Jour. Ind. Soc. Ag. Statistics
Vol. XXXIX, No. 2 (1987), pp. 121-126

ON CERTAIN ALTERNATIVE IPNS SCHEMES

ARIJIT CHAUDHURI and ARUN KUMAR ADHIKARY
Indian Statistical Institute, Calcutta

(Received : March, 1980)

SUMMARY

In taking SRS's of varying sample-sizes in k replicates it is shown how efficiency (i) may be gained in choosing the replicates WOR rather than WR, but (ii) is lost relative to a comparable single SRSWOR if replicates are taken WR.

Keywords: Simple random sampling (SRS), With and without replacement (WR and WOR) sampling, Interpenetrating network of sub-sampling (IPNS) techniques, unequal size sub-samples, replications.

Introduction

In taking a simple random sample without replacement from a finite population of size N to estimate the population total $Y = \sum_{i=1}^{N} Y_i$ in the form of k interpenetrating network of sub-samples (IPNS), Koop [1] showed that for the sake of efficiency one should take k independent (i.e. WR) replicates of SRSWOR's of varying sizes rather than each of a common size. Roy and Singh [4] suggested taking k replicates of a common size each but WOR rather than WR. We consider k replicates WOR but of varying sizes and show that efficiency is thereby gained over Koop's procedure though not necessarily over Roy and Singh's [4].

2. The Results

Let $\bar{Y} = Y/N$, $S^2 = 1/(N-1) \sum_{i=1}^{N} (Y_i - \bar{Y})^2$ denote the population mean and variance and $\bar{y}_i$ the mean for the ith sample-replicate (or sub-sample). When each replicate has a common size n, one has

$$t_i = n \sum_{j=1}^{i-1} \bar{y}_j + [N - n(i-1)]\, \bar{y}_i, \qquad i = 1, \ldots, k, \; (\bar{y}_0 = 0),$$

as an unbiased estimator for Y and when the ith replicate has a size n_i, an unbiased estimator for Y is

$$t'_i = \sum_{j=1}^{i-1} n_j \bar{y}_j + (N - a_i)\, \bar{y}_i \quad \text{where } a_i = \sum_{j=1}^{i-1} n_j, \quad (n_0 = 0).$$

If the replications are taken without replacement from the population, following Raj [3] and Roy and Singh [4] one may work out the variances as

$$V(t_i) = S^2/n\, [N^2 + Nn - n(2N+n)\, i + n^2 i^2] \tag{2.1}$$

$$V(t'_i) = S^2/n_i\, [(N - a_i)(N - a_{i+1})], \quad i = 1, \ldots, k \tag{2.2}$$

and note the zero covariances for the pairs t_i, t_j's and t'_i, t'_j's, $i \neq j$. Roy and Singh [4] recommended the use of the estimator $\bar{t} = 1/k \sum_{i=1}^{k} t_i$ for Y with a variance $V(\bar{t}) = 1/k^2 \sum_{i=1}^{k} V(t_i) = S^2/nk\, [N^2 - Nnk + n^2/3\, (k^2 - 1)]$. Taking $\bar{t}' = 1/k \sum_{i=1}^{k} t'_i$ as a competitor with a variance $V(\bar{t}') = 1/k^2 \sum_{i=1}^{k} V(t'_i)$, the difference between them may be simplified directly from (2.1) and (2.2) to get $D = V(\bar{t}') - V(\bar{t}) = S^2/k^2 \sum_{i=1}^{k} [(N - a_i)(N - a_{i+1})/n_i - (N - in)(N - (i-1)n)/n]$. It is easy to construct examples to show that D may take both positive and negative values. Now, better estimators than $\bar{t}$ and $\bar{t}'$ respectively are immediately available as, using optimal weights W_i, W'_i, $\bar{t}_0 = \sum_{i=1}^{k} W_i t_i$ and $\bar{t}'_0 = \sum_{i=1}^{k} W'_i t'_i$ respectively with $W_i \propto 1/V(t_i)$ and $W'_i \propto 1/V(t'_i)$ giving $V(\bar{t}_0) = 1 \big/ \sum_{i=1}^{k} 1/V(t_i)$ and $V(\bar{t}'_0) = 1 \big/ \sum_{i=1}^{k} 1/V(t'_i)$. Here W_i and W'_i are free of unknown parameters and hence $\bar{t}_0$ and $\bar{t}'_0$ are both usable. Interestingly, one may check that $\bar{t}_0$ and $\bar{t}'_0$ both coincide with the corresponding over-all sample mean based on all the k replicates pooled together giving over-all sample-sizes nk and $\sum_{i=1}^{k} n_i$ respectively. With $nk = \sum_{i=1}^{k} n_i$, there is nothing to choose between $\bar{t}_0$ and $\bar{t}'_0$.

As usual and as is easy to check $1/k(k-1) \sum_{i=1}^{k} (t_i - \bar{t})^2$ and $1/k(k-1) \sum_{i=1}^{k} (t'_i - \bar{t}')^2$ are unbiased estimators respectively for $V(\bar{t})$ and $V(\bar{t}')$.

Moreover, $1/(k-1) \sum_{i=1}^{k} W_i (t_i - \bar{t}_0)^2$ and $1/(k-1) \sum_{i=1}^{k} W'_i \ (t'_i - \bar{t}'_0)^2$ are respectively unbiased estimators for $V(\bar{t}_0)$ and $V(\bar{t}'_0)$.

In case the replicated simple random samples without replacement of sizes $n_1, \ldots, n_k$ are taken WR, then Koop's [1] estimator for Y is $T = N \sum_{i=1}^{k} W''_i \bar{y}_i$ with $W''_i \propto 1/V(\bar{y}_i)$ so that $V(T) = NS^2 \Big/ \sum_{i=1}^{k} n_i/(N - n_i) = NS^2 \Big/ \sum_{i=1}^{k} (f_i)/(1 - f_i)_g$ with $f_i = n_i/N$. But this estimator is less efficient than $\bar{t}'_0$ as we have the Theorem 1. $V(T) > V(\bar{t}'_0)$.

Proof. $V(T) - V(\bar{t}'_0) = S^2 [N \Big/ \sum_{i=1}^{k} n_i/(N - n_i) - 1 \Big/ \sum_{i=1}^{k} n_i/(N - a_i)$ $(N - a_{i+1})] > 0$, because $[N/(N - a_i)\ (N - a_{i+1}) - 1/(N - n_i)] > 0$, as one may easily check.

A non-negative unbiased estimator for $V(T)$ is also given by Koop [1]. One of the principal uses of the IPNS technique introduced by Mahalanobis [2] is that it yields a simple alternative measure of error in estimation. Yet it is still important to examine how one may lose in terms of efficiency compared to an alternative but comparable over-all single sampling procedure.

In following Koop's (1) procedure the expected number of distinct units is $N\left[1 - \prod_{i=1}^{k} (1 - f_i)\right] = m$, say. For simplicity, let us assume this to be an integer and let one take a single SRSWOR of size m getting the sample mean $\bar{y}$ and use the expansion estimator $N\bar{y}$ as the usual unbiased estimator for Y with variance $V = V(N\bar{y}) = N(N - m)\ S^2/m$. To compare the relative efficiency of T versus $N\bar{y}$ we may consider $R = V(N\bar{y})/$

$V(T)$. One may note that

$R = \left(\prod_{i=1}^{k} (1-f_i)/\left[1 - \prod_{i=1}^{k} (1-f_i)\right]\right) \sum_{i=1}^{k} (f_i/(1-f_i))$ to get the Theorem 2. $R < 1$.

Proof. Recalling that if $x_i > 0$, $i = 1, \ldots, k$, one has $\prod_{i=1}^{k} (1 + x_i) > 1 + \sum_{i=1}^{k} x_i$, and hence

$$1 = \prod_{i=1}^{k} \prod_{i=1}^{k} (1 - f_i)/(1 - f_i) = \left(\prod_{i=1}^{k} (1-f_i)\right) \prod_{i=1}^{k} (1 + f_i/(1 - f_i))$$

$$> \prod_{i=1}^{k} (1 - f_i)\left[1 + \sum_{i=1}^{k} f_i/(1 - f_i)\right],$$ which gives

$$\left[1 - \prod_{i=1}^{k} (1 - f_i)\right] > \left[\prod_{i=1}^{k} (1 - f_i)\right]\left(\sum_{i=1}^{k} (f_i)/(1 - f_i)\right),$$

implying the result. Of course the lower the value of R, the less the efficiency of T relative to $N\bar{y}$ and with increasing sampling fractions f_i, $i = 1, \ldots, k$, for a fixed k, T suffers loss in its efficiency as one may observe from the following Theorem 3. *For a fixed k, with f_i $(0 < f_i < 1)$ increasing, R decreases monotonically.*

Proof. We will write $B = \prod_{i=1}^{k} (1 - f_i)$ and note

$$\frac{\partial R}{\partial f_i} = -\frac{1}{\left[1 - \prod_{i=1}^{k} (1 - f_i)\right]^2}\left[\prod_{j \neq i} (1 - f_j)\right]\left[\sum_{i=1}^{k} (f_i/(1 - f_i)\right]$$

$$+ \left[\frac{\prod_{i=1}^{k} (1 - f_i)}{1 - \prod_{i=1}^{k} (1 - f_i)}\right]\Big/(1 - f_i)^2$$

$$= -\frac{B/(1 - f_i)}{(1 - B)^2} \sum_{i=1}^{k} (f_i/(1 - f_i)) + \frac{B}{1 - B} \frac{1}{(1 - f_i)^2}$$

$$= \left(\frac{B}{(1-B)(1-f_i)} \left[\frac{1}{(1-f_i)} - \frac{1}{(1-B)} \sum_{i=1}^{k} (f_i/(1-f_i) \right] \right.$$

$$= \frac{B}{(1-B)^2} \frac{1}{(1-f_i)} \left[\frac{1-B}{1-f_i} - \sum_{i=1}^{k} (f_i/(1-f_i) \right]$$

$$= \frac{B}{(1-B)^2 (1-f_i)} \left[1 - \prod_{j \neq i} (1-f_i) - \sum_{j \neq i} (f_i/(1-f_i)) \right]$$

Now, since (i) $0 < f_i < 1$, $(f_i/(1-f_i)) > f_i$, and (ii) for $z_i > 0$ for every i, $\prod_{i=1}^{k} (1-z_i) > 1 - \sum_{i=1}^{k} z_i$, we have $1 - \prod_{j \neq i} (1-f_j) - \sum_{j \neq i} \frac{f_j}{(1-f_j)}$

$< 1 - \prod_{j \neq i} (1-f_j) - \sum_{j \neq i} f_j < 0.$

Hence $\frac{\partial R}{\partial f_i} < 0$, proving the theorem.

We may conclude with the following remarks about the usefulness of the findings above. Choosing the replicates WR one sacrifices efficiency but in choosing them WOR it is not necessarily the case keeping in view a comparable procedure based on the over-all sample. In the latter case, one need not wait to complete the processing of all the k replicates for the estimation because with r $(<k)$ replicates at hand one may use

$\bar{t}_r = \frac{1}{r} \sum_{j=1}^{r} t_j$ or $\bar{t}' = \frac{1}{r} \sum_{j=1}^{r} t'_j$ to estimate Y and $\frac{1}{r(r-1)} \sum_{1}^{r} (t_i - \bar{t}_r)^2$,

$\frac{1}{r(r-1)} \sum_{1}^{r} (t'_i - \bar{t}'_r)^2$

to unbiasedly estimate $V(\bar{t}_r)$, $V(\bar{t}'_r)$ respectively. In this case it is easy to choose appropriate n_i keeping formulae (2.1), (2.2) in mind to choose between the use of either $\bar{t}_r$ or $\bar{t}'_r$ in each particular situation. Hence the utility of our discussion in this article in dealing with an efficient IPNS scheme in estimating the population total admitting as well a simple unbiased variance estimator.

ACKNOWLEDGEMENT

The authors are grateful to the referees for their helpful critical comments which led to this version improving on two earlier drafts.

REFERENCES

[1] Koop, J. C. (1967) : Replicated (or interpenetrating) samples of unequal sizes, *Ann. Math. Stat.*, **38** : 1142-1147.

[2] Mahalanobis, P. C. (1946) : Recent experiments in statistical sampling in the Indian Statistical Institute, *J. Roy. Stat. Soc.* (*A*), **109** : 325-378.

[3] Raj, Des (1956) : Some estimators in sampling with varying probabilities without replacement, *J. Amer. Stat. Ass.*, **51** : 269-284.

[4] Roy, A. S. and Singh, M. P. (1973) : Interpenetrating sub-samples—with and without replacement, *Metrika*, **20** : 230-239.