ESTIMATION UNDER BALANCED SYSTEMATIC SAMPLING FOR PARABOLIC POPULATIONS

*S. Sampath & M.A. Basker**

ABSTRACT

[Construction of estimators of finite population total for populations exhibiting known trend has attracted the attention of many researchers. In this paper, two estimators for total of finite populations exhibiting parabolic trend are proposed under balanced systematic sampling and their performances have been assessed with the help of appropriate super population models.

Key words: finite population, systematic sampling, super population models, average variances]

1. Introduction

The problem of estimating a finite population total has received considerable attention for nearly six decades. One of the oldest strategies used for estimating a finite population is simple random sampling – expansion estimator. Pointing out the fact that this strategy fails to take care of inherent characteristics of the given population, sample survey researchers now and then propose various sampling-estimating

---

* Loyola College, Chennai - 600034.

strategies and study their properties. While proposing new strategies survey researchers attempt to use the knowledge of some supplementary variable which is closely related to the variable under study either in the sampling process or in the estimation stage or in both stages. While this has been the direction of study pursued by some researchers, another group works towards developing strategies with increased precision by making use of information regarding labels of population units. The second line of approach has been pursued by researchers who propose systematic sampling schemes in various forms. The oldest method of systematic sampling is Linear Systematic Sampling (LSS) due to Yates(1948). The LSS is shown to perform better than Simple Random Sampling (SRS) particularly when variability with in systematic groups is more significant when compared to overall variation present in the population with respect to the variable under study(Cochran(1977)). In addition to systematic sampling schemes Sampath(1991),Suresh Chandra and Sampath(1992), Suresh Chandra, Sampath and Balasubramani (1992), Sampath and Uthayakumaran(1996,1998) have proposed a variety of sampling designs which use label information and made in depth studies on them. Apart from this, it has been proved that when population units exhibit linear trend, LSS is more efficient than SRS. Sensing that further improvement is possible for populations exhibiting linear trend, researchers started working vigorously and started coming out with many strategies wherein the estimated total coincides with population total making the mean square error zero. To name some of them, Madow(1953), Sethi(1965), Singh, Jindal and Garg(1968). Again here it has to be mentioned that there are two approaches in this direction of getting error free estimators in the presence of linear trend. While in one approach new sampling methods have been proposed, the second approach concentrates on using an estimator different from expansion estimator to estimate the total with out error. The first in this direction is the estimator due to Yates (1948) which estimates the population total in the presence of linear trend under linear systematic sampling. A similar corrected estimator has been constructed under circular systematic sampling (Cochran(1977)). Apart from populations with linear trend there are cases where a survey researcher will encounter populations with parabolic trend. Examples of such populations are discussed in Bellhouse(1981). For populations exhibiting parabolic trend similar corrected estimators have been developed by Agrawal and Jain (1988) and Sampath and Suresh Chandra(1991) under linear systematic sampling. In this paper, corrected estimators for the population total in the presence of parabolic trend is developed corresponding to Balanced Systematic Sampling(BSS).

2. Estimators under Balanced Systematic Sampling

Under balanced systematic sampling the population units are divided into $\frac{n}{2}$ groups (assuming the sample size n is even) of $2k$ units each and a pair of units equidistant from the ends are selected from each group in a systematic manner. The balanced systematic sampling can be mathematically described as follows:

Draw a random number i at random from 1 to k. The units corresponding to the numbers $\{i| + 2jk, 2(j+1)k - i + 1\}, j = 0, 1, 2, ..., \frac{n}{2} - 1\}$ will be the sample corresponding to the random start i when n is even. When the sample size n is odd, the balanced systematic sample of size n corresponding to the random start i is given by the units with labels $\{i + 2jk, 2(j+1)k - i + 1, i + (n-1)k\}, j = 0, 1, 2, ..., \frac{n-3}{2}\}$. Essentially, balanced systematic sampling is meant for estimating a finite population total without error in the presence of linear trend. Hence when population values fail to exhibit linear trend, the performance of balanced systematic sampling is not likely to be the best. In this paper we shall develop an estimator for the population total under balanced systematic sampling which estimates the population total without error in the presence of parabolic trend. In our discussion we assume that the sample is even. When the sample size is even two cases arise. That is, $\frac{n}{2}$ can be either odd or even. We treat these two cases independently. First let us consider the case where $\frac{n}{2}$ is odd. It may be noted that, balanced systematic sampling can be viewed as dividing the population into $\frac{n}{2}$ groups and choosing pairs of units from each group in a systematic and balanced manner. We give weight λ_1 to the first (left unit in the pair) unit in the pair coming from group 1, λ_2 to the pair of units coming from the middlemost (which happens to be unique when $\frac{n}{2}$ is odd) group and λ_3 to the second (right unit in the pair) unit in the pair coming from the last group namely $\frac{n}{2}$th group leading to the estimator

$$\hat{Y}_{BSS}^{(1)} = \frac{N}{n}\left[\lambda_1 Y_r + \lambda_2 \left\{Y_{r + \frac{(n-2)k}{2}} + Y_{\frac{(n-2)k}{2} - r + 1}\right\} + \lambda_3 Y_{N-r+1} + Y_{2k-r+1} + Y_{r+(n-2)k}\right.$$

$$+ \sum_{i=2}^{\frac{n-2}{4}} \{Y_{r+2(i-1)k} + Y_{2ik-r+1}\} + \sum_{i=\frac{n+6}{4}}^{\frac{n-2}{2}} \{Y_{r+2(i-1)k} + Y_{2ik-r+1}\}] \tag{1}$$

for the population total. The weights λ_1, λ_2 and λ_3 are determined so that the estimator coincides with the population total when the population values are modeled by

$$Y_i = \alpha + \beta i + \gamma i^2, i = 1, 2, \ldots, N. \tag{2}$$

Substituting the values given under (2) in (1) we get,

$$\frac{N}{n}\left[\lambda_1(\alpha + \beta r + \gamma r^2) + \lambda_2\left(\alpha + \beta\left[r + \frac{(n-2)k}{2} \right] + \gamma\left[r + \frac{(n-2)k}{2} \right]^2 \right) \right.$$

$$+ \lambda_2\left(\alpha + \beta\left[\frac{(n+2)k}{2} - r + 1 \right] + \gamma\left[\frac{(n+2)k}{2} - r + 1 \right]^2 \right)$$

$$+ \lambda_3\left(\alpha + \beta(nk - r + 1) + \gamma(nk - r + 1)^2 \right)$$

$$+ \left(\alpha + \beta(2k - r + 1) + \gamma(2k - r + 1)^2 \right) + \left(\alpha + \beta[r + (n-2)k] + \gamma[r + (n-2)k]^2 \right)$$

$$+ \sum_{i=2}^{\frac{n-2}{4}}\left[\alpha + \beta(r + 2(i-1)k) + \gamma(r + 2(i-1)k)^2 + \alpha + \beta(2ik - r + 1) + \gamma(2ik - r + 1)^2 \right]$$

$$\left. + \sum_{i=\frac{n+6}{4}}^{\frac{n-2}{2}}\left[\alpha + \beta(r + 2(i-1)k) + \gamma(r + 2(i-1)k)^2 + \alpha + \beta(2ik - r + 1) + \gamma(2ik - r + 1)^2 \right] \right]$$

Equating the population total under model (2) with the above expression and comparing the coefficients of α, β and γ we get

$$\lambda_1 + 2\lambda_2 + \lambda_3 = 4$$

$$\lambda_1 r + \lambda_2(nk + 1) + \lambda_3(nk - r + 1) = 2(nk + 1)$$

$$\lambda_1 r^2 + \lambda_2\left[\left(r + \frac{n-2}{2}k \right)^2 + \left(\frac{nk}{2} - r + k + 1 \right)^2 \right] + \lambda_3(nk - r + 1)^2$$

68

$$= \frac{n}{6} + n^2 k \left(\frac{3}{2}k\right) + nk\left(2 - \frac{2k}{3}\right) + \left[r^2 + (1-r)^2\right]\left(2 - \frac{n}{2}\right) + 2k(k - 2r + 1)$$

Solving these three equations, the values of λ_1, λ_2 and λ_3 are obtained as follows.

$$\lambda_1^{(1)} = \frac{\delta(r) - 2\mu(r)}{\chi(r) - \mu(r)} \tag{3}$$

$$\lambda_2^{(1)} = \frac{2\chi(r) - (r)}{\chi(r) - \mu(r)} \tag{4}$$

$$\lambda_3^{(1)} = \frac{\delta(r) - 2\mu(r)}{\chi(r) - \mu(r)} \tag{5}$$

where
$$\mu(r) = \left(r + \frac{n-2}{2}k\right)^2 + \left(\frac{nk}{2} - r + k + 1\right)^2,$$

$$\delta(r) = \frac{n}{6} + n^2 k\left(\frac{3}{2}k\right) + nk\left(2 - \frac{2k}{3}\right) + \left[r^2 + (1-r)^2\right]\left(2 - \frac{n}{2}\right) + 2k(k - 2r + 1)$$

and $\chi(r) = (nk - r + 1)^2 + r^2$.

On using these weights in the corrected estimator given in (1) one can make the estimated value coincide with the population total in the presence of parabolic trend.

Now let us turn our attention to the case where $\frac{n}{2}$ is even. In this case the number of groups under balanced systematic sampling will be even and hence there will be two middle most groups. We give weight λ_1 for the first unit in the pair coming from the first group; the weight λ_2 is given to the second unit of the pair coming from the first middle group ($\frac{n}{2}$th group) and the first unit coming from the second middle most group($\frac{n}{2}$ + 1st group) and the weight λ_3 is given to the second unit in the pair coming from the last group. Thus we have the corrected estimator

$$\hat{Y}_{BSS}^{(2)} = \frac{N}{n}\left[\lambda_1 Y_r + \lambda_2\left\{ Y_{\frac{nk}{2}-r+1} + Y_{\frac{nk}{2}+r} \right\} + \lambda_3 Y_{N-r+1} + Y_{\frac{(n-4)k}{2}+r} + Y_{\frac{(n+4)k}{2}-r+1} \right.$$

$$\left. + Y_{2k-r+1} + Y_{(n-2)k+r} + \sum_{j=2}^{\frac{n-4}{4}}\left\{ Y_{r+2(j-1)k} + Y_{2jk-r+1} \right\} + \sum_{j=\frac{n+8}{4}}^{\frac{n-2}{2}}\left\{ Y_{r+2(j-1)k} + Y_{2jk-r+1} \right\} \right] \quad (6)$$

Proceeding as in the previous case we form the following three equations

$$\lambda_1 + 2\lambda_2 + \lambda_3 = 4$$

$$\lambda_1 r + \lambda_2(nk+1) + \lambda_3(nk-r+1) = 2(nk+1)$$

$$\lambda_1 r^2 + \lambda_2\left[\left(\frac{nk}{2}-r+1\right)^2 + \left(\frac{nk}{2}+r\right)^2 \right] + \lambda_3(nk-r+1)^2$$

$$= \frac{n}{6} + n^2 k\left(\frac{3}{2}k\right) + nk\left(2-\frac{2k}{3}\right) + \left[r^2 + (1-r)^2\left(2-\frac{n}{2}\right)\right]$$

Solving these equations, we get λ_1, λ_2 and λ_3 as

$$\lambda_1^{(2)} = \frac{\delta(r) - 2\mu(r)}{\chi(r) - \mu(r)} \quad (7)$$

$$\lambda_2^{(2)} = \frac{2\chi(r) - \delta(r)}{\chi(r) - \mu(r)} \quad (8)$$

$$\lambda_3^{(2)} = \frac{\delta(r) - 2\mu(r)}{\chi(r) - \mu(r)} \quad (9)$$

where $\mu(r) = \left(r+\frac{nk}{2}\right)^2 + \left(\frac{nk}{2}-r+1\right)^2$,

$$\delta(r) = \frac{n}{6} + n^2 k\left(\frac{3}{2}k\right) + nk\left(2-\frac{2k}{3}\right) + \left[r^2 + (1-r)^2\left(2-\frac{n}{2}\right)\right]$$

and $\chi(r) = (nk-r+1)^2 + r^2$.

From the above discussion, it is inferred that the estimator coincides the population total in the presence of parabolic trend when the weights derived above are used in the corrected estimator given in (6). The weights used in the estimator do not require the knowledge of the parameters α, β and γ involved in this model.

70

3. Average Variances

In sample surveys, values assumed by the survey variables over various population units are assumed to be constants. The primary aim of survey statistician is to develop a sampling estimating strategy which minimizes the variance or mean squared error of an estimator, which is proposed to estimate unknown parametric values. Under the super-population approach this problem is being studied in an alternate angle. The values taken by the population units with respect to the survey variables are assumed to be the realized values of random variables having a known probability distribution which involves both known as well as unknown parametric values. This approach is called the super- population approach.

Bellhouse and Rao(1975) have compared the efficiencies of different strategies with the help of super-population models. Later Agarwal and Jain (1988) also have done studies similar to Bellhouse and Rao(1975) in general super-population models.

To asses the performances of strategies developed in the previous section, we shall compare them with the help of super – population model namely

$$Y_i = \alpha + \beta i + \gamma i^2 + e_i$$

where $E(e_i) = 0,\ V(e_i) = \sigma^2 i^g$ and $Cov(e_i, e_j) = 0, i \ne j = 1,2,..,N$

where g is a predetermined constant. The average mean square error of estimators corresponding to balanced systematic sampling are derived and the corresponding expressions are given in the following theorems.

Theorem 1

The average mean square error of the estimator $\hat{Y}_{BSS}^{(1)}$ under the above super population model is

$$\frac{\sigma^2}{k} \sum_{r=1}^{k} \left\{ k^2 \left(\lambda_1^2 r^g + \lambda_2^2 \left[\left(\frac{n-2}{2}k + r \right)^g + \left(\frac{n+2}{2}k - r + 1 \right)^g \right] + \lambda_3^2 (nk - r + 1)^g \right. \right.$$

$$\left. + (2k - r + 1)^g + ((n-2)k + r)^g + \sum_{j=2}^{\frac{n-2}{4}} \left[(2(j-1)k + r)^g + (2jk - r + 1)^g \right] \right]$$

$$+ \sum_{i=\frac{n+6}{4}}^{\frac{n-2}{2}} \left[(2(i-1)k+r)^g + (2ik-r+1)^g \right] \Bigg) + \sum_{j=1}^{N} j^g$$

$$- 2k \left(\lambda_1 r^g + \lambda_2 \left[\left(\frac{n-2}{2}k+r \right)^g + \left(\frac{n+2}{2}k-r+1 \right)^g \right] + \lambda_3 (nk-r+1)^g \right.$$

$$+ (2k-r+1)^g + ((n-2)k+r)^g + \sum_{i=2}^{\frac{n-2}{4}} \left[(2(i-1)k+r)^g + (2ik-r+1)^g \right]$$

$$+ \left. \sum_{i=\frac{n+6}{4}}^{\frac{n-2}{2}} \left[(2(i-1)k+r)^g + (2ik-r+1)^g \right] \right) \Bigg\}$$

Theorem 2

The average variance of the estimator $\hat{Y}_{BSS}^{(2)}$ is

$$\frac{\sigma^2}{k} \sum_{r=1}^{k} \left\{ k^2 \left(\lambda_1^2 r^g + \lambda_2^2 \left[\left(\frac{nk}{2}-r+1 \right)^g + \left(\frac{nk}{2}+r \right)^g \right] + \lambda_3^2 (nk-r+1)^g + \left[\frac{(n-4)k}{2}+r \right]^g \right.$$

$$+ \left[\frac{(n+4)k}{2}-r+1 \right]^g + (2k-r+1)^g + ((n-2)k+r)^g$$

$$+ \sum_{i=2}^{\frac{n-4}{4}} \left[(2(i-1)k+r)^g + (2ik-r+1)^g \right] + \sum_{i=\frac{n+8}{4}}^{\frac{n-2}{2}} \left[(2(i-1)k+r)^g + (2ik-r+1)^g \right] \right)$$

$$+ \sum_{j=1}^{N} j^g - 2k \left(\lambda_1 r^g + \lambda_2 \left[\left(\frac{nk}{2}-r+1 \right)^g + \left(\frac{nk}{2}+r \right)^g \right] + \lambda_3 (nk-r+1)^g \right.$$

$$+ \left[\frac{(n-4)k}{2}+r \right]^g + \left[\frac{(n+4)k}{2}-r+1 \right]^g + (2k-r+1)^g + ((n-2)k+r)^g$$

$$+ \sum_{i=2}^{\frac{n-4}{4}} \left[(2(i-1)k+r)^g + (2ik-r+1)^g \right] + \sum_{i=\frac{n+8}{4}}^{\frac{n-2}{2}} \left[(2(i-1)k+r)^g + (2ik-r+1)^g \right] \right) \Bigg\}$$

72

4. Comparitive study

In view of the difficulties involved in comparing the above average mean square errors theoretically, because of their complex nature, a computer based comparative study is carried out to asses their performance. The empirical comparisons have been made for various choices of N, n, k and g. The following tables gives the relative efficiencies of the corrected estimators under balanced systematic sampling over the Yates corrected estimator corrected under linear systematic sampling given in Sampath and Suresh Chandra(1991) for parabolic populations.

Percentage efficiencies of BSS Over LSS

	k=5					k=8			
n	g=0	g=1	g=2	g=3	n	g=0	g=1	g=2	g=3
12	101.919	101.913	102.603	103.273	12	101.802	101.796	102.447	103.072
14	101.619	101.616	102.231	102.835	14	101.519	101.516	102.097	102.661
16	101.398	101.396	101.949	102.495	16	101.312	101.31	101.831	102.342
18	101.229	101.228	101.729	102.225	18	101.153	101.152	101.624	102.089
20	101.096	101.096	101.553	102.007	20	101.028	101.028	101.458	101.884
22	100.989	100.989	101.408	101.826	22	100.928	100.927	101.323	101.714
24	100.901	100.901	101.289	101.676	24	100.845	100.845	101.21	101.573
26	100.827	100.827	101.187	101.547	26	100.775	100.775	101.115	101.452
28	100.764	100.764	101.101	101.437	28	100.717	100.716	101.033	101.349
30	100.71	100.71	101.026	101.342	30	100.666	100.666	100.963	101.259
32	100.663	100.663	100.961	101.258	32	100.622	100.622	100.902	101.18
34	100.622	100.622	100.903	101.184	34	100.583	100.583	100.847	101.111
36	100.586	100.586	100.852	101.118	36	100.549	100.549	100.799	101.049

	k=10					k=12			
n	g=0	g=1	g=2	g=3	n	g=0	g=1	g=2	g=3
12	101.762	101.757	102.395	103.005	12	101.736	101.73	102.36	102.961
14	101.486	101.483	102.052	102.603	14	101.464	101.461	102.022	102.565
16	101.283	101.281	101.792	102.291	16	101.263	101.262	101.766	102.257
18	101.128	101.127	101.589	102.043	18	101.111	101.11	101.565	102.013
20	101.006	101.005	101.427	101.843	20	100.99	100.99	101.405	101.815
22	100.907	100.907	101.294	101.677	22	100.893	100.893	101.275	101.652
24	100.826	100.826	101.184	101.538	24	100.814	100.813	101.166	101.515
26	100.758	100.758	101.09	101.42	26	100.747	100.747	101.074	101.399
28	100.701	100.701	101.011	101.319	28	100.69	100.69	100.996	101.3
30	100.651	100.651	100.942	101.231	30	100.641	100.641	100.928	101.213
32	100.608	100.608	100.882	101.154	32	100.599	100.599	100.869	101.137
34	100.57	100.57	100.829	101.086	34	100.562	100.562	100.816	101.07
36	100.537	100.537	100.782	101.026	36	100.529	100.529	100.77	101.011

5. Conclusion

A careful analysis of the results obtained from the computer study clearly indicates that the suggested Yates corrected estimator under Balanced systematic sampling estimator is more efficient than the corresponding estimators. The efficiency increases as g increases.

Hence we conclude that the corrected estimators developed for balanced systematic sampling in this paper are more efficient than the Yates corrected estimator developed by Sampath and Suresh Chandra(1991) for populations exhibiting parabolic trend.

6. References

1. Agrawal,M.C. and Jain,N. (1988) : Comparision of some sampling strategies in the presence of trend, Jour. Ind. Soc. Ag. Statistics, 40, 191-201

2. Bellhouse, D.R. (1981) : Spatial sampling in the presence of a trend. Jour. of Stat. Plann. and Inf., 5, 365 – 375.

3. Bellhouse,D.R. and Rao,J,N,K. (1975): Systematic sampling in the presence of trend, Biometrika, 62, 694-697

4. Cochran,W.G. (1977) : Sampling Techniques, Wiley Eastern Limited, New Delhi

5. Madow,W.G. (1953): 'On the theory of systematic sampling III', Ann. Math. Stat., 24, 101-106

6. Sampath,S. and Suresh Chandra,K.(1991): Estimation of population mean in the presence of parabolic trend, Jour. Ind. Soc. of Agri. Stat., 43, 321-328.

7. Sampath,S.(1991): Some contributions to inference for finite populations, unpublished Ph.D. thesis, submitted to University of Madras, Chennai, India

8. Sampath,S. and Uthayakumaran,N.(1996): Markov systematic sampling using auxiliary information, Statistica, LVI, 439-443

9. Sampath,S. and Uthayakumaran,N.(1998): Markov systematic sampling, Biometrical Journal, 40, 883-895

10. Sethi,V.K.(1965): On optimum pairing of units, Sankhya, Ser B, 27, 315-320

11. Singh,P., Jindal,K.K. and Garg,J.N.(1968) : On modified systematic sampling, Biometrika, 55, 541-546

12. Suresh Chandra,K. and Sampath,S.(1992): Markov sampling with auxiliary information, Statistical Papers, 33, 83-91

13. Suresh Chandra,K., Sampath,S. and Balasubramani,G,K, (1992): Markov sampling for finite populations, Biometrika, 79, 210-213

14. Yates, F.(1948): Systematic sampling, Phil. Trans. Roy. Soc., London, A241. 345-377