Some overview of the ratio type estimators

Aleksandras Plikusas[1]

[1] Institute of Mathematics and Informatics, Lithuania
e-mail: plikusas@ktl.mii.lt

Abstract

An overview of some ratio type estimators of the finite population total is presented. The generalization of a dual to ratio-cum-product estimator for the arbitrary sample design is given.

1 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 (Singh et al (2005)). 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 Singh et al (2005) 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.

2 Main notation, ratio-cum-product estimator

Assume that simple random sample s of size n is drawn from the population $\mathcal{U}$. The ratio and product estimators are

$$\hat{t}_R = \frac{\bar{y}}{\bar{x}} t_x, \quad \hat{t}_P = \frac{\bar{y}\,\bar{z}}{t_z},$$

where

$$\bar{y} = \frac{1}{n}\sum_{k\in s} y_k, \quad \bar{x} = \frac{1}{n}\sum_{k\in s} x_k, \quad \bar{z} = \frac{1}{n}\sum_{k\in s} z_k, \quad t_x = \sum_{k=1}^{N} x_k, \quad t_z = \sum_{k=1}^{N} z_k.$$

The ratio-cum-product estimator is defined as

$$\hat{t}_{RP} = \hat{t}_y \frac{\bar{z}}{\bar{x}} \frac{t_x}{t_z}, \quad \hat{t}_y = \frac{N}{n}\sum_{k\in s} y_k.$$

The approximate variances of the estimators $\hat{t}_R$, $\hat{t}_P$, $\hat{t}_{RP}$ are

$$AVar(\hat{t}_R) = \Big(1 - \frac{n}{N}\Big)\frac{t_y^2}{n}\Big(c_y^2 + c_x^2\left(1 - 2\rho_{xy}\frac{c_y}{c_x}\right)\Big),$$

$$AVar(\hat{t}_P) = \Big(1 - \frac{n}{N}\Big)\frac{t_y^2}{n}\Big(c_y^2 + c_z^2\big(1 + 2\rho_{xz}\frac{c_y}{c_z}\big)\Big),$$

$$AVar(\hat{t}_{RP}) = \Big(1 - \frac{n}{N}\Big)\frac{t_y^2}{n}\Big(c_y^2 + c_z^2\big(1 - 2\rho_{yz}\frac{c_y}{c_z}\big) + c_x^2\Big(1 - 2\big(\rho_{xy}\frac{c_y}{c_x} + \rho_{yz}\frac{c_y}{c_z}\big)\Big)\Big),$$

where

$$c_y = \frac{s_y}{\mu_y}, \quad s_y^2 = \frac{1}{N-1}\sum_{k=1}^{N}(y_k - \mu_y)^2, \quad \mu_y = \frac{1}{N}\sum_{k=1}^{N} y_k,$$

$$\rho_{xy} = \frac{s_{xy}}{s_x s_y}, \quad s_{xy} = \frac{1}{N-1}\sum_{k=1}^{N}(y_k - \mu_y)(x_k - \mu_x),$$

c_x, c_z, ρ_{xz}, ρ_{yz}, μ_x, μ_z are defined analogously and respective to the subscripts used.

3 Dual to ratio-cum-product estimator estimator

3.1 Simple random sample case

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 (2005).

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 all (2005) is

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

The approximate variance of $\hat{t}_{RP}^*$ is

$$AVar(\hat{t}_{RP}^*) = \left(1 - \frac{n}{N}\right)\frac{t_y^2}{n}\big(c_y^2 + g c_z^2(g + 2K_{yz}) + g c_x^2(g - 2gK_{zx} - 2K_{yx})\big),$$

where

$$K_{yx} = \rho_{yx} c_y / c_x, \quad K_{zx} = \rho_{zx} c_z / c_x, \quad K_{yz} = \rho_{yz} c_y / c_z\,.$$

It is shown in Singh et al (2005) that for $1 - g > 0$ ($N > 2n$), $\hat{t}_{RP}^*$ is more efficient than Horvitz-Thompson estimator $\hat{t}_y$ and ratio-cum-product estimator $\hat{t}_{RP}$ when

$$\frac{1}{2}\,g < \frac{K_{yx}c_x^2 - K_{yz}c_z^2}{c_x^2 + c_z^2 - 2K_{zx}c_x^2} < \frac{1}{2}\,(1+g).$$

3.2 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:

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

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

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

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

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

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 is defined as

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

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

$$\hat{t}_{x\pi}^* = \sum_{k \in s} \frac{x_k^*}{\pi_k}, \quad \hat{t}_{z\pi}^* = \sum_{k \in s} \frac{z_k^*}{\pi_k},$$

the dual to ratio-cum-product estimator is defined as

$$\hat{t}_{RP\pi}^* = \hat{t}_y \, \frac{\hat{t}_{x\pi}^*}{t_x} \, \frac{t_z}{\hat{t}_{z\pi}^*}.$$

4 Some other ratio-type estimators

Below some ratio-type estimators are listed. Most of them are considered by different authors and only for simple random sample case.

Sisodia and Dwivedi (1981) suggested an estimator incorporating the coefficient of variation of the auxiliary variable x:

$$\hat{t}_{SD} = \hat{t}_y \frac{\mu_y + c_x}{\bar{x} + c_x}.$$

Singh and Kakran (1993) modified this estimator by replacing the coefficient of variation c_x by the population kurtosis of the variable x:

$$\hat{t}_{SK} = t_y \frac{m_x + \beta_x}{\bar{x} + \beta_x}, \quad \beta_x = \frac{m_4}{s_x^4} - 3, \quad m_4 = \frac{1}{N} \sum_{k=1}^{N} x_k^4.$$

Upadhyaya and Singh (1999) considered both coefficient of variation and kurtosis in their ratio-type estimators:

$$\hat{t}_{US}^{(1)} = \hat{t}_y \frac{\mu_x \, \beta_x + c_x}{\bar{x} \, \beta_x + c_x}, \quad \hat{t}_{US}^{(2)} = \hat{t}_y \frac{\mu_x \, c_x + \beta_x}{\bar{x} \, c_x + \beta_x}.$$

Ray and Singh (1981) introduced a class of ratio type estimators

$$\hat{t}_{RS} = \frac{\bar{y} + \hat{b}(\bar{x}^\alpha - \mu_x^\alpha)}{\bar{x}^\gamma} \, t_x^\gamma,$$

where

$$\hat{b} = \frac{\hat{s}_{xy}}{\hat{s}_x^2}, \quad \hat{s}_{xy} = \frac{1}{n-1} \sum_{k \in s} (x_k - \bar{x})(y_k - \bar{y}).$$

The special case of this estimator when $\alpha = \gamma = 1$ was analyzed by Kadilar and Cingi (2004).

Similar estimators were considered by Singh (2000):

$$\hat{t}_{rat} = w_1 \, \hat{t}_y + w_2 \frac{\bar{y}}{\bar{x}} \, t_x,$$

where the weights w_1 and w_2 are suitably chosen constants whose sum need not be unity.

Most of the listed estimators are considered for simple random sample. The efficiency of such ratio-type estimators for more general sample designs may be of interest.

References

Bandyopadhyay, S. (1980) Improved ratio and product estimators. *Sankhyā Series C*, **42**(2), 45-49.

Kadilar, C., Cingi, H. (2004) Ratio estimators in simple random sample. *Applied Mathematics and Computation*, **151**, 893-902.

Ray, S. K., Singh, R. K. (1981) Difference-cum-ratio type estimators. *Journal of Indian Statistical Association*, **19**, 147-151.

Singh, G. N. (2000) General class of ratio-type estimators under super-population model. *Biometrical Journal*, **42**(3), 363-375.

Singh, H. P., Kakran, M. S. (1993) A modified ratio estimator using known coefficient of kurtosis of an auxiliary character. *Unpublished*.

Singh, H. P., Singh, R., Espejo, M. R., Pineda, M. D., Nadarajan, S. (2005) On the efficiency of the dual to ratio-cum-product estimator. *Mathematical Proceedings of the Royal Irish Academy*, **105A**(2), 51-56.

Sisodia, B. V. S., Dwivedi, V. K. (1981), A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of Indian Society Agricultural Statistics*, **33**, 13-18.

Srivenkataramana, T. (1980) A dual to ratio estimator in sample surveys. *Biometrica*, **67**, 199-204.

Upadhyaya, L. N., Singh, H. P. (1999) Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, **41**(5), 627-636.

Upadhyaya, L. N., Singh, H. P., Vos, J. V. E. (1985) On the estimation of population means and ratios using supplementary information. *Statistica Neerlandica*, **39**(3), 309-318.