CHAPTER 5

GENERALIZATION OF MURTHY'S SCHEME AND ESTIMATOR

5.1. Introduction

We consider a Population-of N units labeled $I = 1, 2, \cdots, N$. Associated with each unit I are two values X_I and Y_I believed to be roughly in proportion. X_I is known for each unit I in advance. Only those units Y_I are made known which are in a sample S drawn according to some sampling scheme described by P(S), so that $\sum_S P(S) = 1$, Let $\sum_1^N X_I = X$ and $\sum_1^N Y_I = Y$, the unknown Population-total of Y's. For ready reference Raj and Murthy estimator are given below, Raj (1956) developed an ingenious method of estimating the Population-total and unbiased estimator of its variance for a general sample size n. The sampling scheme that he employed draws the first unit i with probability proportional to size i.e. $P_i = X_i/X$. The second unit $j \neq i$ is then drawn with probability proportional to size of the remaining units i.e. $P_{j|i} = P_i/(1-P_i)$. He then used the order of drawing the units to produce (for n = 2)

$$t_1 = y_i/p_i$$

$$t_2 = y_i + \frac{y_j}{p_j}(1-p_i)$$

Both t_1 and t_2 are unbiased estimator of Y and what is more that t_1 and t_2 are uncorrelated. He then choose to get his estimator $t_{mean} = \frac{1}{2}(t_1 + t_2)$ with the variance:

$$\text{Var}(y'_R) = \frac{1}{8}\sum_{I=1}^{N}\sum_{\substack{J=1 \\ I \neq J}}^{N} P_I P_J (2 - P_I P_J)\left(\frac{Y_I}{P_I} - \frac{Y_J}{P_J}\right)^2 \qquad (1.5.6)$$

An unbiased estimator of this is:

$$\text{var}\left(y'_R\right) \equiv \frac{\left(1-p_i\right)^2}{4}\left(\frac{y_i}{p_i}-\frac{y_j}{p_j}\right)^2 \tag{1.5.9}$$

Murthy (1957) using ideas of Basu (1958) suggested that since (y_i, y_j) without regard to their order of drawing is sufficient. Rao-Blackwellization can then be used to improve Raj's estimator. The improved estimator so obtained by Murthy (1957) is:

$$t_{symm} = \frac{\frac{y_i}{p_i}\left(1-p_j\right)+\frac{y_i}{p_j}\left(1-p_i\right)}{\left(1-p_i\right)+\left(1-p_j\right)} \tag{1.5.11}$$

with variance:

$$\text{Var}\left(y'_M\right) = \frac{1}{2}\sum_{I=1}^{N}\sum_{\substack{J=1\\I\neq J}}^{N} P_I P_J \frac{1-p_I-p_J}{2-p_I-p_J}\left(\frac{T_I}{P_I}-\frac{Y_J}{P_J}\right)^2 \tag{1.5.12}$$

An unbiased variance estimator of this is:

$$\text{var}\left(y'_M\right) = \frac{\left(1-p_i\right)\ \left(1-p_j\right)\ \left(1-p_i-p_j\right)}{\left(2-p_i-p_j\right)^2}\left(\frac{y_i}{P_i}-\frac{y_j}{P_j}\right)^2 \tag{1.5.13}$$

We will attempt to generalize Murthy's estimator in several ways in section 2.

5.2 Generalization

We start with a very general set up. Let n =2 so that two units are drawn one after another without replacement. The probability of drawing i^{th} unit at the first draw is q_i such that $\sum_{1}^{N} q_i = 1$ The probability of drawing j^{th} unit ($j \neq i$) at the second draw given that i^{th} unit was drawn at the first draw is $q(j \mid i)$, $\sum_{\substack{j=1\\j\neq i}}^{N} q(j \mid i) = 1$ Proceeding as in Raj (1956) define

$$\left.\begin{array}{l} t_1 = y_i/q_i \\ t_2 = y_i + \dfrac{y_j}{q(j|i)} \end{array}\right\} \qquad (5.2.1)$$

A linear combination of t_1 and t_2 gives

$$T_{12} = C \ y_i/q_i + (1-C)\left(y_i + \frac{y_j}{q(j|i)}\right), \qquad 0 \le C \le 1 \qquad (5.2.2)$$

Murthy (1957) observed that $(y_i, y_j) \equiv (y_j, y_i) \equiv S$ is sufficient i.e. the order in which units i and j are drawn are irrelevant. Using the process of Rao-Blackwellisation we have two estimators depending on their order of drawing

$$T_{12} = c \ y_i/q_i + (1-c)\left(y_i + \frac{y_j}{q(j|i)}\right) \qquad \text{with probability } q_i\, q(j \mid i)$$

and

$$T_{21} = c \ y_j/q_j + (1-c)\left(y_j + \frac{y_i}{q\ (i|j)}\right) \qquad \text{with probability } q_j\, q(i \mid j)$$

$$\text{Now } E(T_{lm} \mid S) = \frac{(T_{12})\ q_i q\ (j|i) + (T_{21})\ q_j q\ (i|j)}{q_i q\ (j|i) + q_j q\ (i|j)} \quad \text{where } l, m = 1,2$$

and

$$l \ne m = \frac{\left\{c y_i/q_i + (1-c)\left(y_i + \frac{y_j}{q(j|i)}\right)\right\} q_i q(j|i) + \left\{c y_j/q_j + (1-c)\left(y_j + \frac{y_i}{q(i|j)}\right)\right\} q_j q_{(i|j)}}{q_i q\ (j|i) + q_j q\ (i|j)}$$

(5.2.3)

Of course it would be ideal to free the estimator from c so that any linear combination will lead to the same estimator. This avoids the arbitrariness of choosing c = 1/2 in Raj's scheme. For this to happen the expression in (5.2.3) should be free from c for which the condition is

$$\left\{y_i/q_i - \left(y_i + \frac{y_j}{q(j|i)}\right)\right\} q_i q(j|i) + \left\{y_j/q_j - \left(y_j + \frac{y_i}{q(i|j)}\right)\right\} q_j q(i|j) = 0$$

for all y_i and y_j

Equating Coefficient of y_i and y_j to zero we get

$$\left(\frac{1}{q_i}-1\right)q_i q(j|i)-\frac{q_j q(i|j)}{q(i,j)}=0 \Rightarrow q(j|i)=\frac{q_j}{1-q_i} \text{ as in Raj (1956)} \quad (5.2.4)$$

However there are two essential differences. These are

i) q_i need not be equal to $P_I = X_I/X$ and this can be exploited

ii) If $q(j|i)=\frac{q_j}{1-q_i}$ then Murthy's (1957) estimator turns out to be $\left(\frac{y_i}{q_i}\frac{1}{1-q_i}+\frac{y_j}{q_j}\frac{1}{1-q_j}\right)\bigg/\left(\frac{1}{1-q_i}+\frac{1}{1-q_j}\right)$ which is precisely Murthy's estimator if $q_i = p_i = X_i/X$

Pathak (1961) and independently Samiuddin and Hanif (1978) proved strictly in relation to Raj's estimator that any linear combination would lead to the same Murthy's estimator. The condition is thus sufficient. We now see it to be necessary as well in a more general context. We have considered several ways of generalizing Murthy's estimator. These are as follows:

a) Select first unit i with probability $q_i = \frac{p_i(1-p_i)}{1-ap_i}\bigg/\sum_{1}^{N}\frac{p_i(1-p_i)}{1-ap_i}$ and second unit $j \neq i$ with probability $q_j/(1-q_i)$. Note that in this case starting with Raj's estimate we get the same Murthy estimator for all values of c. For each selected value of **a** we get a Murthy estimator.

b) Select first unit i with probability $q_i = \frac{p_i(1-p_i)}{1-ap_i}\bigg/\sum_{1}^{N}\frac{p_i(1-p_i)}{1-ap_i}$ and second unit j with probability $p_j/(1-p_i)$. We have chosen c = 1/2 to construct Murthy's estimator starting from Raj's estimator. Each value of **a** leads to a distinct Murthy estimator.

c) Based on Shahbaz et al (2003) selection procedure, the first unit i is selected with probability $\frac{p_i(1-p_i)}{1-2p_i}\bigg/\sum_{1}^{N}\frac{p_i(1-p_i)}{1-2p_i}$, the second unit $j \neq i$ with probability proportional to $p_j\left[\frac{1}{1-2p_i}+\frac{1}{1-2p_j}\right]$.

Here we have selected different values of c to construct Murthy's estimator starting from Raj.

d) Following Samiuddin et al (1992) we select first unit i with probability proportional to p_i and the second unit $j \neq i$ with equal probability without replacement from the remaining units and use the generalized Murthy's estimator $\sum_{i=1}^{n} y_i P\ (S|i)/P\ (S)$ where having selected the first unit i, the rest of the sample S is selected in what ever way we like. In the above case the resulting estimator is $\left[(y_i+y_j)/(x_i+x_j)\right]X$ which is the well-known ratio estimator. Notice that this estimator is both design and model unbiased in this particular case. We also note that there are other ways in which the second unit j can be selected (such as $p_j/(1-p_i)$ which will lead to usual Murthy's estimator).

5.3. Empirical Study

In this study 50 real populations are considered. The variance of the resulting Murthy's estimator is calculated in each case. These variances are then ranked giving rank 1 to the lowest variance and highest rank to the largest variance. These comparisons are therefore based on ranks (and not on relative efficiencies). Table 5.1 gives the consolidated frequency table of ranks for each value of **a** in group of estimator in (**a**). Table 5.2 gives the consolidated frequency table of rank for each value of **a** (and c = 1/2) in the groups of estimate in (b). The case **a** = 0 outperforms all others. The first units i in this is selected with probability $p_i(1-p_i)\Big/\left(1-\sum_{i=1}^{N} p_i^2\right)$. Table 5.3 gives the frequency table of ranks for resulting Murthy estimator in group (c) for various values of constant c. The last row in each of the above Tables gives the average rank. We notice in Table 1 that **a** = 0, **a** = 0.5 and **a** = 1.0 (Murthy) in the last row give the lowest average rank.

In the Table 5.3, c = 1 leads to a clear lowest average rank.

Finally we need to compute the best in each group with each other. Notice that they altogether make up 2 in group (a), 2 in group (b), 1 in Table 5.3, Murthy estimator (Common to group (a) and (b)) and finally 1 in group (d). So altogether we have seven estimators to be compared. This is done in the form of frequency table and average rank. The results are given in Table 5.4. A look at table 5.4 suggests that both the best performers belong to group (a) for **a** = 0 and **a** = 0.5. Although the average rank for **a** = 0.5 (3.18) is smaller than those for **a** = 0.0 (3.30). There is something to recommend **a** = 0.0 in some aspects, which are commendable. For example it has the maximum number of cases (19) where it performs best of all the 7 types considered here

Table 5.1:
Frequency table of ranks for estimators in group (a)

	Values of "a"								
Rank	0	0.5	1	1.5	2	2.5	3	3.5	4
1	28	6	3	2	4	0	2	0	5
2	4	31	3	2	1	4	0	5	0
3	2	0	35	5	1	0	5	2	0
4	3	2	2	34	2	7	0	0	0
5	0	2	0	1	42	2	2	0	1
6	3	2	0	6	0	37	1	0	1
7	2	0	7	0	0	0	40	1	0
8	1	7	0	0	0	0	0	40	0
9	7	0	0	0	0	0	0	0	38
Mean	3.14	3.08	3.42	3.96	4.54	5.36	6.26	7.15	7.96

Table 5.2:
Frequency table of ranks for estimators in group (b)

	Values of "a"								
Rank	0	0.5	1	1.5	2	2.5	3	3.5	4
1	31	1	3	0	2	2	1	2	8
2	3	33	0	2	0	3	2	7	0
3	1	2	32	3	2	1	7	0	2
4	2	0	2	35	0	8	1	1	1
5	0	1	2	2	44	0	0	1	0
6	1	2	2	8	1	35	1	0	0
7	1	1	9	0	0	1	33	1	1
8	1	10	0	0	0	0	1	33	1
9	10	0	0	0	1	0	1	1	33
Mean	3.18	3.54	3.84	4.22	4.86	5.20	6.04	6.63	7.17

Table 5.3:
Frequency table of ranks for estimators in group (c)

Rank	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	9	1	0	1	0	0	2	0	1	3	33
2	0	9	3	0	0	1	0	1	0	33	3
3	1	1	8	0	1	0	0	2	36	1	0
4	0	1	1	11	0	0	1	36	0	0	0
5	1	0	0	0	10	1	36	0	0	1	1
6	1	0	0	0	1	47	0	0	0	1	0
7	0	1	0	0	37	1	11	0	0	0	0
8	0	0	1	37	0	0	0	11	1	0	0
9	0	0	37	0	1	0	0	0	11	0	1
10	0	37	0	0	0	0	0	0	1	11	1
11	38	0	0	1	0	0	0	0	0	0	11
Mean	8.82	8.06	7.5	7.04	6.54	5.92	5.26	4.8	4.52	3.86	3.68

Table 5.4:
Frequency table of ranks chosen from the member of different groups

Rank	Group a A=0.0	Group a a=0.5	Group b a=0.0	Group b a=0.5	Group c c=1.0	Group d	Generalized Murthy or Ratio estimator
1	19	3	0	0	8	16	4
2	9	20	4	2	2	1	12
3	2	10	21	15	2	0	0
4	3	2	21	19	3	2	0
5	0	12	3	12	6	3	14
6	9	3	1	1	13	3	20
7	8	0	0	1	16	25	0
Average	3.30	3.18	3.52	3.96	5.00	4.68	4.36