



Thailand Statistician
July 2010; 8(2) : 223-233
<http://statassoc.or.th>
Contributed paper

A New Ratio Estimator in Stratified Adaptive Cluster Sampling

Nipaporn Chutiman

Department of mathematics, Faculty of Science, Mahasarakham University,
Mahasarakham 44150, Thailand.
E-mail: nipaporn.c@msu.ac.th

Received: 28 April 2010

Accepted: 30 June 2010

Abstract

Kadilar and Cingi considered the ratio estimators for the population mean of the variable of interest in stratified random sampling. In this paper we focus on the new ratio estimator (based on Kadilar and Cingi in stratified adaptive cluster sampling. Simulations showed the proposed estimator had the smallest estimated mean square error when compared to the ratio estimators in stratified sampling and the ratio estimator, based on Hansen-Hurwitz estimator, in stratified adaptive cluster sampling.

Keywords: ratio estimator, stratified adaptive cluster sampling.

1. Introduction

Stratified adaptive cluster sampling, proposed by Thompson [1], is an efficient method for sampling rare and hidden clustered populations. In each stratum of stratified adaptive cluster sampling, an initial sample of units is selected by simple random sampling. If the value of the variable of interest from a sampled unit satisfies a pre-specified condition C then the unit's neighborhood will also be added to the sample. If any other units that are "adaptively" added also satisfy the condition C , then their neighborhoods are also added to the sample. This process is continued until no more units that satisfy the condition are found. The set of all units selected and all neighboring units that satisfy the condition is called a network. The adaptive sample units, which do

not satisfy the condition, are called edge units. A network and its associated edge units are called a cluster.

In some situations the researcher obtains observations from more than one variable, from the variable of interest and the auxiliary variable. Use of an auxiliary variable is a common method to improve the precision of estimates of a population parameter. Kadilar and Cingi [2,3] developed ratio estimators in stratified random sampling. In this paper, we will study the ratio estimator in stratified adaptive cluster sampling [3]. Some comparisons are made using a simulation.

2. Ratio Estimator in Stratified Radom Sampling

Let y_{hi} be the variable of interest associated with the i th unit of stratum h and x_{hi} be the auxiliary variable associated with the i th unit of stratum h . Let N_h represent the number of units in stratum h and n_h represent the number of units in the

sample from that stratum. The total of units in the population is $N = \sum_{h=1}^L N_h$, L is the

total number of strata and the total of units in the sample is $n = \sum_{h=1}^L n_h$. The population

mean of y -values in stratum h is $\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$ and the population mean of x -

values in stratum h is $\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$. The population mean of y -values is

$\bar{Y} = \frac{1}{N} \sum_{h=1}^L N_h \bar{Y}_h$ and the population mean of x -values is $\bar{X} = \frac{1}{N} \sum_{h=1}^L N_h \bar{X}_h$.

The combined ratio estimator is

$$\bar{y}_R = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X} = \hat{R} \bar{X}, \tag{1}$$

where $\bar{y}_{st} = \sum_{h=1}^L \frac{N_h}{N} \bar{y}_h$; $\bar{x}_{st} = \sum_{h=1}^L \frac{N_h}{N} \bar{x}_h$, \bar{y}_h is the sample mean of the variable of interest in stratum h , \bar{x}_h is the sample mean of the auxiliary variable in stratum h and

$\hat{R} = \frac{\bar{y}_{st}}{\bar{x}_{st}}$ is the sample ratio of means.

The variance of \bar{y}_R is

$$V(\bar{y}_R) = \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh}^2 - 2RS_{xyh} + R^2S_{xh}^2), \tag{2}$$

where $\gamma_h = \frac{1-(n_h/N_h)}{n_h}$, $R = \frac{\bar{Y}}{\bar{X}}$ is the population ratio of means, S_{yh}^2 is the population variance of the variable of interest in stratum h , S_{xh}^2 is the population variance of the auxiliary variable in stratum h and S_{xyh} is the population covariance between auxiliary variable and variable of interest in stratum h [4,5].

Kadilar and Cingi [3] proposed a ratio estimator [6] as

$$\bar{y}_{Rp} = \kappa \bar{y}_R. \tag{3}$$

where κ minimizes the mean square error of \bar{y}_{Rp} .

The mean square error of \bar{y}_{Rp} is

$$MSE(\bar{y}_{Rp}) = \kappa^2 \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh}^2 - 2RS_{xyh} + R^2S_{xh}^2) + (\kappa - 1)^2 \bar{Y}^2. \tag{4}$$

The bias of \bar{y}_{Rp} is

$$B(\bar{y}_{Rp}) = (\kappa - 1)\bar{Y} + \frac{1}{\bar{X}} \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (RS_{xh}^2 - \kappa S_{xyh}). \tag{5}$$

The κ which minimizes the MSE is the solution to:

$$\frac{\partial MSE(\bar{y}_{Rp})}{\partial \kappa} = 2\kappa \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh}^2 - 2RS_{xyh} + R^2S_{xh}^2) + 2(\kappa - 1)\bar{Y}^2 = 0.$$

From this equation, we obtain

$$\kappa = \frac{\bar{Y}^2}{\bar{Y}^2 + \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh}^2 - 2RS_{xyh} + R^2S_{xh}^2)}, \tag{6}$$

where $0 < \kappa < 1$.

3. Ratio Estimator in Stratified Adaptive Cluster Sampling

For stratified adaptive cluster sampling, the population consists of N units partitioned into L strata based on prior information about units that are similar, and it is assumed that the population ignores crossover of networks between strata.

The ratio estimator in stratified adaptive cluster sampling based on the Hansen-Hurwitz estimator is

$$\bar{y}_{R_sac} = \frac{\bar{y}_{st_sac}}{\bar{x}_{st_sac}} \bar{X} = \hat{R}_{st_sac} \bar{X}, \tag{7}$$

where $\bar{y}_{st_sac} = \sum_{h=1}^L \frac{N_h}{N} \bar{w}_{yh}$, $\bar{w}_{yh} = \frac{\sum_{i=1}^{n_h} w_{yhi}}{n_h}$ where n_h is the number of sample in

stratum h , w_{yhi} is the average of the variable of interest in the network that includes

unit i of the initial sample in stratum h , that is, $w_{yhi} = \frac{y_{hi}}{m_{hi}}$. $\bar{x}_{st_sac} = \sum_{h=1}^L \frac{N_h}{N} \bar{w}_{xh}$,

$\bar{w}_{xh} = \frac{\sum_{i=1}^{n_h} w_{xhi}}{n_h}$ where w_{xhi} is the average of the auxiliary variable in the network that

includes unit i of the initial sample in stratum h , that is, $w_{xhi} = \frac{x_{hi}}{m_{hi}}$ and m_{hi} is the

number of units in that network.

The variance of \bar{y}_{R_sac} is

$$V(\bar{y}_{R_sac}) = \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h \left(\frac{\sum_{i=1}^{n_h} (w_{yhi} - R w_{xhi})^2}{N_h - 1} \right).$$

The alternative form of $V(\bar{y}_{R_sac})$ can be written as

$$V(\bar{y}_{R_sac}) = \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h \left[S_{yh_sac}^2 + R^2 S_{xh_sac}^2 - 2RS_{xyh_sac} \right], \tag{8}$$

where $S_{ih_sac}^2$, $S_{jh_sac}^2$, S_{ijh_sac} ($i, j = x, y$ and $i \neq j$) be the variances and the

covariance term in stratum h , that is, $S_{xh_sac}^2 = \sum_{i=1}^{N_h} (w_{xhi} - \bar{X}_h)^2 / (N_h - 1)$,

$$S_{yh_sac}^2 = \sum_{i=1}^{N_h} (w_{yhi} - \bar{Y}_h)^2 / (N_h - 1) \text{ and}$$

$$S_{xyh_sac} = \sum_{i=1}^{N_h} (w_{xhi} - \bar{X}_h)(w_{yhi} - \bar{Y}_h) / (N_h - 1).$$

The bias of \bar{y}_{R_sac} is

$$B(\bar{y}_{R_sac}) = \frac{1}{X} \left\{ R \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h S_{xh_sac}^2 - \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h S_{xyh_sac} \right\}. \tag{9}$$

4. The Proposed estimator

The new ratio estimator in stratified adaptive cluster sampling [3] is

$$\bar{y}_{Rp_sac} = \kappa^* \bar{y}_{R_sac} \tag{10}$$

The MSE of this estimator is

$$\begin{aligned} MSE(\bar{y}_{Rp_sac}) &= E(\bar{y}_{Rp_sac} - \bar{Y})^2 \\ &= E(\kappa^* \bar{y}_{R_sac} - \bar{Y})^2 \\ &= E(\kappa^{*2} \bar{y}_{R_sac}^2 - 2\bar{Y} \kappa^* \bar{y}_{R_sac} + \bar{Y}^2) \\ &= \kappa^{*2} E(\bar{y}_{R_sac}^2) - 2\bar{Y} \kappa^* E(\bar{y}_{R_sac}) + \bar{Y}^2 \\ &\cong \kappa^{*2} E(\bar{y}_{R_sac}^2) - 2\kappa^* \bar{Y}^2 + \bar{Y}^2 + \kappa^{*2} \bar{Y}^2 - \kappa^{*2} [E(\bar{y}_{R_sac})]^2 \\ &= \kappa^{*2} [E(\bar{y}_{R_sac}^2) - [E(\bar{y}_{R_sac})]^2] + \bar{Y}^2 (\kappa^* - 1)^2 \end{aligned}$$

$$\begin{aligned}
 MSE(\bar{y}_{Rp_sac}) &= \kappa^{*2}V(\bar{y}_{R_sac}) + \bar{Y}^2(\kappa^* - 1)^2 \\
 &\cong \kappa^{*2} \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh_sac}^2 + R^2 S_{xh_sac}^2 - 2RS_{xyh_sac}) + \bar{Y}^2(\kappa^* - 1)^2. \quad (11)
 \end{aligned}$$

To find κ^* which minimizes $MSE(\bar{y}_{Rp_sac})$ take derivative of $MSE(\bar{y}_{Rp_sac})$ with respect to κ^* and set it equal to zero.

$$\begin{aligned}
 \frac{\partial}{\partial \kappa^*} MSE(\bar{y}_{Rp_sac}) &= 2\kappa^* \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh_sac}^2 + R^2 S_{xh_sac}^2 - 2RS_{xyh_sac}) + 2\bar{Y}^2(\kappa^* - 1) \\
 &= 0
 \end{aligned}$$

$$\kappa^* \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh_sac}^2 + R^2 S_{xh_sac}^2 - 2RS_{xyh_sac}) + \bar{Y}^2 \kappa^* - \bar{Y}^2 = 0$$

$$\therefore \kappa^* = \frac{\bar{Y}^2}{\bar{Y}^2 + \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (S_{yh_sac}^2 + R^2 S_{xh_sac}^2 - 2RS_{xyh_sac})} ; \quad 0 < \kappa^* < 1 \quad (12)$$

The bias of \bar{y}_{Rp_sac} is

$$\begin{aligned}
 E(\bar{y}_{Rp_sac} - \bar{Y}) &= E(\kappa^* \bar{y}_{R_sac} - \bar{Y}) \\
 &= E\left(\kappa^* \frac{\bar{y}_{st_sac}}{\bar{x}_{st_sac}} \cdot \bar{X} - \bar{Y}\right) \\
 &= \bar{X} E\left(\frac{(\kappa^* \bar{y}_{st_sac} - R\bar{x}_{st_sac})}{\bar{x}_{st_sac}}\right)
 \end{aligned}$$

$$\begin{aligned}
 & \cong E \left[\left(1 - \frac{(\bar{x}_{st_sac} - \bar{X})}{\bar{X}} \right) (\kappa^* \bar{y}_{st_sac} - R\bar{x}_{st_sac}) \right] \\
 & = E(\kappa^* \bar{y}_{st_sac} - R\bar{x}_{st_sac}) - E \left[\frac{\kappa^* \bar{y}_{st_sac} (\bar{x}_{st_sac} - \bar{X})}{\bar{X}} \right] + RE \left[\frac{\bar{x}_{st_sac} (\bar{x}_{st_sac} - \bar{X})}{\bar{X}} \right] \\
 & = \kappa^* \bar{Y} - \bar{Y} - \frac{\kappa^*}{\bar{X}} E[(\bar{y}_{st_sac} - \bar{Y})(\bar{x}_{st_sac} - \bar{X})] + \frac{R}{\bar{X}} E[(\bar{x}_{st_sac} - \bar{X})(\bar{x}_{st_sac} - \bar{X})] \\
 & = (\kappa^* - 1)\bar{Y} + \frac{R}{\bar{X}} \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h S_{xh_sac}^2 - \frac{\kappa^*}{\bar{X}} \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h S_{xyh_sac} \\
 & B(\bar{y}_{Rp_sac}) = (\kappa^* - 1)\bar{Y} + \frac{1}{\bar{X}} \sum_{h=1}^L \frac{N_h^2}{N^2} \gamma_h (RS_{xh_sac}^2 - \kappa^* S_{xyh_sac}) \tag{13}
 \end{aligned}$$

5. Simulation Study

In this section, the simulated x -values and y -values from [7] were studied. The simulation data from $20 \times 20 = 400$ units was divided into 4 strata of equal size. Each stratum has $20 \times 5 = 100$ units formed from 20 rows and 5 columns. The total of y -values is 254; the mean of y -values (\bar{Y}) is 0.635 and the total of x -values is 196; the mean of the x -values (\bar{X}) is 0.490, and the population ratio (R) is 1.2959. The populations were shown in figure 1-2.

0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	4	6	0	1	0	0	0	4	3	4	2	3
0	0	1	0	0	0	0	4	6	11	4	3	1	0	0	2	6	8	3	7
0	0	0	0	0	0	1	4	4	9	11	4	0	0	2	1	8	10	0	2
0	0	0	0	0	0	1	2	7	4	3	1	1	0	0	0	2	2	0	0

Figure 1. X values

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	6	0	1
0	0	1	0	0	0	0	0	6	18	3	0	1	0	0	0	9	27	9	0
1	0	0	0	0	0	0	1	14	46	11	0	0	0	0	1	8	25	14	1
0	0	0	0	0	0	0	0	4	7	2	0	0	0	0	0	1	2	1	0

Figure 2. Y values

Table 1. Data Statistics

	Stratum			
	1	2	3	4
\bar{X}_h	0.1	0.7	0.37	0.79
\bar{Y}_h	0.11	1.05	0.21	1.17
S_{xh}^2	0.0909	3.7273	1.7506	3.4807
S_{yh}^2	0.1191	26.6338	1.3595	17.1324
S_{xyh}	0.0293	7.0758	1.3761	5.7027
$S_{xh_sac}^2$	0.0909	3.2588	1.3466	2.1847
$S_{yh_sac}^2$	0.1191	10.8472	0.8679	7.3049
S_{xyh_sac}	0.0293	5.5442	0.9417	3.1650

For each iteration, an initial sample of units in each stratum is selected by simple random sampling without replacement. A neighborhood of a unit is defined as the four spatially adjacent units, that is to the left, right, top and bottom (north, south, east and west) of that unit. For each estimator 10,000 iterations were performed to obtain an accuracy estimate. The condition for added units in the sample is defined by $C = \{y : y > 0\}$. Initial SRS sizes in each stratum were varied $n_h = 2, 5, 10, 15, 20, 25, 30, 40$ and 50 were used, then the initial sample sizes from all stratum (n) were $8, 20, 40, 60, 80, 100, 120, 160,$ and 200 . The estimated mean square error of the estimate total is

$$M\hat{S}E(\bar{y}) = \frac{1}{10,000} \sum_{i=1}^{10,000} (\bar{y}_i - \bar{Y})^2,$$

where \bar{y}_i is the value for the relevant estimator for sample i . Let v denote the final sample size in stratified adaptive cluster sampling.

The results of the ratio estimator in stratified sampling ($\bar{y}_R = \hat{R}\bar{X}$), the ratio estimator based on Prasad [6] ($\bar{y}_{Rp} = \kappa\bar{y}_R$), The ratio estimator in stratified adaptive cluster sampling ($\bar{y}_{R_sac} = \hat{R}_{st_sac}\bar{X}$) and The new ratio estimator in stratified adaptive cluster sampling based on Kadilar and Cingi [3] ($\bar{y}_{Rp_sac} = \kappa^*\bar{y}_{R_sac}$) are as follows:

Table 2. The biased of the estimators for the population mean of the variable of the interest.

n_h	n	$E(v)$	$B(\bar{y}_R)$	$B(\bar{y}_{Rp})$	$B(\bar{y}_{R_sac})$	$B(\bar{y}_{Rp_sac})$
2	8	17.81	-0.2301	-0.4903	-0.1924	-0.3104
5	20	42.19	-0.0403	-0.2847	-0.0565	-0.0416
10	40	74.81	-0.0160	-0.1697	0.1245	0.0694
15	60	104.26	-0.0214	-0.1270	0.0940	0.0597
20	80	129.48	-0.0064	-0.0868	0.0889	0.0645
25	100	153.04	-0.0023	-0.0609	0.0820	0.0637
30	120	174.73	-0.0083	-0.0577	0.0777	0.0635
40	160	215.14	-0.0094	-0.0420	0.0743	0.0651
50	200	252.32	-0.0094	-0.0351	0.0717	0.0655

Table 3. The estimated MSE of the estimators for the population mean of the variable of the interest.

n_h	n	$E(v)$	$M\hat{S}E(\bar{y}_R)$	$M\hat{S}E(\bar{y}_{Rp})$	$M\hat{S}E(\bar{y}_{R_sac})$	$M\hat{S}E(\bar{y}_{Rp_sac})$
2	8	17.81	0.48195	0.29209	0.19604	0.18081
5	20	42.19	0.35952	0.20948	0.11854	0.08774
10	40	74.81	0.13436	0.10850	0.07323	0.05529
15	60	104.26	0.09875	0.08208	0.02904	0.02199
20	80	129.48	0.06272	0.05576	0.01828	0.01400
25	100	153.04	0.04437	0.04086	0.01356	0.01059
30	120	174.73	0.03632	0.03404	0.01122	0.00905
40	160	215.14	0.02195	0.02099	0.00803	0.00675
50	200	252.32	0.01416	0.01386	0.00675	0.00592

6. Conclusions

Stratified adaptive cluster sampling is an efficient method for sampling rare and hidden clustered populations. The numerical study showed that the bias for the ratio estimators in stratified sampling are less than the bias for the ratio estimators in stratified

adaptive cluster sampling. the estimated mean square error for the ratio estimators in stratified adaptive cluster sampling $\left(M\hat{S}E\left(\bar{y}_{R_sac}\right)\right)$ and $M\hat{S}E\left(\bar{y}_{Rp_sac}\right)$ are less than the estimated mean square error for the ratio estimators in stratified sampling $\left(M\hat{S}E\left(\bar{y}_R\right)\right)$ and $M\hat{S}E\left(\bar{y}_{Rp}\right)$. Moreover the estimated mean square error of the proposed estimator , \bar{y}_{Rp_sac} , is less than the estimated mean square error of the ratio estimator in stratified adaptive cluster sampling based on Hansen-Hurwitz estimator, \bar{y}_{R_sac} .

7. Acknowledgements

This research was supported by Research Institute of Mahasarakham University. We would also like to profoundly thanks Mr. Paveen Chutiman for his programming advice.

References

- [1] Thompson, S.K., Stratified Adaptive Cluster Sampling. *Biometrika*. 1991;78: 389-397.
- [2] Kardilar, C. and Cingi, H., Ratio Estimators in Stratified Random Sampling. *Biometrical Journal*. 2003; 45:218-225.
- [3] Kardilar, C. and Cingi, H., A New Ratio Estimator in Stratified Random Sampling. *Communications in Statistics-Theory and Methods*. 2005; 34:597-602.
- [4] Cochran, W.G., *Sampling Techniques*: New York: John Wiley and Sons, 1977.
- [5] Thompson, S.K., *Sampling*. New York : Wiley, 2002.
- [6] Prasad, B., Some Improved Ratio Type Estimators of Population Mean and Ratio in Finite Population. *Sample Surveys*. 1989; 18: 379-392.
- [7] Chao, C., Ratio Estimation on Adaptive Cluster Sampling. *Journal of the Chinese Statistical Association*. 2004; 42:307-327.