Thailand Statistician
July 2017; 15(2): 105-110
http://statassoc.or.th
Contributed paper

Estimators for Population Mean in Adaptive Cluster Sampling

Faryal Younis*, **and Javid Shabbir**
Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan.
*Corresponding author; e-mail: faryal.younis@stat.qau.edu.pk

Received: 10 December 2016
Accepted: 2 March 2017

Abstract

In this article, we propose modified ratio estimators to estimate finite population mean under adaptive cluster sampling when initial sample is taken by simple random sampling without replacement. The estimators utilize information on known parameters of auxiliary variable. Mean square error equations for proposed estimators are obtained up to first degree of approximation. Proposed estimators are more efficient than estimators proposed by Thompson (1990) and Chutiman (2013) under certain realistic conditions. Simulation study is carried out to support the theoretical findings.

Keywords: adaptive cluster sampling, standard deviation, covariance.

1. Introduction

Adaptive cluster sampling (ACS) introduced by Thompson (1990) is an efficient sampling scheme for rare and clustered populations. ACS is useful for studies of contagious diseases, drug addicts, endangered species of animals, rare plants, minerals and natural resources. ACS begins by taking an initial sample using traditional sampling designs e.g. simple random sampling with or without replacement (Thompson, 1990), stratified sampling (Thompson, 1991) or systematic sampling (Dryver et al., 2012) and then adaptively increasing the sample using information of neighboring units which satisfy the prespecified condition.

Chao (2004) recommended generalized ratio estimator based on modified Horvitz-Thompson estimator under ACS design. Dryver and Chao (2007) suggested ratio estimators based on modified Hansen-Hurwitz estimator and based on Hartley-Ross estimator. Chao et al. (2011) proposed the improved ratio estimators using Rao-Blackwell theorem. Chutiman (2013) introduced some modified ratio estimators along with a regression and difference estimator in ACS. Gattone et al. (2016) suggested product estimator based on both modified Hansen-Hurwitz and Horvitz-Thompson estimator under ACS. The use of auxiliary information helps to enhance efficiency of estimators. The estimators proposed by Chutiman (2013) utilize coefficient of variation and kurtosis of the auxiliary variable. Motivated by Kadilar (2004), we suggest modified ratio estimators in adaptive cluster sampling based on Hansen-Hurwitz estimator. The estimators take advantage of available auxiliary information in terms of standard deviation and covariance, assuming that units of study and auxiliary variable are same. The estimators are aimed to estimate finite population mean under ACS using simple random sampling without replacement for initial sample selection.

2. Adaptive Cluster Sampling

Consider a finite population of N units in which study variable Y takes values $y_1, y_2, ..., y_N$. Our objective is to estimate the unknown population mean $\mu_y = \frac{1}{N}\sum_{i=1}^{N}(y_i)$. Adaptive cluster sampling starts by selecting an initial sample of n units from population by simple random sampling without replacement. All n selected units are examined. If any of these initially selected units satisfy the prespecified condition C then its neighborhoods are included in the sample. Further if any of these adaptively selected units satisfy condition C, then its neighbors are also included in the sample and this process continues until no more units satisfy the condition C. Neighborhood and the condition C are properly defined before a survey is conducted. Set of units that fulfill the condition C around the initially selected unit is known as network.

Thompson (1990) proposed an unbiased estimator for population mean based on Hansen-Hurwitz estimator as

$$(\widehat{\mu}_y)_{HH} = \frac{1}{n}\sum_{i=1}^{n}(w_y)_i \tag{1}$$

where

$$(w_y)_i = \frac{1}{m_i}\sum_{j\in\Psi_i} y_j.$$

Ψ_i is the network that includes unit i, $(w_y)_i$ is the average of y values and m_i is the size of that network, The variance of $(\widehat{\mu}_y)_{HH}$ is

$$V((\widehat{\mu}_y)_{HH}) = \frac{1-f}{n}S^2_{w_y} \tag{2}$$

where

$$S^2_{w_y} = \frac{1}{N-1}\sum_{i=1}^{N}((w_y)_i - \mu_y)^2.$$

Chutiman (2013) suggested the following ratio estimators:

$$\widehat{\mu}_2 = \frac{(\widehat{\mu}_y)_{HH}}{(\widehat{\mu}_x)_{HH} + C_{w_x}}(\mu_x + C_{w_x}) \tag{3}$$

$$\widehat{\mu}_3 = \frac{(\widehat{\mu}_y)_{HH}}{(\widehat{\mu}_x)_{HH} + \beta_{2(w_x)}}(\mu_x + \beta_{2(w_x)}) \tag{4}$$

$$\widehat{\mu}_4 = \frac{(\widehat{\mu}_y)_{HH}}{(\widehat{\mu}_x)_{HH}C_{w_x} + \beta_{2(w_x)}}(\mu_x C_{w_x} + \beta_{2(w_x)}). \tag{5}$$

C_{w_x} is population coefficient of variation and $\beta_{2(w_x)}$ is population coefficient of kurtosis of w_x. The approx. mean square error of above said estimators are

$$MSE(\widehat{\mu}_i) \approx \frac{1-f}{n}(S^2_{w_y} + \phi_i^2 R^2 S^2_{w_x} - 2\phi_i R S_{w_x w_y}) \quad \text{for } i = 2,3,4, \tag{6}$$

where

$$\begin{aligned} S^2_{w_x} &= \frac{1}{N-1}\sum_{i=1}^{N}((w_x)_i - \mu_x)^2, S_{w_x w_y} = \frac{1}{N-1}\sum_{i=1}^{N}((w_x)_i - \mu_x)((w_y)_i - \mu_y) \\ \text{and} \quad \phi_2 &= \frac{\mu_x}{\mu_x + C_{w_x}}, \phi_3 = \frac{\mu_x}{\mu_x + \beta_{2(w_x)}}, \phi_4 = \frac{\mu_x C_{w_x}}{\mu_x C_{w_x} + \beta_{2(w_x)}}. \end{aligned}$$

3. Proposed Estimators

The modified ratio estimators for population mean in ACS are suggested as:

$$\widehat{\mu}_{p1} = \frac{(\widehat{\mu}_y)_{HH} + b_w(\mu_x - (\widehat{\mu}_x)_{HH})}{(\widehat{\mu}_x)_{HH} + S_{w_x}}[\mu_x + S_{w_x}] \tag{7}$$

$$\widehat{\mu}_{p2} = \frac{(\widehat{\mu}_y)_{HH} + b_w(\mu_x - (\widehat{\mu}_x)_{HH})}{(\widehat{\mu}_x)_{HH} + \sqrt{S_{w_x w_y}}}\left[\mu_x + \sqrt{S_{w_x w_y}}\right] \tag{8}$$

where

$$b_w = \frac{s_{w_x w_y}}{s^2_{w_x}}, s_{w_x w_y} = \frac{1}{n-1}\sum_{i=1}^{n}((w_x)_i - \overline{w}_x)((w_y)_i - \overline{w}_y), s^2_{w_x} = \frac{1}{n-1}\sum_{i=1}^{n}((w_x)_i - \overline{w}_x)^2,$$

S_{w_x} is population standard deviation of w_x and $S_{w_x w_y}$ is population covariance between w_x and w_y. To find mean square error of proposed estimators, we define the following notations:

$$\begin{aligned} (\widehat{\mu}_y)_{HH} &= \mu_y(1+\overline{e}_{w_y}), (\widehat{\mu}_x)_{HH} = \mu_x(1+\overline{e}_{w_x}), E(\overline{e}_{w_y}) = E(\overline{e}_{w_x}) = 0, \\ E(\overline{e}^2_{w_y}) &= \frac{1-f}{n}(\frac{S^2_{wy}}{u^2_y}), E(\overline{e}^2_{w_x}) = \frac{1-f}{n}(\frac{S^2_{wx}}{u^2_x}), E(\overline{e}_{w_y}\overline{e}_{wx}) = \frac{1-f}{n}(\frac{S_{w_y w_x}}{u_y u_x}) \end{aligned}$$

Using the above notations and considering first order approximation, MSE of proposed estimator $\widehat{\mu}_{p1}$ is derived as:

$$\widehat{\mu}_{p1} = (\mu_y(1+\overline{e}_{w_y}) + b_w(\mu_x - \mu_x(1+\overline{e}_{w_x})))\frac{\mu_x + S_{w_x}}{\mu_x(1+\overline{e}_{w_x}) + S_{w_x}} \tag{9}$$

$$\begin{aligned} \widehat{\mu}_{p1} &= (\mu_y + \mu_y\overline{e}_{w_y} - b_w\mu_x\overline{e}_{w_x})(1 - \theta_1\overline{e}_{w_x} + \theta_1^2\overline{e}^2_{w_x}) \\ E(\widehat{\mu}_{p1} - \mu_y)^2 &= E(\mu_y\overline{e}_{w_y} - (\mu_y\theta_1 + b_w\mu_x)\overline{e}_{w_x})^2 \end{aligned}$$

$$MSE(\widehat{\mu}_{p1}) = \frac{1-f}{n}(S^2_{wy}(1-\rho^2_{w_x w_y}) + \theta_1^2 R^2 S^2_{wx}). \tag{10}$$

Similarly MSE for $\widehat{\mu}_{p2}$ is given by

$$MSE(\widehat{\mu}_{p2}) = \frac{1-f}{n}(S^2_{wy}(1-\rho^2_{w_x w_y}) + \theta_2^2 R^2 S^2_{wx}) \tag{11}$$

where

$$\theta_1 = \frac{\mu_x}{\mu_x + S_{w_x}}, \theta_2 = \frac{\mu_x}{\mu_x + \sqrt{S_{w_x w_y}}}$$

4. Efficiency Comparison

The proposed modified ratio estimators are theoretically compared in terms of efficiency, with

- Estimator by Thompson (1990) ($(\widehat{\mu}_y)_{HH}$)
- Estimator by Chutiman (2013) ($\widehat{\mu}_2, \widehat{\mu}_3, \widehat{\mu}_4$)

The proposed estimator $\widehat{\mu}_{pi}$ is efficient than $(\widehat{\mu}_y)_{HH}$, when $MSE(\widehat{\mu}_{pi}) < MSE((\widehat{\mu}_y)_{HH})$ i.e.

$$\frac{1-f}{n}(S^2_{w_y}(1-\rho^2_{w_xw_y}) + \theta_i^2 R^2 S^2_{w_x}) < \frac{1-f}{n} S^2_{w_y}$$

which implies

$$\rho^2_{w_xw_y} - \frac{\theta_i^2 R^2 S^2_{w_x}}{S^2_{w_y}} > 0, i = 1, 2. \quad (12)$$

The proposed estimator $\widehat{\mu}_{pi}$ is efficient than $(\widehat{\mu}_j)$, when $MSE(\widehat{\mu}_{pi}) < MSE(\widehat{\mu}_j)$ i.e.

$$\frac{1-f}{n}(S^2_{w_y}(1-\rho^2_{w_xw_y}) + \theta_i^2 R^2 S^2_{w_x}) < \frac{1-f}{n}(S^2_{w_y} + \phi_j^2 R^2 S^2_{w_x} - 2\phi_j R S_{w_xw_y})$$

which implies

$$S^2_{w_y}\rho^2_{w_xw_y} - R^2 S^2_{w_x}(\theta_i^2 - \phi_j^2) - 2\phi_j R S_{w_xw_y} > 0 \text{ for } i = 1, 2, j = 2, 3, 4. \quad (13)$$

5. Simulation Study

Simulation study is conducted to support the theoretical findings. For this purpose, simulation y-values and x-values are studied from Chutiman (2013). Correlation between study and auxiliary variable at network level is positive. All the efficiency conditions obtained in Section 4 are satisfied so we expect the proposed estimators $\widehat{\mu}_{pi}$ to perform better than $(\widehat{\mu}_y)_{HH}$ and $\widehat{\mu}_j$ for $j = 2, 3, 4$.

Simple random sampling without replacement is used in each iteration to have an initial sample of size n. y-values are used to obtain a network. Condition $C = \{y : y \geq 1\}$ is used to adaptively add units in the sample. Neighborhood is defined as four adjacent units i.e. left, right, top and bottom. Initial sample size $n = 5, 10, 20, 30, 40$ are used for comparison. Five thousand iterations are performed for each estimator to obtain accurate estimates. The following expressions are used for calculation of mean square error and percentage relative efficiency for different estimators.

$$MSE(\widehat{\mu}^*) = \frac{1}{5000}\sum_{i=1}^{5000}(\widehat{\mu}_i^* - \mu_y)^2$$

and

$$PRE(\widehat{\mu}^*) = \frac{MSE((\widehat{\mu}_y)_{HH})}{MSE(\widehat{\mu}^*)} \times 100$$

where $\widehat{\mu}^* = (\widehat{\mu}_y)_{HH}, \widehat{\mu}_j$ for $j = 2, 3, 4$ and $\widehat{\mu}_{pj}$ for $j = 1, 2$. Table 1 presents simulation results for mean square error and Table 2 presents percentage relative efficiencies calculated for various existing estimators and proposed estimators.

Table 1 Mean square error of estimators

n	$(\widehat{\mu}_y)_{HH}$	$\widehat{\mu}_2$	$\widehat{\mu}_3$	$\widehat{\mu}_4$	$\widehat{\mu}_{p1}$	$\widehat{\mu}_{p2}$
5	2.213817	1.206003	2.208236	2.019381	0.15055	0.13084
10	1.120247	0.715977	1.136250	1.034834	0.11911	0.10242
20	0.555398	0.384914	0.555427	0.533658	0.09354	0.08075
30	0.352095	0.264994	0.352565	0.330627	0.08011	0.06905
40	0.262028	0.188395	0.258996	0.245047	0.07470	0.06327

Table 2 Percentage relative efficiency

n	$(\widehat{\mu}_y)_{HH}$	$\widehat{\mu}_2$	$\widehat{\mu}_3$	$\widehat{\mu}_4$	$\widehat{\mu}_{p1}$	$\widehat{\mu}_{p2}$
5	100	183.56	100.25	109.62	1470.48	1692.00
10	100	156.46	98.59	108.25	940.51	1093.78
20	100	144.29	99.99	104.07	593.75	687.79
30	100	132.86	99.86	106.49	439.51	509.91
40	100	139.08	101.12	106.92	350.77	414.14

6. Conclusion

We have developed modified ratio type estimators for the population mean when population under study is rare and clustered. Auxiliary information in terms of standard deviation and covariance is used to increase precision of the estimators. The mean square error of proposed estimators are obtained up to first degree of approximation and compared with existing estimators. All the efficiency conditions are satisfied for population under study. On the basis of results presented in Table 1 and 2, we conclude that proposed ratio estimators are preferable. It is also noticed that proposed estimator $\widehat{\mu}_{p2}$ has least mean square error and highest percentage relative efficiency among all estimators.

Acknowledgement

The authors are thankful to the editor in chief and anonymous referees for their valuable suggestions, which helped to improve the article. The first author is funded by Higher Education Commission, Pakistan.

References

Chao CT. Ratio estimation on adaptive cluster sampling. J. Chinese Stat. Assoc. 2004; 42: 307-327.

Chao CT. Dryver AL. Chiang TZ. Leveraging the Rao-Blackwell theorem to improve ratio estimators in adaptive cluster sampling. Environ. Ecol. Stat. 2011; 18: 543-568.

Chutiman N. Adaptive cluster sampling using auxiliary variable. J. Math. Stat. 2013; 9: 249-255.

Dryver AL. Chao CT. Ratio estimators in adaptive cluster sampling. Environmetrics. 2007; 18: 607-620.

Dryver AL. Nethran U. Smith DR. Partial systematic adaptive cluster sampling. Environmetrics. 2012; 23: 306-316.

Gattone SA. Mohamed E. Dryver AL. Munnich RT. Adaptive cluster sampling for negatively correlated data. Environmetrics. 2016; 27: E103-E113.

Kadilar C. Cingi H. Ratio estimator in simple random sampling. Appl. Math. Comput. 2004; 151: 893-902.

Thompson SK. Adaptive cluster sampling. J. Am. Stat. Assoc. 1990; 85: 1050-1059.

Thompson SK. Stratified adaptive cluster sampling. Biometrika. 1991; 78: 389-397.