VARIANCE ESTIMATION USING QUADRATIC PROGRAMMING TECHNIQUE

Raju Kumar*, Kaustav Aditya*, Hukum Chandra*, Pradeep Basak* and Shrila Das**
*ICAR-Indian Agricultural Stattistics Research Institute, New Delhi-12
** ICAR-Indian Agricultural Research Institute, New Delhi-12

1. Introduction

The calibration approach is frequently used to develop estimators of various population parameters of interest and very frequently used by the statisticians for generation of official and Agricultural statistics in India because it gives precise estimates along with the Standard error which is needed for policy planning. The calibration approach involves deriving revised weights and using these weights in the form of a weighted estimator. The revised weights are developed by minimizing a distance function subject to the constraint that the weighted sample mean of auxiliary variables is equal to the population mean/total of the auxiliary variable. Generally, a chi-square type distance function is used for deriving revised weights. The approach is called as lower level calibration. A higher order calibration approach has been developed by Singh (1998) to improve the estimate of variance of the Horvitz-Thompson (1952) estimator. The higher level calibration approach makes use of known population variance of the auxiliary character. Many a times the chi-square type distance function gives negative weights. The negative weights when used in the variance estimator will give negative variance estimator. We use quadratic programming approach to obtain non-negative weights.

2. Calibration Estimation

The most commonly used estimator of population total /mean is the generalized linear regression (GREG) estimator. Let us the simplest case of the GREG where information on only one auxiliary variable is available. Suppose a population $U=\{1,2,...,N\}$ is the set of labels for the infinite population. Let y_i be the value of the variable of interest, y, for the i^{th} population element, with which also is associated an auxiliary variable x_i. The objective is to estimate the population total $Y=\sum_{i=1}^{N} y_i$ effectively using the known population totals $X=\sum_{i=1}^{N} x_i$. Let s= {1, 2, . . . , n}be the set of sampled units under a general sampling, p, $\pi_i = p_r(i \in s)$ and $\pi_{ij} = p_r(i \text{ and } j \in s)$ be the first order inclusion probabilities and second order inclusion probabilities respectively. Then the well known Horvitz and Thompson (1952) estimator of population total, $\hat{Y}_{HT} = \sum_{i=1}^{n} d_i y_i$ is define by Deville and Sӓrndal (1992) using calibration approach as $\hat{Y}_{DS} = \sum_{i=1}^{n} w_i y_i$, where calibrated weights w_i*'s* are modified from the basic design weights $d_i = 1/\pi_i$ by minimizing a distance function measure between the w_i*'s* and d_i's subject to constraints $\sum_{i=1}^{n} w_i x_i = X$. The most commonly used distance measure is the chi-square distance $\phi = \sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i q_i}$ where q_i are suitably chosen weights. In most of the situations, the value $q_i = 1$ is taken. The form of the estimator is depends upon the choice of $q_{i.}$ By minimizing the $\sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i q_i}$ subject to calibration equation $\sum_{i=1}^{n} w_i x_i = X$, the weights *wᵢ's* is obtained as $w_i = d_i + \frac{d_i q_i x_i}{\sum_{i=1}^{n} d_i q_i x_i^{2}}\left(X - \sum_{i=1}^{n} d_i x_i \right)$ **.** Substitution of the value of w_i in estimator results in $\hat{Y}_{DS} = \sum_{i=1}^{n} w_i y_i$ give $\hat{Y}_{DS} = \sum_{i=1}^{n} d_i y_i + \frac{\sum_{i=1}^{n} d_i Q_i x_i y_i}{\sum_{i=1}^{n} d_i Q_i x_i^{2}}\left(X - \sum_{i=1}^{n} d_i x_i \right)$.

3. Variance Estimation

For improved variance estimators in survey sampling, the higher order calibration approach makes use of known variance or second order moment of the estimator of the population total of auxiliary character whereas low level calibration utilizes only know total of auxiliary character. In the case of single auxiliary variables, Singh *et al.* (1998) proposed a higher level calibration approach for variance estimation.

The Yates and Grundy form of variance of the Horvitz-Thompson (1952) estimator is given as

$$V_{YG}\left(\hat{Y}_{HT}\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^{2}.$$

An unbiased estimate of variance of $\hat{Y}_{HT}$ is

$$\hat{V}_{YG}\left(\hat{Y}_{HT}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left(\pi_i \pi_j - \pi_{ij}\right)}{\pi_{ij}}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^{2}.$$

Singh, Horn and Yu (1998) proposed an estimator of variance of $\hat{Y}_{HT}$ is

$\hat{V}_{ss}\left(\hat{Y}_{HT}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Omega_{ij}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^{2}$ where Ω_{ij} are modified from $D_{ij} = \left(\pi_i \pi_j - \pi_{ij}\right) / \pi_{ij}$ by minimizing the chi-square type distance $D = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left(\Omega_{ij} - D_{ij}\right)^{2}}{D_{ij} Q_{ij}}$ subject to the constraint

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Omega_{ij}\left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)^{2} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j}\right)^{2}.$$

Then modified optimal weights is

$$\Omega_{ij} = D_{ij} + \frac{D_{ij}\Omega_{ij}(d_i x_i - d_j x_j)^{2}}{\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} D_{ij}\Omega_{ij}(d_i x_i - d_j x_j)^{4}}\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\Omega_{ij}(d_i x_i - d_j x_j)^{2} - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} D_{ij}(d_i x_i - d_j x_j)^{2}\right].$$

A problem with chi-square type distance function is that the revised weights may have negative values. The negative weights used in $\hat{V}_{ss}\left(\hat{Y}_{HT}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Omega_{ij}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^{2}$ will produce negative estimate of variance. One of the ways to tackle the problem of negative weights is to consider some other suitable distance function. Another way this problem can be tackled is by using the quadratic programming. approach.

4. Results and Discussion

We demonstrate the use of the quadratic programming approach with the help of a data. For illustration purpose we have used the data given in Horvitz and Thompson (1952). The data pertains to *20* population units. The study variable, y, is the number of households on *i*-th block and known auxiliary character, x, is the eye-estimated number of households in the *i*-th block. All possible sample of size *n=5* were selected by SRSWOR, which resulted in ${}^{N}C_{n} = 15,504$ samples. For each sample calibrated weights were determined by using the formula.

$$\Omega_{ij} = D_{ij} + \frac{D_{ij}\Omega_{ij}(d_i x_i - d_j x_j)^2}{\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} D_{ij}\Omega_{ij}(d_i x_i - d_j x_j)^4} \left[\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\Omega_{ij}(d_i x_i - d_j x_j)^2 - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} D_{ij}(d_i x_i - d_j x_j)^2\right].$$

The data is given in Table 4.1.2.1 below:

Table-4.1.2.1

block	Number of household on *i*th block	Eye estimated number of house-holds on *i*th block
1	19	18
2	9	19
3	17	14
4	14	12
5	21	24
6	22	25
7	27	23
8	35	24
9	20	17
10	15	14
11	18	18
12	37	40
13	12	12
14	47	30
15	27	27
16	25	26
17	25	21
18	13	9
19	19	19
20	12	12

Approximately in 4.48% cases negative weights were obtained. These weights are determined using the quadratic programming approach. The quadratic programming approach is implemented through the SAS package. In all the cases i.e. 15,504 the weights were found to be positive.

5. Conclusion

It is possible to obtain non-negative calibrated weights through the use of quadratic programming approach for minimization of the chi-square type distance function. The recomputed weights using the quadratic programming approach were found positive.

References (if any)

1. Bartkus, I. and Pumputis, D. (2010). Estimation of finite population ratio using calibration of strata weight. *Workshop on Survey Sampling Theory and Methodology*, Lithuania. 23-27, 2010.
2. Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**, 376-382.
3. Estevao, V. M. and Särndal, C.E. (2002). The ten cases of auxiliary information for calibration in two phase sampling. *Journal of Official Statistics*, **18**, 233-255.
4. Krapavickaite, D. and Plikusas, A. (2005) Estimation of a Ratio in the Finite Population. *Informatica*, **16**, 347-364.
5. Plikusas, A. and Pumputis, D. (2007). Calibrated estimators of the population covariance. *Acta Appllication Mathematicae*, **97**, 177-187.
6. Plikusas, A. (2001). Calibrated estimators of the ratio. *Lithuanian Mathematical Journal*, (special issue) **41**, 457-462.
7. Plikusas, A. (2003) Calibrated weights for the estimators of the ratio. *Lithuanian Mathematica Journal*, **43**, 543-547.
8. Singh, S., Horn, S. and Yu, F. (1998). Estimation of variance of the general regression estimator: Higher level calibration approach. *Survey Methodology*, **24**, 41-50.
9. Singh, S., Horn, S., Choudhury, S. and Yu, F. (1999). Calibration of the estimators of variance, *Australian and New Zealand Journal of Statistics*, **41**(2), 199-212.
10. Sitter, R. R. and Wu, C. (2002). Efficient estimation of quadratic finite population functions. *Journal of the American Statistical Association*, **97**, 535-543.
11. Tracy, D. S., Singh, S. and Arnab, R. (2003). Note on calibration in stratified and double sampling. *Survey Methodology*, **29**, 99–104.

Terms - Do not remove or change this section (It should be emailed back to us as is)

• This form is for genuine submissions related to biotechnology topics only.
• You should be the legal owner and author of this article and all its contents.
• If we find that your article is already present online or even containing sections of copied content then we treat as duplicate content - such submissions are quietly rejected.
• If your article is not published within 3-4 days of emailing, then we have not accepted your submission. Our decision is final therefore do not email us enquiring why your article was not published. We will not reply. We reserve all rights on this website.
• Do not violate copyright of others, you will be solely responsible if anyone raises a dispute regarding it.
• Similar to paper based magazines, we do not allow editing of articles once they are published. Therefore please revise and re-revise your article before sending it to us.
• Too short and too long articles are not accepted. Your article must be between 500 and 5000 words.
• We do not charge or pay for any submissions. We do not publish marketing only articles or inappropriate submissions.
• Full submission guidelines are located here: http://www.biotecharticles.com/submitguide.php
• Full Website terms of service are located here: http://www.biotecharticles.com/privacy.php

As I send my article to be published on BiotechArticles.com, I fully agree to all these terms and conditions.