

# Improved Ratio Estimators of Population Mean In Adaptive Cluster Sampling

Subhash Kumar Yadav<sup>1</sup>, Sheela Misra<sup>2</sup>, Sant Saran Mishra<sup>1,\*</sup> and Nipaporn Chutiman<sup>3</sup>

<sup>1</sup> Department of Mathematics and Statistics (A Centre of Excellence on Advanced Computing)Dr. RML Avadh University, Faizabad-224001, U.P., India

<sup>2</sup> Department of Statistics, University of Lucknow, Lucknow-226007, U.P., India

<sup>3</sup> Department of Mathematics, Faculty of Science, Mahasarakham University, Maha Sarakham-44150, Thailand

Received: 13 Aug. 2015, Revised: 18 Sep. 2015, Accepted: 3 Oct. 2015

Published online: 1 Jan. 2016

**Abstract:** In this paper, we study the estimators of the population mean in adaptive cluster sampling by using the information of the auxiliary variable, that is, the population coefficient of variation, the coefficient of skewness kurtosis of the auxiliary variable and the correlation coefficient between main variable and the auxiliary variable. The large sample properties of the proposed estimators have been studied up to the first order of approximation. A numerical study is also carried out to judge the theoretical findings. The numerical example showed that if the population is rare and hidden clustered population, all estimators in adaptive cluster sampling are more efficient than the estimators in simple random sampling with the same condition.

**Keywords:** Adaptive Cluster Sampling, Auxiliary Variable, Mean squared error, Efficiency

## 1 Introduction

Adaptive cluster sampling, proposed by Thompson(1990), is an efficient method for sampling rare and hidden clustered populations. In 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  $C$ , that is  $(i, y_i > c)$  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. If a unit is selected in the initial sample and does not satisfy the condition  $C$ , then there is only one unit in the network.

It is well known that the variable about which we have full information is known as auxiliary variable and the information is known as auxiliary information which is highly (positively or negatively) correlated with the variable under study. Whenever auxiliary variable (information) is known, one would like to use it at the design or estimation stage since it is well known and established that the use of auxiliary information in sampling theory enhances the efficiency of the estimators and it is in use since the use of sampling itself. In this paper, we will study the estimator of population mean in adaptive cluster sampling using an auxiliary variable.

## 2 Estimators Under Simple Random Sampling

Let  $(x_i, y_i), i = 1, 2, \dots, n$  be the  $n$  pair of observations for the auxiliary and study variables, respectively for the population of size  $N$  using Simple Random Sampling With Out Replacement (*SRSWOR*). Let  $\mu_x$  and  $\mu_y$  be the

\* Corresponding author e-mail: [sant\\_x2003@yahoo.co.in](mailto:sant_x2003@yahoo.co.in)

population means of auxiliary and study variables respectively and  $\bar{x}$  and  $\bar{y}$  be the respective sample means. Ratio estimators are used when the line of regression of  $y$  on  $x$  passes through origin and the variables  $X$  and  $Y$  are positively correlated to each other, while product estimators are used when  $X$  and  $Y$  are negatively correlated to each other, otherwise regression estimators are used to estimate the population parameters under consideration.

Cochran (1940) proposed the classical ratio estimator for estimating the population mean,  $\mu_y = \frac{1}{N} \sum_{i=1}^N y_i$  of the study variable as follows:

$$\bar{y}_R = \bar{y} \left( \frac{\mu_x}{\bar{x}} \right) \quad (1)$$

where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  and  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ , assuming that  $\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$ , the population mean of the auxiliary variable is known.

The expressions for the Mean Squared Error (MSE) of the estimator given in (1) up to the first order of approximation are respectively, as follows:

$$MSE(\bar{y}_R) = \frac{(1-f)}{n} \mu_y^2 [C_y^2 + C_x^2 - 2\rho C_y C_x] \quad (2)$$

where  $C_y = \frac{S_y}{\mu_y}$ ,  $C_x = \frac{S_x}{\mu_x}$ ,  $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu_y)^2$ ,  $f = \frac{n}{N}$ ,  $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2$ ,  $\rho_{xy} = \frac{S_{xy}}{S_y S_x}$ ,

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

Sisodia and Dwivedi (1981) proposed the following estimator using the coefficient of variation of auxiliary variable as:

$$\bar{y}_{R1} = \bar{y} \left( \frac{\mu_x + C_x}{\bar{x} + C_x} \right) \quad (3)$$

The MSE of the estimator  $\bar{y}_{R1}$ , up to the first order of approximation respectively are:

$$MSE(\bar{y}_{R1}) \approx \frac{(1-f)}{n} \mu_y^2 [C_y^2 + \theta_1^2 C_x^2 - 2\theta_1 \rho_{xy} C_y C_x], \text{ where } \theta_1 = \frac{\mu_x}{\mu_x + C_x} \quad (4)$$

Upadhyay and Singh (1999) proposed two ratio type estimators utilizing the coefficient of variation and the coefficient of kurtosis of auxiliary variable ( $\beta_{2(x)}$ )

$$\bar{y}_{R2} = \bar{y} \left[ \frac{\beta_{2(x)} \mu_x + C_x}{\beta_{2(x)} \bar{x} + C_x} \right] \quad (5)$$

$$\bar{y}_{R3} = \bar{y} \left[ \frac{C_x \mu_x + \beta_{2(x)}}{C_x \bar{x} + \beta_{2(x)}} \right] \quad (6)$$

The MSE of the estimators  $\bar{y}_{R2}$  and  $\bar{y}_{R3}$ , up to the first order of approximation respectively are:

$$MSE(\bar{y}_{R2}) \approx \frac{(1-f)}{n} \mu_y^2 [C_y^2 + \theta_2^2 C_x^2 - 2\theta_2 \rho_{xy} C_y C_x] \quad (7)$$

$$MSE(\bar{y}_{R3}) \approx \frac{(1-f)}{n} \mu_y^2 [C_y^2 + \theta_3^2 C_x^2 - 2\theta_3 \rho_{xy} C_y C_x] \quad (8)$$

where  $\theta_2 = \frac{\beta_{2(x)} \mu_x}{\beta_{2(x)} \mu_x + C_x}$  and  $\theta_3 = \frac{C_x \mu_x}{C_x \mu_x + \beta_{2(x)}}$

Singh and Tailor (2003) proposed a ratio type estimator using correlation coefficient between auxiliary variable and the variable under study as:

$$\bar{y}_{R4} = \bar{y} \left( \frac{\mu_x + \rho_{xy}}{\bar{x} + \rho_{xy}} \right) \quad (9)$$

The MSE of the estimator  $\bar{y}_{R4}$ , up to the first order of approximation respectively are:

$$MSE(\bar{y}_{R4}) \approx \frac{(1-f)}{n} \mu_y^2 [C_y^2 + \theta_4^2 C_x^2 - 2\theta_4 \rho_{xy} C_y C_x], \text{ where } \theta_4 = \frac{\mu_x}{\mu_x + \rho_{xy}} \quad (10)$$

Kadilar and Cingi (2003) suggested the following estimator utilizing the auxiliary variable as:

$$\bar{y}_{R5} = \bar{y} \left( \frac{\mu_x^2}{\bar{x}^2} \right) \tag{11}$$

The MSE of the estimator  $\bar{y}_{R5}$ , up to the first order of approximation respectively are:

$$MSE(\bar{y}_{R5}) \approx \frac{(1-f)}{n} \mu_y^2 [C_y^2 + 4C_x^2 - 4\rho_{xy}C_yC_x] \tag{12}$$

Yan and Tian (2010) proposed two ratio type estimators using coefficients of skewness ( $\beta_{1(x)}$ ) and kurtosis ( $\beta_{2(x)}$ ) of auxiliary variable as:

$$\bar{y}_{R6} = \bar{y} \left[ \frac{\beta_{2(x)}\mu_x + \beta_{1(x)}}{\beta_{2(x)}\bar{x} + \beta_{1(x)}} \right] \tag{13}$$

$$\bar{y}_{R7} = \bar{y} \left[ \frac{\beta_{1(x)}\mu_x + \beta_{2(x)}}{\beta_{1(x)}\bar{x} + \beta_{2(x)}} \right] \tag{14}$$

The MSE of the estimators  $\bar{y}_{R6}$  and  $\bar{y}_{R7}$ , up to the first order of approximation respectively are:

$$MSE(\bar{y}_{R6}) \approx \frac{(1-f)}{n} \mu_y^2 [C_y^2 + \theta_6^2 C_x^2 - 2\theta_6 \rho_{xy} C_y C_x] \tag{15}$$

$$MSE(\bar{y}_{R7}) \approx \frac{(1-f)}{n} \mu_y^2 [C_y^2 + \theta_7^2 C_x^2 - 2\theta_7 \rho_{xy} C_y C_x] \tag{16}$$

where  $\theta_6 = \frac{\beta_{2(x)}\mu_x}{\beta_{2(x)}\mu_x + \beta_{1(x)}}$  and  $\theta_7 = \frac{\beta_{1(x)}\mu_x}{\beta_{1(x)}\mu_x + \beta_{2(x)}}$

### 3 Estimators Under Adaptive Cluster Sampling

Let the population consists of  $N$  distinct identifiable units labeled from  $1, 2, \dots, N$ . Let  $y_i$  and  $x_i (i = 1, 2, \dots, N)$  denote the observation on the characteristic  $x$  and  $y$  respectively, under study for the  $i^{th}$  unit of the population.

Let  $n$  denote the initial sample size and  $v$  denote the final sample size. Let  $\Psi_i$  denote the network that includes unit  $i$  and  $m_i$  be the number of units in that network. The initial sample of units is selected by simple random sampling without replacement.

The estimator of the population mean for the variable of interest under adaptive cluster sampling based on Hansen-Hurwitz estimator as,

$$\bar{y}_{ac} = \frac{1}{n} \sum_{i=1}^n (w_y)_i \tag{17}$$

where,  $(w_y)_i$  is the average of the variable  $y$  under study in the network that includes unit  $i$  of the initial sample, that is:  $(w_y)_i = \frac{1}{m_i} \sum_{j \in \Psi_i} (y_j)$

The variance of  $\bar{y}_{ac}$  is given by,

$$V(\bar{y}_{ac}) = \frac{N-n}{N(N-1)} \sum_{i=1}^N [(w_y)_i - \mu_y]^2 \tag{18}$$

Dryver and Chao (2007) proposed the following ratio estimator in adaptive cluster sampling as,

$$\bar{y}_{acR} = \frac{\bar{y}_{ac}}{\bar{x}_{ac}} \mu_x \tag{19}$$

where  $\bar{x}_{ac} = \frac{1}{n} \sum_{i=1}^n (w_x)_i$  and  $(w_x)_i$  is the average of the auxiliary variable  $x$  in the network that includes unit  $i$  of the initial sample, that is,  $(w_x)_i = \frac{1}{m_i} \sum_{j \in \mathcal{Y}_i} (x_j)$

The first order approximated MSE of  $\bar{y}_{acR}$  is

$$MSE(\bar{y}_{acR}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + C_{wx}^2 - 2\rho_{wx.wy} C_{wy} C_{wx}] \quad (20)$$

Chutiman (2013) proposed the following ratio type estimators of population mean using parameters of the auxiliary information based on Sisodia and Dwivedi (1981) estimator and Upadhyaya and Singh (1999) two estimators under adaptive cluster sampling as,

$$\bar{y}_{acR1} = \bar{y}_{ac} \left( \frac{\mu_y + C_{wx}}{\bar{x}_{ac} + C_{wx}} \right) \quad (21)$$

$$\bar{y}_{acR2} = \bar{y}_{ac} \left[ \frac{\mu_x \beta_{2(w_x)} + C_{wx}}{\bar{x}_{ac} \beta_{2(w_x)} + C_{wx}} \right] \quad (22)$$

$$\bar{y}_{acR3} = \bar{y}_{ac} \left[ \frac{\mu_x + \beta_{2(w_x)}}{\bar{x}_{ac} + \beta_{2(w_x)}} \right] \quad (23)$$

where  $C_{wx}$  is the population coefficient of variation of  $w_x$ ,  $\beta_{2(w_x)}$  is the population coefficient of kurtosis of  $w_x$ . Where  $S_{wy}^2 = \frac{1}{N-1} \sum_{i=1}^N [(w_y)_i - \mu_y]^2$ ,  $S_{wx}^2 = \frac{1}{N-1} \sum_{i=1}^N [(w_x)_i - \mu_x]^2$ ,  $S_{wx.wy} = \frac{1}{N-1} \sum_{i=1}^N [(w_x)_i - \mu_x][(w_y)_i - \mu_y] = \rho_{wx.wy} S_{wx} S_{wy}$

The mean square errors of above estimators using Taylor series method up to the first order of approximations respectively are,

$$MSE(\bar{y}_{acR1}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + \theta_{w1}^2 C_{wx}^2 - 2\theta_{w1} \rho_{wx.wy} C_{wy} C_{wx}] \quad (24)$$

$$MSE(\bar{y}_{acR2}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + \theta_{w2}^2 C_{wx}^2 - 2\theta_{w2} \rho_{wx.wy} C_{wy} C_{wx}] \quad (25)$$

$$MSE(\bar{y}_{acR3}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + \theta_{w3}^2 C_{wx}^2 - 2\theta_{w3} \rho_{wx.wy} C_{wy} C_{wx}] \quad (26)$$

where  $\theta_{w1} = \frac{\mu_y}{\mu_y + C_{wx}}$ ,  $\theta_{w2} = \frac{\mu_x \beta_{2(w_x)}}{\mu_x \beta_{2(w_x)} + C_{wx}}$ ,  $\theta_{w3} = \frac{\mu_x}{\mu_x + \beta_{2(w_x)}}$  and  $R = \frac{\mu_x}{\mu_y}$

## 4 Proposed Estimators

Motivated by Singh and Tailor (2003), Kadilar and Cingi (2003) and Yan and Tian (2010) estimators of population mean in simple random sampling given above, we proposed the estimators based on these mentioned estimators of population mean in adaptive cluster sampling as,

$$\bar{y}_{acR4} = \bar{y}_{ac} \left( \frac{\mu_y + \rho_{wx.wy}}{\bar{x}_{ac} + \rho_{wx.wy}} \right) \quad (27)$$

$$\bar{y}_{acR5} = \bar{y}_{ac} \left( \frac{\mu_x^2}{\bar{x}_{ac}^2} \right) \quad (28)$$

$$\bar{y}_{acR6} = \bar{y}_{ac} \left( \frac{\beta_{2(w_x)} \mu_x + \beta_{1(w_x)}}{\beta_{2(w_x)} \bar{x}_{ac} + \beta_{1(w_x)}} \right) \quad (29)$$

$$\bar{y}_{acR7} = \bar{y}_{ac} \left( \frac{\beta_{1(w_x)} \mu_x + \beta_{2(w_x)}}{\beta_{1(w_x)} \bar{x}_{ac} + \beta_{2(w_x)}} \right) \quad (30)$$

Using the method discussed by Chutiman (2013), the mean square errors of above estimators up to the first order of approximations respectively are,

$$MSE(\bar{y}_{acR4}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + \theta_{w4}^2 C_{wx}^2 - 2\theta_{w4} \rho_{wx,wy} C_{wy} C_{wx}] \tag{31}$$

$$MSE(\bar{y}_{acR5}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + 4C_{wx}^2 - 4\rho_{wx,wy} C_{wy} C_{wx}] \tag{32}$$

$$MSE(\bar{y}_{acR6}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + \theta_{w6}^2 C_{wx}^2 - 2\theta_{w6} \rho_{wx,wy} C_{wy} C_{wx}] \tag{33}$$

$$MSE(\bar{y}_{acR7}) \approx \frac{(1-f)}{n} \mu_y^2 [C_{wy}^2 + \theta_{w7}^2 C_{wx}^2 - 2\theta_{w7} \rho_{wx,wy} C_{wy} C_{wx}] \tag{34}$$

where  $\theta_{w4} = \frac{\mu_x}{\mu_x + \rho_{wx,wy}}$ ,  $\theta_{w6} = \frac{\mu_x \beta_{2(w_x)}}{\mu_x \beta_{2(w_x)} + \beta_{1(w_x)}}$ ,  $\theta_{w7} = \frac{\mu_x \beta_{1(w_x)}}{\mu_x \beta_{1(w_x)} + \beta_{2(w_x)}}$

### 5 Numerical Illustration

In this section, the simulation  $x$ -values and  $y$ -values from Chutiman and Kumphon (2008) were studied. The data statistics of this populations were shown in Table 1. We take the sample size as  $n = 20$  in Table 2, value of MSE which are computed using equations in Section 2-4, are given.

**Table 1: Data Statistics**

$N = 400$	$n = 20$	$\mu_y = 1.2225$	$\mu_x = 0.5550$
$S_y = 5.050$	$\theta_1 = 0.114$	$S_{wy} = 3.562$	$\theta_{w1} = 0.137$
$S_x = 2.400$	$\theta_2 = 0.876$	$S_{wx} = 1.948$	$\theta_{w2} = 0.9357$
$S_{xy} = 11.037$	$\theta_3 = 0.042$	$S_{wx,wy} = 6.428$	$\theta_{w3} = 0.006$
$\rho_{xy} = 0.910$	$\theta_4 = 0.379$	$\rho_{wx,wy} = 0.926$	$\theta_{w4} = 0.375$
$C_y = 4.131$	$\theta_6 = 0.817$	$C_{wy} = 2.914$	$\theta_{w6} = 0.864$
$C_x = 4.325$	$\theta_7 = 0.064$	$C_{wx} = 3.510$	$\theta_{w7} = 0.046$
$\beta_{1(x)} = 6.832$	$\beta_{2(x)} = 55.090$	$\beta_{1(w_x)} = 7.953$	$\beta_{2(w_x)} = 91.369$

**Table 2: MSE Values of Estimators**

$MSE(\bar{y}_{R1}) = 0.966$	$MSE(\bar{y}_{R2}) = 0.407$	$MSE(\bar{y}_{R3}) = 1.117$	$MSE(\bar{y}_{R4}) = 0.527$	$MSE(\bar{y}_{R5}) = 1.904$
$MSE(\bar{y}_{R6}) = 0.311$	$MSE(\bar{y}_{R7}) = 1.068$	$MSE(\bar{y}_{acR1}) = 0.435$	$MSE(\bar{y}_{acR2}) = 0.309$	$MSE(\bar{y}_{acR3}) = 0.595$
Proposed Estimators	$MSE(\bar{y}_{acR4}) = 0.221$	$MSE(\bar{y}_{acR5}) = 1.010$	$MSE(\bar{y}_{acR6}) = 0.093$	$MSE(\bar{y}_{acR7}) = 0.542$

### 6 Conclusion

In the present manuscript, we developed new ratio type estimators for the estimation of population mean by using auxiliary information in adaptive cluster sampling scheme. The bias and the mean squared error of proposed estimators have been obtained up to the first order of approximation. An empirical study is carried out and from the estimated MSE of the estimators in Table-2, it is clear that if the population is rare and hidden clustered population, all estimators in adaptive cluster sampling are more efficient than the estimators of population mean in simple random sampling, given the same condition. Further among the all mentioned estimators of population mean along with all proposed estimators in adaptive cluster sampling, the proposed estimator,  $\bar{y}_{acR6}$  has the smallest estimated mean square error. Therefore, it should preferably be adopted for the estimation of population mean in adaptive cluster sampling scheme.

## Acknowledgement

The authors are very much thankful to the editor in chief and the unknown learned referee for critically examining the manuscript and giving valuable suggestions to improve it.

## References

- [1] N. Chutiman , Adaptive cluster sampling using auxiliary variable, *J.Math.Stat.*, **9**, 249-255 (2013).
- [2] N. Chutiman and B. Chutiman, Ratio estimator using auxiliary variable for adaptive cluster sampling, *Journal of Thai Statistical Association*, **6(2)**,241-256 (2008)
- [3] W.G.Cochran, The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce, *Jour. Agri. Sci.*,**59**,1225-1226 (1940).
- [4] A.L. Dryver and C. Chao , Ratio estimators in adaptive cluster sampling, *Environmetrics*,**18**, 607-620 (1967).
- [5] C. Kadilar, and H. Cingi, A study on the chain ratio-type estimator, *Hacettepe Journal of Mathematics and Statistics*,**32**,105-108 (2003).
- [6] H. P. Singh and R. Tailor, Use of known correlation coefficient in estimating the finite population mean, *Statistics in Transition* ,**6**,555-560 (2003).
- [7] B.V.S. Sisodia and V.K. Dwivedi, A modified ratio estimator using coefficient of variation of auxiliary variable, *J. Indian Soc. Agric. Statist.*, **33**, 13-18 (1981).
- [8] S.K. Thompson, Adaptive cluster sampling, *J. Am. Statist. Assoc.*, **85**, 1050-1059 (1990).
- [9] L.N. Upadhyaya and H.P. Singh, Use of transformed auxiliary variable in estimating the finite population mean, *Biometri. J.*, **41**, 627-636(1999).
- [10] Z. Yan and B. Tian, Ratio Method to the Mean Estimation Using Co-efficient of Skewness of Auxiliary Variable, , *ICICA 2010, Part II, CCIS 106*, 103-110(2010).



**Subhash Kumar Yadav** is working as assistant professor in the Department of Mathematics and Statistics, Dr Ram Manohar Lohia Avadh University, Faizabad. He has got published many research papers in the field sampling of Statistics.



**Sheela Misra** , the supervisor of Dr Subhash Kumar Yadav is working as Professor in the Department of Statistics, University of Lucknow, Lucknow. She is a very good academician as well as the administrator. She has successfully organized many international conferences in Statistics. She has done a lot of work in different field of Statistics.



**Sant Sharan Mishra** is Associate Professor at Department of Mathematics and Statistics, Dr Ram Manohar Lohia Avadh University, Faizabad. He is a senior faculty in the Department. He has supervised many Ph.D. in different fields of Mathematics and Statistics. He has got published a lot of research papers in different journals of repute.



**Nipaporn Chutiman** is working as senior Faculty in the Department of Mathematics, Maharakham University, Maha Sarakham, Thailand. She is doing very good work in the field sampling of Statistics. She has got published many research papers in different reputed journals of Statistics.