Jour. Ind. Soc. Ag. Statistics
45 (3), 1993 : 307-318

On Certain Alternative Estimators for Multiple Characteristics in Varying Probability Sampling

T. J. Rao*
University of California, Santa Barbara, California, USA
(Received : March, 1990)

Summary

Bansal and Singh [3] andAmahia, et al. [2] have developed alternative estimators for probability proportional to size with replacement sampling scheme when some characteristics under study are poorly correlated with the selection probabilities. In this paper, these results are extended to other varying probability sampling designs.

Key Words : Multipurpose surveys, Probability proportional to size sampling with replacement and without replacement, Horvitz-Thompson estimator, Rao-Hartley-Cochran estimator, Murthy's estimator.

Introduction

In large scale sample surveys where it is of interest to estimate parameters relating to several characteristics, it is sometimes found that some of the study variables are poorly correlated with the selection probabilities when sampling is done by probability proportional to size (pps) sampling method. Rao [6] has suggested alternative estimators when the study variable and the size measure are unrelated and demonstrated the efficiency of these alternative estimators. Recently, Bansal and Singh [2] observed that the model of Rao's [6] is not always applicable "since the correlation in the population is never exactly equal to zero", and obtained a new estimator of the population total for characteristics that are poorly correlated with the selection probabilities for pps sampling with replacement. Amahia, et al., [2] considered much simpler alternative estimators for the situations discussed in [3], which are intuitively simple and easy to interpret. They have also demonstrated the efficiency and robustness of these estimators. In this paper, we shall discuss the efficiency of certain other estimators in pps sampling without replacement for multiple characteristics.

* On leave from Indian Statistical Institute, Calcitta, India.

estimator, say $\hat{Y}_{R,HT}$ while $\rho = 1$ gives $p_i^* = p_i$ and then (2.1.3) reduces to the conventional HT estimator (2.1.1), $\hat{Y}_{C,HT}$. This estimator is thus motivated by the fact that for values of ρ near 1, it is "close" to the HT estimator and for values of ρ near zero it is "close" to Rao's [6] alternative estimator and the other values of ρ generate a class of alternatives.

On the other hand motivated by the Bias criterion, one can define an alternative estimator given by Amahia et al., [2]

$$\hat{Y}'_{P,HT} = (1-\rho)\,\hat{Y}_{R,HT} + \rho\,\hat{Y}_{C,HT}$$

which immediately gives

$$B(\hat{Y}'_{P,HT}) \le B\left(\hat{Y}_{R,HT}\right)$$

This estimator can be written as

$$\hat{Y}'_{P,HT} = \sum_{i=1}^{n} \frac{y_i\, p_i}{p'_i\, \pi_i} \tag{2.1.4}$$

where $$P'_i = \left[(1-\rho)\,N + \left(\frac{\rho}{p_i}\right)\right]^{-1}$$

It is easy to see that

$$V(\hat{Y}_{P,HT}) = \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) \frac{p_i^2}{p_i^{*2}}\, Y_i^2 + \sum_{i \ne j}^{N} \sum^{N} \left[\frac{\pi_{ij}}{\pi_i\, \pi_j} - 1\right] \frac{p_i\, p_j}{p_i^*\, p_j^*}\, Y_i Y_j \tag{2.1.5}$$

with a similar expression for $V(\hat{Y}'_{P,HT})$.

First we observe that

$$B(\hat{Y}'_{P,HT}) = (1-\rho)\, B(\hat{Y}_{R,HT})$$

$$= (1-\rho) \sum_{i=1}^{N} Y_i\,(Np_i - 1)$$

and $$B(\hat{Y}_{P,HT}) = \sum_{i=1}^{n} Y_i \left(\frac{p_i}{p_i^*} - 1\right)$$

Note that

$$a'_{P,HT} - a_{P,HT} = a \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) p_i^{g+2} \left(\frac{1}{p_i'^2} - \frac{1}{p_i^{*2}}\right) \quad (2.1.10)$$

$$\geq 0$$

where p'_i and p_i^* are respectively the weighted h.m. and the weighted a.m. of $1/N$ and p_i. Thus

$$\& \; V(\hat{Y}'_{P,HT}) > V(\hat{Y}_{P,HT})$$

if

$$V \sum_{i=1}^{n} \frac{P_i}{np'_i} > V \sum_{i=1}^{n} \frac{P_i}{np_i^*}. \quad (2.1.11)$$

Also from (2.1.7) and (2.1.9) it can be concluded that the conventional estimator can be better than the proposed estimator $\hat{Y}_{P,HT}$ whenever

$$a_{P,HT} > a_{C,HT}$$

i.e. whenever

$$\sum_{i=1}^{N} \frac{1 - np_i}{np_i} \; p_i^g \; \frac{p_i^2 - p_i^{*2}}{p_i^{*2}} > 0. \quad (2.1.12)$$

We shall now state a lemma to investigate (2.1.12).

Lemma 2.1. (Royall, [9]: Let $0 < b_1 \leq b_2 \leq \ldots \leq b_n$ and $c_1 \leq c_2 \leq \ldots \leq c_n$ satisfying $\sum_{i=1}^{n} c_i \geq 0$, then $\sum_{i=1}^{n} b_i c_i \geq 0$.

We write

$$a_{P,HT} - a_{C,HT} = \sum_{i=1}^{n} b_i c_i \quad (2.1.13)$$

where $b_i = \dfrac{1 - np_i}{np_i} \; p_i^g \dfrac{(p_i + p_i^*)}{p_i^{*2}}$ and $c_i = p_i - p_i^*$

and using lemma 2.1, we get that

Following Amahia et al., [2] we suggest two alternative estimators as in the case of Horvitz–Thompson estimator, viz.,

$$\hat{Y}_{PR} = \sum_{i=1}^{n} \frac{y_i \mu_i}{p_i^*} \tag{2.2.3}$$

and

$$\hat{Y}'_{PR} = \sum_{i=1}^{n} \frac{y_i \mu_i}{p_i'} \tag{2.2.4}$$

which are obtained by replacing y_i in (2.2.1) by $y_i \frac{p_i}{p_i^*}$ and $y_i \frac{p_i}{p_i'}$ respectively where, as before, $p_i^* = \rho p_i + \frac{(1-\rho)}{N}$ and $p_i' = \left((1-\rho)N + \frac{\rho}{p_i}\right)^{-1}$. Once again it is easily verified that $\rho = 0$ in (2.2.3) corresponds to Rao's [6] alternative estimator and $\rho = 1$ corresponds to the conventional estimator (2.2.1).

As before, we notice that

$$B(\hat{Y}_{PR}) = \sum_{i=1}^{N} Y_i \left[\frac{p_i}{p_i^*} - 1\right]$$

and

$$B(\hat{Y}'_{PR}) = \sum_{i=1}^{N} Y_i \left[\frac{p_i}{p_i'} - 1\right]$$

$$= (1-\rho) \sum_{i=1}^{N} Y_i (Np_i - 1)$$

so that

$$B(\hat{Y}'_{PR}) - B(\hat{Y}_{PR}) = \rho(1-\rho) \sum_{i=1}^{N} \frac{Y_i(N p_i - 1)^2}{(1 - \rho + N \rho p_i)}$$

$$> 0, \text{ when } \rho > 0.$$

As before we propose two alternative estimators for this case given by

$$\hat{Y}_{PM} = \frac{1}{2 - p_1 - p_2}\left\{(1 - p_2)\frac{y_1}{p_1^*} + (1 - p_1)\frac{y_2}{p_2^*}\right\}. \qquad (2.3.2)$$

and

$$\hat{Y}'_{PM} = \frac{1}{2 - p_1 - p_2}\left\{(1 - p_2)\frac{y_1}{p_1'} + (1 - p_1)\frac{y_2}{p_2'}\right\}, \qquad (2.3.3)$$

where p_i^* and p_i' are as defined before. Also, it can be shown that $B(\hat{Y}'_{PM}) > B(\hat{Y}_{PM})$ as in the other cases.

It now follows from

$$V(\hat{Y}_{CM}) = \sum_{i}^{N} \sum_{<j}^{N} p_i p_j \frac{(1 - p_i - p_j)}{(2 - p_i - p_j)}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

that

$$V(\hat{Y}_{PM}) = \sum_{i}^{N} \sum_{<j}^{N} p_i p_j \frac{(1 - p_i - p_j)}{(2 - p_i - p_j)}\left(\frac{y_i}{p_i^*} - \frac{y_j}{p_j^*}\right)^2$$

and

$$V(\hat{Y}'_{PM}) = \sum_{i}^{N} \sum_{<j}^{N} p_i p_j \frac{(1 - p_i - p_j)}{(2 - p_i - p_j)}\left(\frac{y_i}{p_i'} - \frac{y_j}{p_j'}\right)^2,$$

by substituting $\frac{y_i p_i}{p_i^*}$ and $\frac{y_i p_i}{p_i'}$ respectively in the place of y_i.

Under the superpopulation model (2.1.6) we obtain

$$\mathcal{E}V(\hat{Y}_{CM}) = 2a \sum_{i<j}\sum \frac{(1 - p_i - p_j)}{(2 - p_i - p_j)} p_i^{g-1} p_j$$

$$\mathcal{E}V(\hat{Y}_{PM}) = 2a \sum_{i<j}\sum \frac{(1 - p_i - p_j)}{(2 - p_i - p_j)} \frac{p_i^{g+1} p_j}{p_i^{*2}} + \beta_{PM}$$

and $$\mathcal{E}V(\hat{Y}'_{PM}) = 2a \sum_{i<j}\sum \frac{(1 - p_i - p_j)}{(2 - p_i - p_j)} \frac{p_i^{g+1} p_j}{p_i'^2} + \beta'_{PM}$$

Table 1. Data

Population I : $N = 10, \rho = 0.488, n = 2$												
x	25	32	14	70	24	20	32	44	50	44		
y	11	7	5	27	30	6	13	9	14	18		
Population II : $N = 12, \rho = 0.05, n = 2$												
x	41	34	54	39	49	45	41	33	37	41	37	39
y	36	47	41	47	49	45	32	37	40	41	37	48

Table 2. Expected Variances of different estimators

	g	$\mathcal{E}V(\hat{Y}_{PR})$	$\mathcal{E}V(\hat{Y}'_{PR})$	$\mathcal{E}V(\hat{Y}_{CR})$
Pop. I	0	$37.89a + 0.0222\beta^2$	$42.16a + 0.0267\beta^2$	50.04a
	1	$3.75a + 0.0222\beta^2$	$4.11a + 0.0267\beta^2$	4.00a
	2	$0.44a + 0.0222\beta^2$	$0.49a + 0.0267\beta^2$	0.39a
Pop. II	0	$59.79a + 0.0085\beta^2$	$59.91a + 0.0086\beta^2$	61.28a
	1	$5.06a + 0.0085\beta^2$	$5.07a + 0.0086\beta^2$	5.00a
	2	$0.44a + 0.0085\beta^2$	$0.44a + 0.0086\beta^2$	0.42a

The above table gives the expected variances with respect to the model considered. Using the data on y, we can actually compare the exact variances of the estimators considered for the sake of completeness:

Table 3. Exact variances for $n = 2$.

Estimator	Pop. I	Pop. II
$\hat{Y}_{CR}$	3279.31	3922.60
$\hat{Y}_{RR}$	2865.55	1807.90
$\hat{Y}'_{PR}$	2476.21	1805.18
$\hat{Y}_{PR}$	2187.67	1801.05