

SOME CHAIN-TYPE EXPONENTIAL ESTIMATORS OF POPULATION MEAN IN TWO-PHASE SAMPLING

G. N. Singh¹, D. Majhi²

ABSTRACT

Using the information on two-auxiliary variables, three different exponential chain-type estimators of population mean of study variable have been proposed in two-phase (double) sampling. Properties of the proposed estimators have been studied and their performances are examined with respect to several well known chain-type estimators. Empirical studies are carried out to support the theoretical results.

Key words: two-phase, auxiliary information, bias, mean square error.
Mathematics subject classification: 62D05

1. Introduction

Ratio, product and regression methods of estimation require the knowledge of population mean of the auxiliary variable. If population mean of the auxiliary variable is not known, it is customary to move towards two-phase sampling scheme, which provides a cost effective estimate of the unknown population mean of auxiliary variable in first-phase sample. Utilizing the information on known population mean of another auxiliary variable in first-phase sample, Chand (1975) introduced chain-type ratio estimator of population mean of study variable. His work was further extended by Kiregyera (1980, 1984), Mukherjee *et al.* (1987), Srivastava *et al.* (1989), Upadhyaya *et al.* (1990), Singh and Singh (1991), Singh *et al.* (1994), Singh and Upadhyaya (1995), Upadhyaya and Singh (2001), Singh (2001), Pradhan (2005), Gupta and Shabbir (2008) and Singh *et al.* (2011) among others. Motivated with the above works, the aim of the present research is to propose some different structures of chain-type estimators in two-phase sampling which may estimate the population mean in a more precise way in comparison with the contemporary estimators of similar kind.

¹ Department of Applied Mathematics, Indian School of Mines, Dhanbad-826004, India.
E-mail: gnsingh_ism@yahoo.com.

² Department of Applied Mathematics, Indian School of Mines, Dhanbad-826004, India.

2. Two-phase sampling set-up

Consider a finite population U of size N indexed by triplet characters (y, x, z) . We wish to estimate the population mean \bar{Y} of study variable y in the presence of two auxiliary variables x and z . Let x and z be called first and second auxiliary variables respectively such that y is highly correlated with x while in comparison with x it is remotely correlated with z (i.e. $\rho_{yx} > \rho_{yz}$). When the population mean \bar{X} of x is unknown but information on z is available on all the units of the population, we use the following two-phase sampling scheme.

Let us now consider a two-phase sampling where in the first phase a large (preliminary) sample s' ($s' \subset U$) of fixed size n' is drawn following SRSWOR to observe two auxiliary variables x and z to estimate \bar{X} , while in the second phase a sub-sample $s \subset s'$ of fixed size n is drawn by SRSWOR to observe the characteristic y under study.

3. Estimators based on one auxiliary variable

Ratio and regression estimators in two-phase sampling are the traditional estimators utilizing the information on one auxiliary variable and are reproduced below along with their respective mean square errors up to $o(n^{-1})$.

$$\bar{y}_{rd} = \frac{\bar{Y}}{\bar{X}} \bar{x}' \quad (1)$$

$$M(\bar{y}_{rd}) = \bar{Y}^2 \left[f_1 C_y^2 + f_3 (C_x^2 - 2\rho_{yx} C_y C_x) \right] \quad (2)$$

$$\bar{y}_{lrd} = \bar{y} + b_{yx}(n)(\bar{x}' - \bar{x}) \quad (3)$$

$$M(\bar{y}_{lrd}) = S_y^2 \left[f_1 (1 - \rho_{yx}^2) + f_2 \rho_{yx}^2 \right] \quad (4)$$

where $b_{yx}(n)$ is the sample regression coefficient of y on x calculated from the data based on s and

$$\bar{y} = \frac{1}{n} \sum_{i \in s} y_i, \quad \bar{x} = \frac{1}{n} \sum_{i \in s} x_i \quad \text{and} \quad \bar{x}' = \frac{1}{n'} \sum_{i \in s'} x_i, \quad f_1 = \left(\frac{1}{n} - \frac{1}{N} \right), \quad f_2 = \left(\frac{1}{n'} - \frac{1}{N} \right), \quad f_3 = (f_1 - f_2) = \left(\frac{1}{n} - \frac{1}{n'} \right)$$

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2, \quad S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2, \quad C_x = \frac{S_x}{\bar{X}}, \quad C_y = \frac{S_y}{\bar{Y}}, \quad \text{and} \quad \rho_{yx} \text{ be the correlation}$$

coefficient between the variables y and x . \bar{X} and \bar{Y} are the population means of the variables x and y , respectively.

4. Estimators based on two auxiliary variables

Chand (1975) introduced a chain-type ratio estimator under two-phase sampling using two auxiliary variables x and z when the population mean \bar{X} of x is unknown but information on z is available on all the units of the population, which is given as

$$\bar{y}_{rc} = \frac{\bar{y} \bar{x}'}{\bar{x} \bar{z}'} \bar{Z} \tag{5}$$

The mean square error of the estimator \bar{y}_{rc} up to $o(n^{-1})$ is derived as

$$M(\bar{y}_{rc}) = \bar{Y}^2 \left[f_1 C_y^2 + f_3 (C_x^2 - 2\rho_{yx} C_y C_x) + f_2 (C_z^2 - 2\rho_{yz} C_y C_z) \right] \tag{6}$$

where \bar{Z} is the population mean of the variable z , $\bar{z}' = \frac{1}{n} \sum_{i \in s'} z_i$, $C_z = \frac{S_z}{\bar{Z}}$,

$S_z^2 = \frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{Z})^2$ and ρ_{yz} be correlation coefficient between variables y and z .

Kiregyera (1980, 1984) extended the work of Chand (1975) and suggested chain-type ratio to regression, regression to ratio and regression to regression estimators of population mean of study variable y in two-phase sampling which utilized the information on two auxiliary variables. The suggested estimators are given below along with their respective mean square errors up to $o(n^{-1})$.

$$\bar{y}_{k1} = \frac{\bar{y}}{\bar{x}} \left[\bar{x}' + b_{xz} (n') (\bar{Z} - \bar{z}') \right] \tag{7}$$

$$M(\bar{y}_{k1}) = \bar{Y}^2 \left[f_3 (C_x^2 + C_y^2 - 2\rho_{yx} C_y C_x) + f_2 C_y^2 + f_2 \rho_{xz} C_x (\rho_{xz} C_x - 2\rho_{yz} C_y) \right] \tag{8}$$

$$\bar{y}_{k2} = \bar{y} + b_{yx} (n) (\bar{x}'_{rd} - \bar{x}); \bar{x}'_{rd} = \frac{\bar{x}'}{\bar{z}'} \bar{Z} \tag{9}$$

$$M(\bar{y}_{k2}) = \bar{Y}^2 C_y^2 \left[f_1 (1 - \rho_{yx}^2) + f_2 \left(\rho_{yx}^2 + \rho_{yx}^2 \frac{C_z^2}{C_x^2} - 2\rho_{yx} \rho_{yz} \frac{C_z}{C_x} \right) \right] \tag{10}$$

$$\bar{y}_{k3} = \bar{y} + b_{yx} (n) (\bar{x}'_{ld} - \bar{x}); \bar{x}'_{ld} = \left[\bar{x}' + b_{xz} (n') (\bar{Z} - \bar{z}') \right] \tag{11}$$

$$M(\bar{y}_{k3}) = \bar{Y}^2 C_y^2 \left[f_3 (1 - \rho_{yx}^2) + f_2 (1 + \rho_{xz}^2 \rho_{yx}^2 - 2\rho_{yx} \rho_{yz} \rho_{xz}) \right] \tag{12}$$

where $b_{xz}(n')$ is the sample regression coefficient of the variable x on z calculated from the data based on s' and ρ_{xz} be correlation coefficient between variables x and z .

5. Proposed estimators

The intelligible use of auxiliary information at estimation stage is a fascinating act in sample surveys. In presence of auxiliary information Bahl and Tuteja (1991) suggested an exponential structure to estimate the population mean of study variable and their work was extended by Singh and Vishwakarma (2007) in two-phase sampling scheme. Motivated with the work related to the proposition of chain-type estimators in two-phase sampling set-up, and looking on the nice behaviours of the exponential type estimators, we suggest below three different chain-types exponential estimators of population mean \bar{Y} of the study variable y . The suggested estimators are given as

$$T_1 = \frac{\bar{y}}{\bar{x}} \bar{x}' \exp\left(\frac{\bar{Z} - \bar{z}'}{\bar{Z} + \bar{z}'}\right) \quad (13)$$

$$T_2 = \bar{y} + b_{yx}(n) \left\{ \bar{x}' \exp\left(\frac{\bar{Z} - \bar{z}'}{\bar{Z} + \bar{z}'}\right) - \bar{x} \right\} \quad (14)$$

and

$$T_3 = \bar{y} \exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) \left(\frac{\bar{Z}}{\bar{z}'}\right) \quad (15)$$

6. Properties of the estimators T_i ($i=1,2,3$)

Theorem 6.1. Biases of the estimators T_i ($i=1,2,3$) defined in equations (13), (14) and (15) up to $o(n^{-1})$ are obtained as

$$B(T_1) = \bar{Y} \left[f_3 (C_x^2 - \rho_{yx} C_y C_x) + \frac{f_2}{2} \left(\frac{3}{4} C_z^2 - \rho_{yz} C_y C_z \right) \right] \quad (16)$$

$$B(T_2) = \beta_{yx} \left[f_3 \left(\frac{\mu_{300}}{\mu_{200}} - \frac{\mu_{210}}{\mu_{110}} \right) + \frac{f_2}{2} \left(\frac{3}{4} \frac{\bar{X}}{\bar{Z}^2} \mu_{002} - \frac{1}{\bar{Z}} \mu_{101} + \frac{\bar{X}}{\bar{Z}} \frac{\mu_{201}}{\mu_{200}} - \frac{\bar{X}}{\bar{Z}} \frac{\mu_{111}}{\mu_{110}} \right) \right] \quad (17)$$

and

$$B(T_3) = \bar{Y} \left[\frac{f_3}{2} \left(\frac{3}{4} C_x^2 - \rho_{yx} C_y C_x \right) + f_2 \left(C_z^2 - \rho_{yz} C_y C_z \right) \right] \tag{18}$$

where $\mu_{rst} = E \left[(x_i - \bar{X})^r (y_i - \bar{Y})^s (z_i - \bar{Z})^t \right]; (r, s, t) \geq 0$ are integers.

Theorem 6.2. Mean square errors of the estimators $T_i (i=1,2,3)$ defined in equations (13), (14) and (15) up to $o(n^{-1})$ are derived as

$$M(T_1) = \bar{Y}^2 \left[f_1 C_y^2 + f_3 \left(C_x^2 - 2\rho_{yx} C_y C_x \right) + \frac{f_2}{4} \left(C_z^2 - 4\rho_{yz} C_y C_z \right) \right] \tag{19}$$

$$M(T_2) = \bar{Y}^2 C_y^2 \left[f_3 \left(1 - \rho_{yx}^2 \right) + f_2 \left(1 + \frac{\rho_{yx}^2 C_z^2}{4 C_x^2} - \rho_{yx} \rho_{yz} \frac{C_z}{C_x} \right) \right] \tag{20}$$

and

$$M(T_3) = \bar{Y}^2 \left[f_1 C_y^2 + \frac{f_3}{4} \left(C_x^2 - 4\rho_{yx} C_y C_x \right) + f_2 \left(C_z^2 - 2\rho_{yz} C_y C_z \right) \right] \tag{21}$$

7. Comparison of the estimators

In this section we compare the proposed estimators $T_i (i=1,2,3)$ with to respect to the estimators $\bar{y}_{rd}, \bar{y}_{lrd}, \bar{y}_{rc}, \bar{y}_{kl}, \bar{y}_{k2}$ and \bar{y}_{k3} . Preference zones of the estimators T_i are explored and shown below:

(i) $T_i (i=1,2,3)$ are better than \bar{y}_{rd} if $M(T_i) \leq M(\bar{y}_{rd})$, which gives

$$\rho_{yz} \frac{C_y}{C_z} \geq \frac{1}{4} \quad (\text{for } i=1) \tag{22}$$

$$\frac{4(C_x - \rho_{yx} C_y)^2}{C_y^2 \left(\rho_{yx}^2 \frac{C_z^2}{C_x^2} - 4\rho_{yx} \rho_{yz} \frac{C_z}{C_x} \right)} \geq \frac{f_2}{f_3} \quad (\text{for } i=2) \tag{23}$$

$$\frac{(3C_x^2 - 4\rho_{yx} C_y C_x)}{4(C_z^2 - 2\rho_{yz} C_y C_z)} \geq \frac{f_2}{f_3} \quad (\text{for } i=3) \tag{24}$$

(ii) T_i ($i=1, 2, 3$) are preferable over \bar{y}_{Ird} if $M(T_i) \leq M(\bar{y}_{\text{Ird}})$, which gives

$$\frac{4(C_x - \rho_{yx} C_y)^2}{(4\rho_{yz} C_y C_z - C_z^2)} \leq \frac{f_2}{f_3} \quad (\text{for } i = 1) \quad (25)$$

$$\left(\frac{\rho_{yz}}{\rho_{yx}} \right) \frac{C_x}{C_z} \geq \frac{1}{4} \quad (\text{for } i = 2) \quad (26)$$

$$\frac{(C_x - 2\rho_{yx} C_y)^2}{4(2\rho_{yz} C_y C_z - C_z^2)} \leq \frac{f_2}{f_3} \quad (\text{for } i = 3) \quad (27)$$

(iii) T_i ($i=1, 2, 3$) will dominate \bar{y}_{rc} if $M(T_i) \leq M(\bar{y}_{\text{rc}})$ and subsequently we get the conditions

$$\rho_{yz} \frac{C_y}{C_z} \leq \frac{3}{4} \quad (\text{for } i = 1) \quad (28)$$

$$\frac{(C_x - \rho_{yx} C_y)^2}{\left\{ \frac{C_y^2}{4} \left(\frac{\rho_{yx} C_z}{C_x} - 2\rho_{yz} \right)^2 - (C_z - \rho_{yz} C_y)^2 \right\}} \geq \frac{f_2}{f_3} \quad (\text{for } i = 2) \quad (29)$$

$$\rho_{yx} \frac{C_y}{C_x} \leq \frac{3}{4} \quad (\text{for } i = 3) \quad (30)$$

(iv) T_i ($i=1, 2, 3$) are more efficient than \bar{y}_{kl} if $M(T_i) \leq M(\bar{y}_{\text{kl}})$, which gives

$$\left| \frac{(\rho_{yz} C_y - \rho_{xz} C_x)}{(C_z - 2\rho_{yz} C_y)} \right| \geq \frac{1}{2} \quad (\text{for } i = 1) \quad (31)$$

$$\frac{(C_x - \rho_{yx} C_y)^2}{\left\{ C_y^2 \left(\frac{\rho_{yx}^2 C_z^2}{4C_x^2} - \rho_{yx} \rho_{yz} \frac{C_z}{C_x} \right)^2 - (\rho_{xz}^2 C_x^2 - 2\rho_{xz} \rho_{yz} C_y C_x)^2 \right\}} \geq \frac{f_2}{f_3} \quad (\text{for } i = 2) \quad (32)$$

$$\frac{(4\rho_{yx} C_y C_x - 3C_x^2)}{4\left\{ (\rho_{xz} C_x - \rho_{yz} C_y)^2 - (C_z - \rho_{yz} C_y)^2 \right\}} \leq \frac{f_2}{f_3} \quad (\text{for } i = 3) \quad (33)$$

(v) $T_i (i=1, 2, 3)$ are preferable over \bar{y}_{k2} if $M(T_i) \leq M(\bar{y}_{k2})$, which gives

$$\frac{(C_x - \rho_{yx} C_y)^2}{\left[\left\{ C_y \left(\rho_{yx} \frac{C_z}{C_x} - \rho_{yz} \right) \right\}^2 - \left(\frac{C_z}{2} - \rho_{yz} C_y \right)^2 \right]} \leq \frac{f_2}{f_3} \text{ (for } i=1) \tag{34}$$

$$\frac{\rho_{yz} C_x}{\rho_{yx} C_z} \leq \frac{3}{4} \text{ (for } i=2) \tag{35}$$

$$\frac{(C_x - 2\rho_{yx} C_y)^2}{4 \left\{ \left(\rho_{yx} \frac{C_z}{C_x} \right) \left(\rho_{yx} \frac{C_z}{C_x} - 2\rho_{yz} \right) + (2\rho_{yz} C_y C_z - C_z^2) \right\}} \leq \frac{f_2}{f_3} \text{ (for } i=3) \tag{36}$$

(vi) $T_i (i=1, 2, 3)$ will dominate \bar{y}_{k3} if $M(T_i) \leq M(\bar{y}_{k3})$ and subsequently we get the conditions

$$\frac{(C_x - \rho_{yx} C_y)^2}{\left[\left\{ C_y (\rho_{yx} \rho_{xz} - \rho_{yz}) \right\}^2 - \left\{ \frac{1}{2} (C_z - \rho_{yz} C_y) \right\}^2 \right]} \leq \frac{f_2}{f_3} \text{ (for } i=1) \tag{37}$$

$$\left[\rho_{yx} \rho_{xz}^2 - \rho_{yz} \rho_{xz} \right] \geq \left[\frac{1}{4} \left(\frac{C_z}{C_x} \right) (\rho_{yx} - 4\rho_{yz}) \right] \text{ (for } i=2) \tag{38}$$

$$\frac{(C_x - 2\rho_{yx} C_y)^2}{4 \left\{ (\rho_{yx} \rho_{xz}) (\rho_{yx} \rho_{xz} - 2\rho_{yz}) + (2\rho_{yz} C_y C_z - C_z^2) \right\}} \leq \frac{f_2}{f_3} \text{ (for } i=3) \tag{39}$$

8. Empirical studies

To examine the performance of the proposed estimators $T_i (i=1, 2, 3)$, we have computed the percent relative efficiencies of T_i with respect to $\bar{y}_{rd}, \bar{y}_{lrd}, \bar{y}_{rc}, \bar{y}_{k1}$ based on two natural populations which almost satisfy the conditions shown in equations (22)-(33) and presented in Table 1. The percent relative efficiencies of the estimators T_i with respect to an estimator δ are

defined as $PRE = \left[\frac{MSE(\delta)}{MSE(T_i)} \right] \times 100 ; (i=1, 2, 3)$.

Population Source-I: Cochran (1977)

y: Number of 'placebo' children.

x: Number of paralytic polio cases in the placebo group.

z: Number of paralytic polio cases in the 'not inoculated' group

$N = 34$, $n = 10$, $n' = 15$, $\bar{Y} = 4.92$, $\bar{X} = 2.59$, $\bar{Z} = 2.91$, $C_y^2 = 1.0248$,
 $C_x^2 = 1.5175$, $C_z^2 = 1.1492$, $\rho_{yx} = 0.7326$, $\rho_{yz} = 0.6430$ and $\rho_{xz} = 0.6837$.

Population Source-II: Fisher (1936)

Consisting of measurements on three variables, namely sepal width (y), sepal length (x) and petal length (z) for 50 Iris flowers (versicolor) such that:

$N = 50$, $n = 10$, $n' = 20$, $\bar{Y} = 2.770$, $C_y^2 = 0.012566$, $C_x^2 = 0.007343$,
 $C_z^2 = 0.011924$, $\rho_{yx} = 0.5605$, $\rho_{yz} = 0.5259$ and $\rho_{xz} = 0.7540$.

Table 1. The percent relative efficiencies (PRE) based on population I and population II of the proposed estimators T_i ($i=1,2,3$) with respect to the estimators \bar{y}_{rd} , \bar{y}_{lrd} , \bar{y}_{rc} , \bar{y}_{kl} .

Estimators	PRE FOR POPULATION I			
	\bar{y}_{rd}	\bar{y}_{lrd}	\bar{y}_{rc}	\bar{y}_{kl}
T_1	132.7310	115.5850	113.0849	101.8650
T_2	146.8405	127.8718	125.1060	112.6934
T_3	136.7430	119.0787	116.5031	104.9441
PRE FOR POPULATION II				
T_1	114.1981	110.6203	110.2922	100.0537
T_2	116.7494	113.0917	112.7563	102.2890
T_3	104.3392	101.0703	100.7705	91.4159

9. Conclusions

It is visible in Table 1 that the proposed estimators T_i ($i=1,2,3$) are preferable over the estimators \bar{y}_{rd} , \bar{y}_{lrd} , \bar{y}_{rc} and \bar{y}_{kl} except the estimator T_3 which is being dominated by the estimator \bar{y}_{kl} in population II. The proposed estimators

will also be preferable over the estimators \bar{y}_{k2} and \bar{y}_{k3} for the population which satisfies the conditions derived in equations (34)-(39). Hence, looking on the dominance nature of the proposed estimators, they may be recommended for their practical applications.

Acknowledgments

The authors are thankful to the reviewers for their valuable and inspiring suggestions. The authors are also thankful to the UGC, New Delhi and the Indian School of Mines, Dhanbad for providing the financial assistance and necessary infrastructure to carry out the present research work.

REFERENCES

- BAHL, S., TUTEJA, R. K., (1991). Ratio and product type exponential estimator. *Information and optimization sciences*, 12, 159–163.
- CHAND, L., (1975). Some ratio-type estimators based on two or more auxiliary variables. Ph. D. dissertation, Iowa State University, Ames, Iowa.
- COCHRAN, W. G., (1977). *Sampling techniques*. New-York: JohnWiley and Sons.
- FISHER, R. A., (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, 7, 179–188.
- GUPTA, S., SHABBIR, J., (2008). On improvement in estimating the population mean in simple random sampling. *Journal of Applied Statistics*, 35 (5), 559–566.
- KIREGYERA, B., (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika*, 27, 217–223.
- KIREGYERA, B., (1984) Regression-type estimators using the two auxiliary variables and the model of double sampling from finite population. *Metrika*, 31(3-4), 215–226.
- MUKHERJEE, R., RAO, T. J., VIJAYAN, K., (1987). Regression type estimators using multiple auxiliary information. *Australian & New Zealand Journal of Statistics*, 29, 244–254.
- PRADHAN, B. K., (2005). A chain regression estimator in two-phase sampling using multi-auxiliary information. *Bulletin of the malaysian mathematical sciences society*, 28(1), 81–86.

- SINGH, H. P., VISHWAKARMA, G. K., (2007). Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian journal of statistics*, 36, 3, 217–225.
- SINGH, G. N., (2001). On the use of transformed auxiliary variable in the estimation of population mean in two-phase sampling. *Statistics in Transition*, 5(3), 405–416.
- SINGH, G. N., UPADHYAYA, L. N., (1995). A class of modified chain type estimators using two auxiliary variables in two-phase sampling. *Metron*, LIII, 117–125.
- SINGH, V. K., SINGH, G. N., (1991). Chain type regression estimators with two auxiliary variables under double sampling scheme. *Metron*, Vol. XLIX, No. 1–4, 279–289.
- SINGH, V. K., SINGH, H. P., SINGH, H. P., (1994). A general class of chain estimators for ratio and product of two means of a finite population. *Communications in Statistics-Theory and Methods*, 23, 1341–1355.
- SINGH, R., CHAUHAN, P., SMARANDACHE, F., (2011). Improvement in estimating population mean using two-auxiliary variables in two-phase sampling. *Italian Journal of Pure and Applied Mathematics*, 28, 135–142.
- SRIVASTAVA, S. R., SRIVASTAVA, S. P., KHARE, B. B., (1989). Chain ratio type estimators for ratio of two population means using auxiliary characters. *Communications in Statistics-Theory and Methods*, 18, 3917–3926.
- UPADHYAYA, L. N., KUSHWAHA, K. S., SINGH, H. P., (1990) A modified chain ratio-type estimator in two-phase sampling using multi-auxiliary information. *Metron*, 48, 381–393.
- UPADHYAYA, L. N., SINGH, G. N., (2001). Chain type estimators using transformed auxiliary variable in two-phase sampling. *Advances in modeling and analysis*, 38, (1–2), (1–10).