Liet. mat. rink. LMD darbai, **50**, 2009, 275–280

The generalization of ratio-cum product estimator for the arbitrary sample design

Ignas BARTKUS[1], Aleksandras PLIKUSAS[2]
[1]Vilnius Pedagogical University
Studentų g. 39, LT-08001 Vilnius
[2]Institute of Mathematics and Informatics
Akademijos 4, LT-08663 Vilnius
e-mail: ignas.bartkus@gmail.com; plikusas@ktl.mii.lt

Abstract. The generalization of a dual to ratio-cum-product estimator of the finite population total for the arbitrary sample design is given. The approximate variance is calculated and variance estimator is constructed. Some simulation results are presented for the comparison of ratio-cum-product estimator with some known estimators.

Keywords: finite population, ratio estimator, product estimator, dual variable, dual to ratio-cum-product estimator.

Introduction

Many estimators of the finite population parameters are constructed using known auxiliary variables. The classical well known ratio estimator is one of them. Various improvements of this ratio estimator have been considered by many authors. Some estimators use two auxiliary variables. Some other group of estimators are composite. They are constructed by taking a weighted sum of Horvitz-Thompson estimator and some ratio estimator. The sum of corresponding weights need not equal to one. The first of estimators considered in the paper is product, ratio-cum-product (Singh (1965)) and dual to ratio-cum-product estimators [5]. The product estimators are used in case negatively correlated auxiliary variable is available. This estimator behaves similarly as simple ratio estimator. In case two auxiliary variables are known, the ratio-cum-product estimator may be used. In the paper [5] the dual to ratio-cum-product estimator is considered for simple random sample without replacement. It seems this estimator is more effective compare to the ratio-cum-product estimator. In this paper the dual variables for the construction of dual to ratio-cum-product estimator, are defined for the case of stratified simple random sample and arbitrary sample design. The approximate variance of the dual to ratio-cum-product estimator is calculated and variance estimator is presented.

1. Main notation, ratio-cum-product estimator

Consider a finite population $\mathcal{U} = (u_1, u_2, \ldots, u_N)$ of N units. Assume three variables y, x, and z are defined on the population $\mathcal{U}$, taking values $\{y_1, \ldots, y_N\}$, $\{x_1, \ldots, x_N\}$, and $\{z_1, \ldots, z_N\}$. We consider y as unknown study variable, x and z by known auxiliary variables, possibly correlated with the variable y. We are interested in the estimation of the finite population total

$$t_y = \sum_{k=1}^{N} y_k \,.$$

Let simple random sample $s \subset \mathcal{U}$ of size n is drawn from the population $\mathcal{U}$. Denote

$$t_y = \sum_{k=1}^{N} y_k, \quad \mu_y = \frac{1}{N} t_y, \quad \bar{y} = \frac{1}{n} \sum_{k \in s} y_k, \quad \hat{t}_y = \frac{N}{n} \sum_{k \in s} y_k, = N\bar{y}.$$

The values $\hat{t}_x$, $\hat{t}_z$, μ_x, μ_z $\bar{x}$, $\bar{z}$, are defined analogously.

The ratio, product, and ratio-cum-product estimators are defined as follows

$$\hat{t}_R = \frac{\hat{t}_y}{\hat{t}_x} t_x = \frac{\bar{y}}{\bar{x}} t_x, \quad \hat{t}_P = \frac{\hat{t}_y \, \hat{t}_z}{t_z}, \quad \hat{t}_{RP} = \hat{t}_y \frac{\bar{z}}{\bar{x}} \frac{t_x}{t_z}.$$

2. Dual to ratio-cum-product estimator

The dual variable for the improvement of ratio-type estimator was used by Bandyopadhyay (1980) and Srivenkataramana (1980). An estimator with two auxiliary variables were considered by Singh *et al.* [5].

Consider again simple random sampling of size n and introduce the linear transformation of the variables x and z:

$$x_k^* = (1+g)\mu_x - g x_k, \quad z_k^* = (1+g)\mu_z - g z_k,$$

where $g = n/(N-n)$. Then

$$\bar{x}^* = (1+g)\mu_x - g\bar{x}, \quad \bar{z}^* = (1+g)\mu_z - g\bar{z}$$

are unbiased estimators for μ_x and μ_z. It is easy to see that $t_x^* = \sum_{k=1}^{N} x_k^* = t_x$ and $t_z^* = \sum_{k=1}^{N} z_k^* = t_z$. Correlation coefficient between variables y and x^*, $Corr(y, x^*) = -Corr(y, x) = -\rho_{yx}$ and $Corr(y, z^*) = -\rho_{yz}$. The dual to ratio-cum-product estimator suggested in the paper of Singh *et al.* [5] is

$$\hat{t}_{RP}^* = N\,\bar{y}\,\frac{\bar{x}^*}{\mu_x}\,\frac{\mu_z}{\bar{z}^*}.$$

2.1. Stratified simple random sample case

Assume the population $\mathcal{U}$ consists of H strata: $\mathcal{U} = \mathcal{U}_1 \cup \ldots \cup \mathcal{U}_H$. The size of stratum $\mathcal{U}_h$ is N_h, and the size of simple random sample s_h in stratum $\mathcal{U}_h$ is n_h, $h = 1, \ldots, H$. Denote $g_h = n_h/(N_h - n_h)$ for $h = 1, \ldots, H$, and define two transformations of the auxiliary variable x:

$$x_k^*(1) = A_x - g_h x_k, \quad \text{for } k \in \mathcal{U}_h,$$

$$x_k^*(2) = (1+g_h)\mu_{xh} - g_h x_k, \quad \text{for } k \in \mathcal{U}_h,$$

where

$$A_x = N^{-1}\sum_{h=1}^{H}(1+g_h)N_h\mu_{xh}, \quad \mu_{xh} = N_h^{-1}\sum_{k\in\mathcal{U}_h} x_k.$$

The transformations for the variable z are defined analogously. Note that $\sum_{k=1}^{N} x_k^*(j) = \sum_{k=1}^{N} x_k = t_x,\ j=1,2$. The relation $Corr(y,x^*) = -\rho_{xy}$ is not valid in the case of stratified sample. The covariances are given by

$$cov\big(y,x^*(1)\big) = -\frac{1}{N-1}\sum_{h=1}^{H}\big(N_h g_h \mu_{xh}(\mu_{hy}-\mu_y) + g_h(N_h-1)s_{xyh}\big),$$

$$cov\big(y,x^*(2)\big) = -\frac{1}{N-1}\sum_{h=1}^{H}\big(N_h(\mu_{xh}-\mu_x)(\mu_{yh}-\mu_y) - g_h(N_h-1)s_{xyh}\big),$$

where

$$s_{xyh} = \frac{1}{N_h-1}\sum_{k\in\mathcal{U}_h}(x_k-\mu_{xh})(y_k-\mu_{yh}).$$

The dual estimators for the both transformations coincide:

$$\hat{t}^*(1) = \sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} x_k^*(1) = \sum_{h=1}^{H}\frac{N_h}{n_h}\sum_{k\in s_h} x_k^*(2) = \hat{t}_x^*.$$

The dual to ratio-cum-product estimator be

$$\hat{t}_{RP}^* = \hat{t}_y\,\frac{\hat{t}_x^*}{t_x}\,\frac{t_z}{\hat{t}_z^*}.$$

The approximate variance of this estimator can be find by the usual linearization technique. We can expect this estimators be more efficient compare to the corresponding ratio-cum-product estimator.

In the case of the general unequal probability sampling design with the inclusion probability π_k of the element k, the dual variable is defined as

$$x_k^* = \left(\sum_{k\in s}\frac{1}{\pi_k}\right)^{-1}\sum_{k=1}^{N}(1+g_k)x_k - g_k x_k, \quad g_k = \frac{\pi_k}{1-\pi_k}.$$

Using the notation

$$\hat{t}_x^* = \sum_{k\in s}\frac{x_k^*}{\pi_k}, \quad \hat{t}_z^* = \sum_{k\in s}\frac{z_k^*}{\pi_k},$$

one can use the dual to ratio-cum-product estimator for general sample design

$$\hat{t}^*_{RP} = \hat{t}_y \frac{\hat{t}^*_x}{t_x} \frac{t_z}{\hat{t}^*_z}.$$

The approximate variance of the estimator $\hat{t}^*_{RP}$ can be derived using standard linearization technique

$$AVar(\hat{t}^*_{RP}) = \sum_{k,l \in U} (\pi_{kl} - \pi_k \pi_l) \left(\frac{y_k - R_x g_k x_k + R_z g_k z_k}{\pi_k} \right) \left(\frac{y_l - R_x g_l x_l + R_z g_l z_l}{\pi_l} \right).$$

Here π_{kl} is an inclusion probability of the pair of elements u_k, u_l into the sample s.

As the estimator of the variance of dual to ratio-cum-product estimator we can use the following estimator

$$\widehat{Var}\,(\hat{t}^*_{RP}) = \sum_{k,l \in s} \frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}} \left(\frac{y_k - \hat{R}_x g_k x_k + \hat{R}_z g_k z_k}{\pi_k} \right) \left(\frac{y_l - \hat{R}_x g_l x_l + \hat{R}_z g_l z_l}{\pi_l} \right).$$

Here

$$R_x = \frac{t_y}{t_x}, \quad R_z = \frac{t_y}{t_z}, \quad \hat{R}_x = \frac{\hat{t}_y}{\hat{t}_x}, \quad \hat{R}_z = \frac{\hat{t}_y}{\hat{t}_z}.$$

3. Simulation study

In this section some empirical study is presented to observe the behavior of the estimators in the case of stratified simple random sample design. A real populations from some Lithuanian Enterprise survey were used for the simulation. During the simulation study several populations were examined.

It should be noted that both auxiliary variables initially are positively correlated with the study variable. So, first of all we transform the variable z to dual, and consider the transformed variable as given negatively correlated auxiliary variable. Simulation results presented below show that ratio-cum-product estimator may perform efficiently.

Population **I**

y – an income of enterprise, x – number of employees, z – Wages-fund (dual variable). $N = 636$, $t_y = 119060206$, $t_x = 43785$, $t_z = 1827869$, $\rho_{yx} = 0.7795$, $\rho_{yz} = -0.9496$, $\rho_{zx} = -0.7964$.

Population **II**

y – Gross wage, x – average earnings, z – number of employees (dual variable). $N = 150$, $t_y = 6249836$, $t_x = 23199$, $t_z = 11719$, $\rho_{yx} = 0.8462$, $\rho_{yz} = -0.7001$, $\rho_{zx} = -0.5738$.

These populations are stratified into three and two strata respectively by the size of the variable x. In the first population 4000 samples and in the second population 2000 samples were drawn.

Table 1. Simulation results for the Population I

Estimator	Sample size n	Estimated relative bias $\times 10^3$	Average estimate of variance $\times 10^{-13}$	Approximate variance $\times 10^{-13}$	CV
$\hat{t}_{HT}$	100	-1.094	5.1596	5.1330	0.0602
	200	0.472	2.0387	2.0598	0.0381
	300	-0.242	1.0494	1.0639	0.0274
	400	-0.078	0.5517	0.5554	0.0198
$\hat{t}_R$	100	-0.962	4.1413	4.0891	0.0538
	200	-0.024	1.6558	1.6437	0.0341
	300	0.021	0.8313	0.8484	0.0245
	400	0.083	0.4370	0.4435	0.0177
$\hat{t}_{RP}$	100	-3.164	2.9601	2.8936	0.0453
	200	-1.914	1.1656	1.1637	0.0287
	300	-0.687	0.5994	0.6005	0.0206
	400	0.065	0.3101	0.3140	0.0149
$\hat{t}^*_{RP}$	100	-1.944	4.0056	3.9739	0.0531
	200	-0.857	1.1574	1.1568	0.0286
	300	-0.482	0.5488	0.5522	0.0197
	400	1.117	0.8245	0.8075	0.0238

Table 2. Simulation results for the Population II

Estimator	Sample size n	Estimated relative bias $\times 10^3$	Average estimate of variance $\times 10^{-13}$	Approximate variance $\times 10^{-13}$	CV
$\hat{t}_{HT}$	40	-0.092	1.4846	1.4625	0.0612
	50	1.805	1.0362	1.0741	0.0523
	60	1.168	0.8113	0.7815	0.0447
$\hat{t}_R$	40	0.549	1.0378	0.9861	0.0502
	50	1.810	0.7065	0.7304	0.0432
	60	1.294	0.5284	0.5167	0.0363
$\hat{t}_{RP}$	40	-0.192	0.9007	0.8497	0.0466
	50	-0.628	0.6022	0.6281	0.0401
	60	-1.174	0.4420	0.4474	0.0339
$\hat{t}^*_{RP}$	40	-1.611	0.8725	0.8482	0.0467
	50	-0.732	0.5193	0.5409	0.0372
	60	-1.531	0.3711	0.3732	0.0310

The estimated relative bias (the average bias divided by the true total t_y), average estimate of the variance, approximate true variance and coefficient of variation are calculated.

Tables 1 and 2 show that for stratified simple random sampling design dual ratio-cum-product estimator can be more efficient than other estimators considered.

For Population II (Table 2), dual ratio-cum-product estimator has a little bit bigger coefficient of variation then ratio-cum-product estimator.

References

1. S. Bandyopadhyay. Improved ratio and product estimators. *Sankhyā Series C*, 42(2):45–49, 1980.
2. S.K. Ray, R.K. Singh. Difference-cum-ratio type estimators. *Journal of Indian Statistical Association*, 19:147–151, 1981.
3. G.N. Singh. General class of ratio-type estimators under super-population model. *Biometrical Journal*, 42(3):363–375, 2000.
4. H.P. Singh, M.S. Kakran. A modified ratio estimator using known coefficient of kurtosis of an auxiliary character, 1993, unpublished.
5. H.P. Singh, R. Singh, M.R. Espejo, M.D. Pineda, S. Nadarajan. On the efficiency of the dual to ratio-cum-product estimator. *Mathematical Proceedings of the Royal Irish Academy*, 105A(2): 51–56, 2005.
6. T. Srivenkataramana. A dual to ratio estimator in sample surveys. *Biometrica*, 67:199–204, 1980.
7. L.N. Upadhyaya, H.P. Singh. Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41(5):627–636, 1999.
8. L.N. Upadhyaya, H.P. Singh, J.V.E. Vos. On the estimation of population means and ratios using supplementary information. *Statistica Neerlandica*, 39(3):309–318, 1985.

REZIUMĖ

I. Bartkus, A. Plikusas. Dualusis santykinis sandauginis įvertinys bet kuriam imties planui

Darbe apibrėžtas dualusis santykinis sandauginis baigtinės populiacijos sumos įvertinys bet kurio imties plano atveju. Apskaičiuota apytikslė šio įvertinio dispersija, pasiūlytas dispersijos įvertinys. Sukonstruoto įvertinio efektyvumas tikrinamas modeliuojant dviems skirtingoms populiacijoms. Modeliavimo rezultatai rodo, kad pasiūlytas įvertinys gali būti efektyvesnis už kitus žinomus įvertinius.

Raktiniai žodžiai: baigtinė populiacija, santykinis įvertinys, sandauginis įvertinys, dualusis kintamasis, dualusis santykinis sandauginis įvertinys.