijai. *Liet. mat. rink.*, 48/49:288–293, 2008.
2. A.R. Eckler. Rotation sampling. *The Annals of Mathematical Statistics*, 26:664–685, 1955.
3. A.W. Fuller. Estimation for multiple phase samples. In: R.L. Chambers, J.L. Skinner (Eds.), *Analysis of Survey Data*. John Wiley & Sons, Chichester, 2003.
4. D. Krapavickaitė, A. Plikusas. *Imčių teorijos pagrindai*. Technika, Vilnius, 2005.
5. C.E. Särndal, B. Swensson, J. Wretman. *Model Assisted Survey Sampling*. Springer-Verlag, New York, 1992.
6. A.R. Sen. Successive sampling with p auxiliary variables. *The Annals of Mathematical Statistics*, 43(6):2031–2034, 1972.

REZIUMĖ

V. Chadyšas. Sumos vertininimas naudojant papildomą informaciją rotuojamai imčiai

Straipsnyje nagrinėjamas sumos vertinimas esant imties rotacijai. Sumos vertinimui pasiūlyta kelių fazių imties išrinkimo schema. Sukonstruotas sudėtinis santykinis sumos įvertinys naudojantis papildomą informaciją, bei jo apytikslė dispersija, kai kiekvienoje iš fazių renkama paprastoji atsitiktinė imtis.

Raktiniai žodžiai: sumos įvertinys, imties rotacija, kelių fazių ėmimas, paprastasis atsitiktinis ėmimas.