

Some Improved Estimators in Double Sampling Using two Auxiliary Variables

Mohammad S. Ahmed

Department of Mathematics and Statistics, College of Science, Sultan Qaboos University,
P.O.Box 36, Al-Khoud, 123 Muscat, Sultanate of Oman. Email: msahmed@squ.edu.om.

ABSTRACT: Dash and Mishra [1] suggested an improved class of estimators without defining the optimum estimator. However, they gave the wrong Taylor's series expression of their class of estimator and their minimum mean squared error expressions are also incorrect. Here we show that Ahmed *et al.*'s [2] class of chain estimators is more efficient than Dash and Mishra's [1], with minimum mean squared error.

Keywords: Chain based estimator; Double sampling; Auxiliary variable; Mean square error.

بعض المقدرات المتطورة في العينات المزدوجة باستخدام اثنتين من المتغيرات المساعدة

محمد ص. أحمد

ملخص: داش وميشرا [1] اقترحا فئة متطورة للمقدرات بدون تحديد المقدر الأمثل ولكن للأسف أعطوا التعبير الخاطئ لسلسلة تايلور للفئة المقدره وأيضا تعابير الحد الأدنى لمتوسط الخطأ التربيعي لها أيضا غير صحيحه. وقد بينا هنا أن فئة من سلسلة المقدرات لأحمد وآخرون. [2] هي أكثر كفاءة من داش وميشرا [1] مع الحد الأدنى لمتوسط الخطأ التربيعي.

مفتاح الكلمات: تقدير حلقي، العينات المزدوجة، المتغير المساعدة، متوسط الخطأ التربيعي.

1. Introduction

Suppose one is interested in estimating the population mean of a study variable Y from a finite population of size N . If the information on an auxiliary variable X which is highly correlated with Y is readily available on all the units of the population, it is well known that ratio and regression type estimators can be used for increased efficiency, incorporating the knowledge of \bar{X} . However, in certain practical situations \bar{X} is not known a priori, in which case the technique of Double sampling is fruitfully applied. The values of X are assumed to be known over a large simple random sample of size n' . Now suppose that information on yet another auxiliary variable Z is available on all units of the population (\bar{Z} is the population mean of Z). Then we observe the study variable Y from a simple random subsample of that large sample of size n ($n \leq n'$) to estimate the population mean (\bar{Y}) of study variable Y . The first phase sample s' of size n' ($n' \leq N$) is selected to observe X and Z , and the second phase sample s of size n ($n \leq n'$) is selected to observe Y , X and Z .

Suppose \bar{x}' and \bar{z}' are the unbiased estimators of \bar{X} and \bar{Z} , respectively, based on first phase sample s' ; \bar{y} , \bar{x} and \bar{z} are the unbiased estimators of \bar{Y} , \bar{X} and \bar{Z} , respectively, based on second phase sample, s . Thus, a specific class of estimators can be suggested by suitably choosing the auxiliary information. Over a period, a large number of estimators were proposed under different assumptions which can be classified as a particular member or case of some general class of estimators. In section 2, we critically evaluate some of these estimators.

2. Some Classes of Estimators

Using information on two closely related auxiliary variables, X and Z , a good number of classes of improved estimators were suggested by Chand [3], Kiregyera [4, 5], Srivastava *et al.* [6], Sahoo *et al.* [7], Tripathi and Ahmed [8], and Mishra and Rout [9]. A brief review of these estimators may be referred to Das and Mishra [1].

Dash and Mishra [1] suggested an improved class of estimators as defined

$$d_{dm} = h(\bar{y}, \bar{x}, \bar{z}, \bar{x}', \bar{z}') \quad (1.1)$$

Ahmed [10] proposed a wider class of estimators as

$$d_h = h(\bar{y}, u, v, v') \quad (1.2)$$

where $u = \frac{\bar{x}}{\bar{x}'}$, $v = \frac{\bar{z}}{\bar{z}'}$ and $v' = \frac{\bar{z}'}{\bar{z}}$; d_h is a function such that $h(\bar{y}, 1, 1, 1) = \bar{y}$ satisfying some regularity conditions.

Tripathi and Ahmed [8] suggested the general class of estimator:

$$d_{ta} = \bar{y} + t_1(\bar{x}' - \bar{x}) + t_2(\bar{z}' - \bar{z}) + t_3(\bar{Z} - \bar{z}') \quad (1.3)$$

where t_1, t_2 and t_3 are suitably chosen statistics such that $E(t_i) = \tau_i$ or $E(t_i) \approx \tau_i$.

With suitable choices of t_1, t_2 and t_3 , Tripathi and Ahmed [8] showed that Chand [3], Kiregyera [4, 5], Mukerjee *et al.* [11] and Sahoo *et al.* [7] are the particular cases (1.3). Further, Ahmed [12] also showed that the recent estimators of Mishra and Rout [9], Upadhyay and Singh [13], Samiuddin and Hanif [14] and Hanif *et al.* [15] are also the particular case of (1.3).

The optimum values of τ_1, τ_2 and τ_3 are respectively $\tau_{1_{opt}} = B_{yx.z}$, $\tau_{2_{opt}} = B_{yz.x}$ and $\tau_{3_{opt}} = B_{yz}$, where $B_{yx.z}$ and $B_{yz.x}$ are partial regression coefficients and B_{yz} is a the total regression coefficient (See Tripathi and Ahmed [8]).

The optimum estimator (1.3) is

$$d_{ta_0} = \bar{y} + b_{yx.z}(\bar{x}' - \bar{x}) + b_{yz.x}(\bar{z}' - \bar{z}) + b_{yz}(\bar{Z} - \bar{z}') \quad (1.4)$$

where $b_{yx.z}$ and $b_{yz.x}$ are sample partial, and b_{yz} is the sample total, regression coefficients respectively.

Ahmed *et al.* [2] suggested another general class of estimators:

$$d_{at} = \bar{y} \left(\frac{\bar{x}'}{\bar{x}}\right)^{\beta_1} \left(\frac{\bar{z}'}{\bar{z}}\right)^{\beta_2} \left(\frac{\bar{Z}}{\bar{z}'}\right)^{\beta_3} \quad (1.5)$$

where β_1, β_2 and β_3 are suitable chosen constants.

The optimum values of β_1, β_2 and β_3 are respectively,

$$\beta_{1_{opt}} = \frac{B_{yx.z}}{R_x}, \beta_{2_{opt}} = \frac{B_{yz.x}}{R_z} \text{ and } \beta_{3_{opt}} = \frac{B_{yz}}{R_z} \quad (R_x = \frac{\bar{y}}{\bar{x}} \text{ and } R_z = \frac{\bar{y}}{\bar{z}})$$

The optimum estimator (1.3) is

$$d_{at_0} = \bar{y} \left(\frac{\bar{x}'}{\bar{x}}\right)^{b_1} \left(\frac{\bar{z}'}{\bar{z}}\right)^{b_2} \left(\frac{\bar{Z}}{\bar{z}'}\right)^{b_3} \quad (1.6)$$

where $b_1 = \frac{b_{yx.z}}{r_x}$, $b_2 = \frac{b_{yz.x}}{r_z}$, $b_3 = \frac{b_{yz}}{r_z}$ ($r_x = \frac{\bar{y}}{\bar{x}}$ and $r_z = \frac{\bar{y}}{\bar{z}}$)

It can be shown that the suggested estimators of Chand [3], Srivastava *et al.* [6], Samiuddin and Hanif [14] and Hanif *et al.* [15] are the particular cases of the class of estimators (1.5) and that it has minimum mean squared error (MSE).

Ahmed *et al.*[16] showed that both general classes of estimators (1.3), (1.5) and that of Singh *et al.* [8] are a subclass of (1.2).

The minimum MSE of (1.2), (1.3) and (1.5) is the same and it is

$$M_{min} = S_y^2 \left[\left(\frac{1}{n} - \frac{1}{n'}\right) (1 - \rho_{y.zx}^2) + \left(\frac{1}{n'} - \frac{1}{N}\right) (1 - \rho_{yz}^2) \right] \quad (1.7)$$

where ρ_{yz} is the population correlation coefficient between y and z , and $\rho_{y.zx}$ is the multiple correlation coefficient of y on x and z ($\rho_{y.zx} = \sqrt{\frac{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{1 - \rho_{xz}^2}}$).

SOME IMPROVED ESTIMATORS IN DOUBLE SAMPLING

It can be easily shown that the proposed improved class of estimators of Dash and Mishra [1] as given in (1.1) does not accommodate the estimator (1.5), as the known mean \bar{Z} does not contain in (1.1).

They were not able to provide a correct Taylor series expression of the estimator (1.1), which is a parametric function. Besides this, they failed to present an explicit estimator for population mean. In their article, Dash and Mishra [1] presented an incorrect expression of the minimum variance of Sahoo *et al.*'s [7] estimator [see the expression (2.2) in Dash and Mishra [1], pp. 4349 (which may be a typographical error)].

The correct minimum MSE expression should be

$$M(t_8)_{min} = S_y^2 \left[\left(\frac{1}{n} - \frac{1}{n'} \right) (1 - \rho_{yx}^2) + \left(\frac{1}{n'} - \frac{1}{N} \right) (1 - \rho_{yz}^2) \right] \quad (1.8)$$

Dash and Mishra [6] derived a wrong expression for minimum MSE of their estimator as

$$M(d_{dm})_{min} = S_y^2 \left[\left(\frac{1}{n} - \frac{1}{N} \right) (1 - \rho_{y.xz}^2) + \left(\frac{1}{n} - \frac{1}{N} \right) (1 - \rho_{xz}^2) B_{yx.z}^2 \frac{S_x^2}{S_y^2} \right] \quad (1.9)$$

It is true that estimator (1.7) is more efficient than (1.8), as we have

$$M(t_8)_{min} - M_{min} = S_y^2 \left(\frac{1}{n} - \frac{1}{n'} \right) (\rho_{y.xz}^2 - \rho_{yx}^2) \geq 0 \quad (1.10)$$

which is the same as [Dash and Mishra [6], eq. 3.1)] but expressions (1.9) and (1.7) are totally different.

3. Conclusion

This paper provides an overview of classes of estimators in Double sampling using two auxiliary variables. Some limitations and errors of an improved class of estimators recently suggested by Dash and Mishra [1] have been pointed out. The reviews suggest that although a lot of improved classes of estimators have been presented in recent time, they are equivalent or particular cases of the general class of estimators suggested by Tripathi and Ahmed [8] and Ahmed *et al.* [2]. Diana and Tommassic [18] argued that the regression type of estimators as given by Ahmed *et al.*[2] is one of the best estimators, at least at the first order of approximation. They further suggested that no new estimators be proposed, since no estimator can be more efficient than the best estimators proposed by Ahmed *et al.* [2], and some others, at least at the first order of approximation.

References

1. Dash, P.R. and Mishra, G. An improved class of estimators in double sampling using two auxiliary variables. *Commu. Stat.-Theo. Meth.*, 2011, **40(24)**, 4347-4352.
2. Ahmed, M.S., Rahman, M.S. and Ahmed, R. General class of chain estimators for a finite population mean using double sampling. *J. Appl. Statist. Sci.*, 1998, **7**, 185-190.
3. Chand, L. Some ratio-type estimators based on two or more auxiliary variables. Ph.D. thesis, Iowa State University, Ames, Iowa., 1995.
4. Kiregyera, B. A chain ratio-type estimator in finite population two phase sampling using two-auxiliary variables. *Metrika*, 1980, **27**, 217-223.
5. Kiregyera, B. Regression type estimators using two-auxiliary variables and the model of double sampling from finite populations. *Metrika*, 1984, **31**, 215-226.
6. Srivastava, S.Rani., Khare, B.B., and Srivastava, S.R. A generalized chain ratio estimator for mean of a finite population. *J. Ind. Sco. Ag. Stat.*, 1990, **42**, 108-117.
7. Sahoo, J., Sahoo, L.N. and Mohanty, S. A regression approach to estimation in two-phase sampling using two auxiliary variables. *Current Sci.*, 1993, **65(1)**, 73-75.
8. Tripathi, T.P. and Ahmed, M.S. A class of estimators for a finite population mean based on multivariate information and double sampling. *Cal. Stat. Asso. Bull.*, 1995, **45(179-180)**, 203-218.
9. Mishra, G. and Rout, K. A regression estimator in two phase sampling in presence of two auxiliary variables. *Metron*, 1997, **12**, 177-186.
10. Ahmed, M.S. 1997. The general class of chain estimators for the ratio of two means using double sampling. *Commu. Stat.-Theo. Meth.*, 1997, **269**, 2247-2254.
11. Mukerjee, R., Rao, T.J. and Vijayan, K. Regression-type estimators using multiple auxiliary information. *Aust. J. Stat.*, 1987, **29(3)**, 244-254.

12. Ahmed, M.S. Some recent works on two phase sampling for ratio and regression estimation. 12th Islamic Countries Conference on Statistical Sciences ICCS-12, December 19-22, 2012, Qatar University, Doha, Qatar.
 13. Upadhyaya, L.N. and Singh, G.N. Chain type estimators using transformed auxiliary variable in two-phase sampling. *Adv. Model. Anal.*, 2001, **38**, 1-2, 1-10.
 14. Samiuddin, M. and Hanif, M. Estimation of population mean in single and two phase sampling with or without additional information. *Pak. J. Stat.*, 2007, **232**, 99-118.
 15. Hanif, M., Ahmed, Z. and Ahmad, M. Generalized multivariate ratio estimators using multi-auxiliary variables for multi-phase sampling. *Pak. J. Stat.*, 2009, **254**, 615-629.
 16. Ahmed, M.S., Khan, S.U. and Tripathi, T.P. Two general classes of chain ratio and product estimators for a finite population mean based on Double sampling and multivariate information. *J. Stat. Studies.*, 1994, **14**, 86-99.
 17. Singh, V.K., Singh, Hari P., Singh, Housila P., and Shukla, D. A general class of chain estimators for ratio and product of two means of a finite population. *Commu. Stat.-Theo. Meth.*, 1994, **235**, 1341-1355.
 18. Diana, G. and Tommassic, C. Estimation for finite population mean in double sampling. Atti della XLI RIUNIONE SCIENTIFICA DELLA SIS. 599-602, Padova:CLEUP "Coop. Libreria Editrice Università di Padova, Milano, 5-7 June, 2002.
-

Received 14 August 2014

Accepted 8 December 2014