ALMOST UNBIASED ESTIMATORS BASED ON INTERPENETRATING SUB-SAMPLES

By M. N. MURTHY

Indian Statistical Institute

SUMMARY. In this paper a technique is given for estimating unbiasedly any non-linear function of estimable parameters. The technique consists in estimating the bias of the usual estimator using estimates based on interpenetrating sub-samples and then correcting the estimator for its bias.

1. Introduction

The question of evolving a generalized unbiased estimator for any sample design has been considered by Midzuno (1950), Godambe (1955) and Nanjamma, Murthy and Sethi (1959) for certain classes of parameters. Murthy (1962) has suggested a technique of generating unbiased estimators for any sample design for the class of parameters which can be expressed as a sum of single-valued set functions defined over a class of sets of units belonging to the finite population under consideration. Examples of such parameters are the population total Y and the population variance which can be expressed respectively as

$$Y = \sum_{i=1}^{N} Y_i$$

and

$$\sigma^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j>i} (Y_i - Y_j)^2.$$

In this paper we shall supplement the generalized theory of unbiased estimation (Murthy, 1962) by giving a procedure of obtaining unbiased (or almost unbiased) estimators for non-linear functions of parameters each of which can be expressed as single-valued set function defined over a class of sets of units belonging to a finite population. Examples of such parameters are given by ratio of population totals of two characteristics, population standard deviation, correlation coefficient, etc.

The procedure of obtaining this unbiased estimator consists in estimating the bias of the usual estimator which is taken as the same non-linear function of unbiased estimators of the parameters as the parametric function under consideration, on the basis of interpenetrating sub-sample estimates. This procedure is based on the technique used by Murthy and Nanjamma (1959) in estimating the bias of a ratio estimator.

The procedure given in this paper is likely to be of much help in survey practice, since the estimation of relationships between characteristics and between parameters, such as a ratio of population totals of two characteristics or the population coefficient of variation are usually of much interest in sample surveys.

2. Parametric function

Let the parametric function $f(\theta)$ be a single valued non-linear function of the parameters $(\theta_1, \theta_2, \ldots, \theta_k)$, where θ_i $(i = 1, 2, \ldots, k)$ can be expressed as

$$\theta_i = \sum_{a_i \epsilon A_i} f_i(a_i) \qquad \ldots \quad (2.1)$$

where $f_i(a_i)$ is a single-valued set function defined over the class 'A_i' of sets 'a_i' consisting of units belonging to the population X.

Suppose we have defined the sample space 'S' of samples 's' with a suitable probability measure such that it is possible to estimate the parameters $(\theta_1, \theta_2, \ldots, \theta_k)$ unbiasedly using the procedure given by Murthy (1962). That is, it is assumed that the sample space is so specified that each $a_i \epsilon A_i$ $(i = 1, 2, \ldots, k)$ occurs in at least one 's' and that each 's' contains at least one set 'a_i' in 'A_i' $(i = 1, 2, \ldots, k)$. Then a generalized unbiased estimator of θ_i $(i = 1, 2, \ldots, k)$ is given by

$$t_i = \hat{\theta}_i = \sum_{a_i \subset s} f_i(a_i)\phi_i(s, a_i)/P(s) \qquad \ldots \quad (2.2)$$

where

$$\sum_{s \supset a_i} \phi_i(s, a_i) = 1.$$

In fact, we can make the above formulation more general by relaxing the assumption that θ_i's $(i = 1, 2, \ldots, k)$ are estimated from the same samples. In other words, θ_i $(i = 1, 2, \ldots, k)$ may be estimated on the basis of the same, overlapping or non-overlapping samples drawn with the same or different sample designs.

Let $(t_1, t_2, \ldots, t_k)$ be unbiased estimators of the parameters $(\theta_1, \theta_2, \ldots, \theta_k)$. Then an estimator of $f(\theta)$ can be taken as $f(t)$. If $f(\theta)$ is a linear function, obviously $f(t)$ will be unbiased for $f(\theta)$. But here we are taking $f(\theta)$ as a non-linear function of $(\theta_1, \theta_2, \ldots, \theta_k)$ and hence $f(t)$ will, in general, be biased for $f(\theta)$.

3. Bias and mean square error

In this section approximate expressions for the bias and the mean square error of the estimator of $f(t)$ are obtained by using Taylor series symbolically. It may be noted that in statistical practice one is interested not so much in the convergence properties of the infinite series representing a function, but in finding out whether the first few terms of that series will give a good approximation to the function. Because of this, the question of the validity of the application of Taylor series expansion to the case of a finite population estimator will not be considered here. However, it will be assumed that the estimator t_i is such that $\left|\frac{t_i - \theta_i}{\theta_i}\right| < 1$, especially for estimators occurring in the denominator of the function $f(t)$ so that the first few terms of the expansion can be expected to give a good approximation to the function. This latter statement has been empirically verified in the case of applying this expansion to a ratio estimator.

If the sample size is fairly large, the assumption $\left|\frac{t_i-\theta_i}{\theta_i}\right| < 1$ will be valid. Let $t_i=\theta_i(1+e_i)$, $(i=1, 2, \ldots, k)$ and $t=(t_1, t_2, \ldots, t_k)$, $=(\theta_1, \theta_2, \ldots, \theta_k)$, $e=\theta(e_1, e_2, \ldots, e_k)$. Expanding $f(t)$ in a Taylor series about $t = \theta$ and neglecting terms of degree greater than 2 in e's, we get

$$f(t) = f(\theta)+\sum_{i=1}^{k} \theta_i e_i \left(\frac{\partial f}{\partial t_i}\right)_{t=\theta}$$

$$+\frac{1}{2}\left[\sum_{i=1}^{k} \theta_i^2 e_i^2 \left(\frac{\partial^2 f}{\partial t_i^2}\right)_{t=\theta} + 2\sum_{i=1}^{k}\sum_{j>i} \theta_i\theta_j e_i e_j \left(\frac{\partial^2 f}{\partial t_i \partial t_j}\right)_{t=\theta}\right] \quad \ldots \quad (3.1)$$

It may be observed that for certain parameters there will be no terms of degree greater than 2 to neglect. An example of such a parameter is the product $\theta_1\theta_2$ with the estimator t_1t_2. Taking expected value of $f(t)$ in (3.1), we find that the bias of $f(t)$ correct to the second degree of approximation is given by

$$B[f(t)] = \frac{1}{2}\left[\sum_{i=1}^{k} \left(\frac{\partial^2 f}{\partial t_i^2}\right)_{t=\theta} \mu_2(ii)+2\sum_{i=1}^{k}\sum_{j>i} \left(\frac{\partial^2 f}{\partial t_i \partial t_j}\right)_{t=\theta} \mu_2(ij)\right] \quad \ldots \quad (3.2)$$

where $\mu_2(ij) = E(t_i-\theta_i)(t_j-\theta_j), \quad i, j = 1, 2, \ldots, k.$

The mean square error of $f(t)$ to the second degree of approximation is given by

$$M[f(t)] = E[f(t)-f(\theta)]^2 = E\left[\sum_{i=1}^{k} \theta_i e_i \left(\frac{\partial f}{\partial t_i}\right)_{t=\theta}\right]^2$$

$$= \sum_{i=1}^{k} \left(\frac{\partial f}{\partial t_i}\right)^2_{t=\theta} \mu_2(ii)+2\sum_{i=1}^{k}\sum_{j>i} \left(\frac{\partial f}{\partial t_i}\right)_{t=\theta}\left(\frac{\partial f}{\partial t_j}\right)_{t=\theta} \mu_2(ij). \quad \ldots \quad (3.3)$$

4. Biases of two estimators

Suppose the sample on which the estimate t_i of θ_i $(i = 1, 2, \ldots, k)$ is based is selected in the form of n independent interpenetrating sub-samples. Let t_{is} be the unbiased estimate of θ_i based on the s-th independent interpenetrating sub-sample $(i = 1, 2, \ldots, k;\ s = 1, 2, \ldots, n)$. In this case let us consider the following two estimators T_1 and T_n of $f(\theta)$.

$$T_1 = \frac{1}{n}\sum_{s=1}^{n} f(t_s) \quad \ldots \quad (4.1)$$

where $t_s = (t_{1s}, t_{2s}, \ldots, t_{ks}), \quad (s = 1, 2, \ldots, n),$

and

$$T_n = f(\bar{t}) \quad \ldots \quad (4.2)$$

$$\bar{t} = (\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_k), \quad \bar{t}_i = \frac{1}{n}\sum_{s=1}^{n} t_{is}, \quad (i = 1, 2, \ldots, k).$$

Applying the result (3.2) to T_n in (4.2) we get

$$B(T_n) = \frac{1}{2}\left[\sum_{i=1}^{k}\left(\frac{\partial^2 f}{\partial t_i^2}\right)_{\mathbf{t}=\boldsymbol{\theta}} \bar{\mu}_2(ii)+2\sum_{i=1}^{k}\sum_{j>i}\left(\frac{\partial^2 f}{\partial t_i \partial t_j}\right)_{\mathbf{t}=\boldsymbol{\theta}} \bar{\mu}_2(ij)\right]$$

where $$\bar{\mu}_2(ij) = E(\bar{t}_i-\theta_i)(\bar{t}_j-\theta_j) = \frac{1}{n^2}\sum_{s=1}^{n}\mu_{2s}(ij),$$

$$\mu_{2s}(ij) = E(t_{is}-\theta_i)(t_{js}-\theta_j).$$

That is

$$B_n = B(T_n) = \frac{1}{n^2}\sum_{s=1}^{n}\frac{1}{2}\left[\sum_{i=1}^{k}\left(\frac{\partial^2 f}{\partial t_i^2}\right)_{\mathbf{t}=\boldsymbol{\theta}} \mu_{2s}(ii)+2\sum_{i=1}^{k}\sum_{j>i}\left(\frac{\partial^2 f}{\partial t_i \partial t_j}\right)_{\mathbf{t}=\boldsymbol{\theta}} \mu_{2s}(ij)\right]$$

$$= \frac{1}{n^2}\sum_{s=1}^{n} B[f(t_s)]. \quad \dots \quad (4.3)$$

The bias of the estimator T_1 in (4.1) is given by

$$B_1 = B(T_1) = \frac{1}{n}\sum_{s=1}^{n} B[f(t_s)]. \quad \dots \quad (4.4)$$

Comparing (4.3) and (4.4) we find that the bias of the estimator T_1 is n times that of the estimator T_n.

5. Estimation of bias

As observed in Section 4, comparing the biases of the estimators T_1 and T_n, we get

$$B_1 = n\,B_n. \quad \dots \quad (5.1)$$

Using this result we can derive an unbiased estimator of the bias B_1.

$$E(T_1) = f(\theta)+B_1$$

$$E(T_n) = f(\theta)+B_n.$$

Hence $$E(T_1-T_n) = B_1-B_n = (n-1)B_n.$$

Thus an unbiased estimator of B_n is given by

$$\hat{B}_n = \frac{T_1-T_n}{n-1}. \quad \dots \quad (5.2)$$

The variance of the estimator of $\hat{B}_n$ is given by

$$V(\hat{B}_n) = \frac{V(T_1)}{(n-1)^2}(a^2-2\rho a+1) \quad \dots \quad (5.3)$$

where $a^2 = V(T_n)/V(T_1)$, and ρ is the correlation coefficient between the estimators T_1 and T_n. For most of the sample designs a^2 and ρ will tend to 1 as the sample size increases and hence the variance of the bias estimator will tend to 0 as sample size increases. It may be observed that an unbiased estimator of the bias of T_1 is given by

$$\hat{B}_1 = \frac{n}{n-1}\,(T_1 - T_n). \qquad \ldots \quad (5.4)$$

6. (Almost) unbiased estimator

Since an unbiased estimator of the bias of the estimator T_n has been obtained in Section 5, the estimator T_n can be corrected for its bias, thereby obtaining an unbiased or almost unbiased estimator of $f(0)$ according as the third and higher degree terms in 'e' become 0 or not. In the latter case, the estimator is said to be almost unbiased since it is unbiased only to the second degree of approximation. The estimator corrected for its bias is given by

$$T_c = T_n - B_n = T_n - \frac{T_1 - T_n}{(n-1)} = \frac{nT_n - T_1}{(n-1)}. \qquad \ldots \quad (6.1)$$

It may be noted that this is the corrected estimator we get, even if we correct the estimator T_1 for its bias.

The variance of the corrected estimator is

$$V(T_c) = \frac{V(T_1)}{(n-1)^2}\,(n^2a^2 - 2\rho a + 1). \qquad \ldots \quad (6.2)$$

The gain in precision in using T_c instead of T_n is given by

$$G(T_c) = \frac{M_n - V(T_c)}{M_n} = 1 - \frac{n^2a^2 - 2n\rho a + 1}{(n-1)^2(a^2+z^2)} \qquad \ldots \quad (6.3)$$

where z^2 is the ratio of the square of the bias of T_1 to the variance of T_1. If the sub-sample size is large z^2 will be negligibly small. Neglecting z^2 in the above expression, we find that the gain in precision will be positive if

$$(2n-1)a^2 - 2n\rho a + 1 < 0$$

which will be true if 'a' lies between the roots of the equation

$$(2n-1)a^2 - 2n\rho a + 1 = 0. \qquad \ldots \quad (6.4)$$

For given values of a and ρ, the minimum value of n which makes the corrected estimator more efficient and the value of n which maximises the gain are respectively given by

$$\left[\frac{(1-a^2)}{2a(\rho - a)}\right] + 1 \qquad \ldots \quad (6.5)$$

and

$$\frac{(1-\rho a)}{a(\rho - a)} \qquad \ldots \quad (6.6)$$

TABLE 1. SHOWING THE MINIMUM AND MAXIMUM VALUES OF $G(T_c)$ AND THE CORRESPONDING VALUES OF n FOR DIFFERENT VALUES OF ρ AND a ($\rho > a$)

sr. no.	a	ρ	minimum		maximum	
			n	$G(T_c)$	n	$G(T_c)$
1	0.6	0.7	6	0.0089	10	0.0102
2		0.8	3	0.0556	4	0.0988
3		0.9	2	0.0880	3	0.3050
4	0.7	0.8	4	0.0113	7	0.0260
5		0.9	2	0.1020	3	0.1684
6	0.8	0.9	3	0.0469	4	0.0486

Source: Murthy, M. N. and Nanjamma, N. S. (1959): Almost unbiased ratio estimates based on interpenetrating sub-sample estimates, *Sankhyā*, 21, 381-392.

7. Illustrations

In this section, the results derived in the previous sections are applied to some particular cases.

Case (*i*) : $f(\theta) = \theta^k$. Let t be an unbiased estimator of θ based on any sample design. Then an estimator of $f(\theta)$ is given by

$$f(t) = t^k. \quad \dots \quad (7.1)$$

The bias and mean square error of $f(t)$ correct to the second degree of approximation are given by

$$B[f(t)] = \frac{1}{2}\, k(k-1)\, C^2 f(\theta) \quad \dots \quad (7.2)$$

$$M[f(t)] = k^2C^2[f(\theta)]^2 \quad \dots \quad (7.3)$$

where C^2 is the relative variance of $t[= V(t)/\theta^2]$, since

$$\frac{df}{dt} = kt^{k-1} \text{ and } \frac{d^2f}{dt^2} = k(k-1)t^{k-2}.$$

The bias relative to the mean square error is

$$\frac{B^2[f(t)]}{M[f(t)]} = \frac{1}{4}\,(k-1)^2C^2. \quad \dots \quad (7.4)$$

From (7.2) and (7.4) we see that the bias of $f(t)$ and its contribution to the mean square error both decrease as the sample size increases, since for most sample designs C^2 decreases with increase in sample size.

If t_s $(s = 1, 2, ..., n)$ are unbiased estimates of θ based on n independent interpenetrating sub-samples, the following two estimators T_1 and T_n of $f(\theta)$ can be considered.

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} t_s^k \quad \dots \quad (7.5)$$

and

$$T_n = \bar{t}^k, \left(\bar{t} = \frac{1}{n} \sum_{s=1}^{n} t_s \right). \quad \dots \quad (7.6)$$

We have seen that the bias of T_1 is n times that of the bias of T_n. Hence an unbiased estimator of the bias of T_n is given by

$$\hat{B}(T_n) = \frac{\sum_{s=1}^{n} t_s^k - n\bar{t}^k}{n(n-1)} \quad \dots \quad (7.7)$$

and the corrected estimator is given by

$$T_c = \frac{n^2\bar{t}^k - \sum_{s=1}^{n} t_s^k}{n(n-1)}. \quad \dots \quad (7.8)$$

It may be noted that the expression for bias and the corrected estimator will be completely unbiased if k in $f(\theta)$ is 2.

Case (ii) : *Correlation Coefficient* (ρ). The correlation coefficient between two characteristics x and y is

$$\rho = \frac{\text{cov}(x, y)}{\sqrt{V(x)\, V(y)}}. \quad \dots \quad (7.9)$$

In this case the parametric function is of the form

$$f(\theta) = \frac{\theta_1}{\sqrt{\theta_2 \theta_3}} \quad \dots \quad (7.10)$$

and the estimator is given by
$$f(t) = \frac{t_1}{\sqrt{t_2 t_3}} \quad \dots \quad (7.11)$$

where t_1, t_2 and t_3 are unbiased estimators of θ_1, θ_2 and θ_3 respectively. The bias and mean square error of $f(t)$ correct to the second degree of approximation are given by

$$B[f(t)] = \frac{f(\theta)}{8}[3(v_{22}+v_{33})-4(v_{12}+v_{13})+2v_{23}] \quad \dots \quad (7.12)$$

and

$$M[f(t)] = \frac{[f(\theta)]^2}{4}[4v_{11}+(v_{22}+v_{33})-4(v_{12}+v_{13})+2v_{23}] \quad \dots \quad (7.13)$$

where

$$v_{ij} = \frac{E(t_i-\theta_i)(t_j-\theta_j)}{\theta_i\theta_j}.$$

Let t_{is} $(i = 1, 2, 3)$ be unbiased estimates based on the s-th independent interpenetrating sub-sample $(s = 1, 2, ..., n)$. Then using the two estimators

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} \frac{t_{1s}}{\sqrt{t_{2s} t_{3s}}} \qquad \dots \quad (7.14)$$

$$T_n = \frac{\bar{t}_1}{\sqrt{\bar{t}_2 \bar{t}_3}} \qquad \dots \quad (7.15)$$

we get the following corrected estimator of ρ

$$T_c = \frac{n T_n - T_1}{(n-1)}. \qquad \dots \quad (7.16)$$

Case (iii) : *Regression Estimator.* Let y and x be unbiased estimators of the population totals Y and X respectively and let b be a consistent estimator of the regression coefficient obtained by taking the ratio of unbiased estimators of the covariance between x and y and the variance of x.

The regression estimator is $$\hat{y} = y + b(X - x). \qquad \dots \quad (7.17)$$

The estimator in this case is of the form

$$f(t) = t_1 + \frac{t_2}{t_3}(X - t_4). \qquad \dots \quad (7.18)$$

The bias and the mean square error of this estimator, correct to the second degree of approximation, are given by

$$B[f(t)] = \beta X(v_{34} - v_{24}) \qquad \dots \quad (7.19)$$

and $$M[f(t)] = V(y) - 2\beta \operatorname{cov}(x, y) + \beta^2 V(x). \qquad \dots \quad (7.20)$$

By defining the two estimators T_1 and T_n on the basis of n interpenetrating sub-sample estimates we get the corrected estimator as

$$T_c = \frac{nT_n - T_1}{(n-1)}.$$

Case (iv) : *Skewness* $(\beta_2 = \mu_4/\mu_2^2)$. The parametric function is of the form

$$f(\theta) = \theta_1/\theta_2^2$$

and an estimator of $f(\theta)$ is given by

$$f(t) = t_1/t_2^2$$

where t_1 and t_2 are unbiased estimators of θ_1 and θ_2 respectively. The bias and the mean square error of $f(t)$, correct to the second degree of approximation, are given by

$$B[f(t)] = \beta_2(3v_{22} - 2v_{12}) \qquad \dots \quad (7.21)$$

and $$M[f(t)] = \beta_2^2(v_{11} + 4v_{22} - 4v_{12}) \qquad \dots \quad (7.22)$$

where $$v_{ij} = E(t_i - \theta_i)(t_j - \theta_j)/\theta_i\theta_j.$$

Defining suitably the two estimators T_1 and T_n based on n interpenetrating sub-samples we get the corrected estimator, as before, as

$$\hat{\beta}_{2u}' = \frac{nT_n - T_1}{(n-1)}.$$

8. Estimation of bias

General Case. Suppose $f(\theta)$ is the parametric function of the parameters $(\theta_1, \theta_2, \ldots, \theta_k)$ and $f(t)$ is an estimator of $f(\theta)$ based on the estimators $(t_1, t_2, \ldots, t_k)$ which are unbiased for the parameters $(\theta_1, \theta_2, \ldots, \theta_k)$. Let $t_i = \theta_i + h_i$, $i = 1, 2, \ldots, k$. Applying Taylor series expansion to $f(t)$ about $t = \theta$ symbolically and neglecting terms of degrees greater than p in h_i's, we get

$$f(t) = f(\theta) + \sum_{j=1}^{p} \frac{1}{j!} \sum_{i_1, i_2, \ldots, i_j} (h_{i_1} h_{i_2} \ldots, h_{i_j}) \left(\frac{d^j f}{dt_{i_1} dt_{i_2} \ldots dt_{i_j}} \right)_{t=\theta} \quad \ldots \quad (8.1)$$

Taking the expected value of (8.1), we get the bias of $f(t)$ as

$$B[f(t)] = \sum_{j=2}^{p} \frac{1}{j!} \sum_{i_1, i_2, \ldots, i_j} \mu_j(i_1, i_2, \ldots i_j) \left[\frac{d^j f}{dt_{i_1} dt_{i_2} \ldots dt_{i_j}} \right]_{t=\theta}. \quad \ldots \quad (8.2)$$

Suppose t_{is} is an unbiased estimate of θ_i based on the s-th independent interpenetrating sub-sample ($i = 1, 2, \ldots, k$; $s = 1, 2, \ldots n$). Let us consider the following p estimators of $f(\theta)$

$$T_m = \frac{1}{\binom{n}{m}} \Sigma\, f(\bar{t}(m)), \quad m = 1, 2, \ldots p-1, n \quad \ldots \quad (8.3)$$

where $$\bar{t}\ (m) = (\bar{t}_1(m), \bar{t}_2(m), \ldots, \bar{t}_k(m)),$$

$\bar{t}_i(m)$ being the mean of the estimate t_i based on a combination of m sub-samples taken from the n independent interpenetrating sub-samples and Σ denotes summation over all combinations of m sub-samples formed out of n sub-samples.

The bias of T_m to the p-th degree of approximation is given by

$$B_m = B(T_m) = \frac{1}{\binom{n}{m}} \Sigma\, B[f(\bar{t}(m))]$$

$$= \frac{1}{\binom{n}{m}} \Sigma\, E \left[\sum_{j=2}^{p} \frac{1}{j!} \sum_{i_1, i_2, \ldots, i_j} (\bar{h}_{i_1} \bar{h}_{i_2} \ldots \bar{h}_{i_j}) \left(\frac{\partial^j f}{\partial t_{i_1} \partial t_{i_2} \ldots \partial t_{i_j}} \right) \right]_{t=\theta} \quad \ldots \quad (8.4)$$

where $\bar{h}_{i_r} = \frac{1}{m} \sum_{s=1}^{m} h_{i_r s}$. After simplification the bias of T_m may be expressed in the form

$$B_m = \sum_{j=2}^{p} \frac{A_j}{m^{j-1}} \quad (m = 1, 2, \ldots, p-1, n) \qquad \ldots \quad (8.5)$$

where A_j is a function of the j-th order moments and product moments of the estimators $(t_1, t_2, \ldots, t_k)$ and of terms of the form

$$\left[\frac{d^r f}{dt_{i_1} dt_{i_2} \ldots dt_{i_r}} \right]_{t=\theta}, \; (r \geqslant j).$$

From (8.5) we see that in the series of estimation $\{T_m\}$, $B(T_{m+1}) < B(T_m)$.

Since $$E(T_m) = f(\theta) + B_m,$$

we get $$E(T_1 - T_m) = B_1 - B_m = \sum_{j=2}^{p} \left(1 - \frac{1}{m^{j-1}}\right) A_j. \qquad \ldots \quad (8.6)$$

Let $D_m = (T_1 - T_m)$. The equation (8.6) can be written as

$$E(D) = A\,\Lambda, \qquad \ldots \quad (8.7)$$

where $$D = (D_2, D_3, \ldots, D_{p-1}, D_n)$$
$$A = (A_2, A_3, \ldots, A_{p-1}, A_p)$$

$$\Lambda = \begin{pmatrix} 1-\frac{1}{2} & 1-\frac{1}{3} & \ldots & 1-\frac{1}{p-1} & 1-\frac{1}{n-1} \\ 1-\frac{1}{2^2} & 1-\frac{1}{3^2} & \ldots & 1-\frac{1}{(p-1)^2} & 1-\frac{1}{(n-1)^2} \\ \\ 1-\frac{1}{2^{p-1}} & 1-\frac{1}{3^{p-1}} & \ldots & 1-\frac{1}{(p-1)^{p-1}} & 1-\frac{1}{(n-1)^{p-1}} \end{pmatrix}.$$

It may be noted that in (8.7) we have $(p-1)$ equations in $(p-1)$ unknowns. It may be observed that we are considering p estimators since there are $(p-1)A$'s and $f(\theta)$ to be estimated. Solving (8.7) for A we get

$$A = E(D)\,\Lambda^{-1}. \qquad \ldots \quad (8.8)$$

Taking $B = (B_2, B_3, \ldots, B_{p-1}, B_n)$, we get

$$B = A(e + \Lambda) \qquad \ldots \quad (8.9)$$

where 'e' is a $(p-1, p-1)$ matrix whose elements are all equal to 1. Substituting in (8.9) the solution for A obtained in (8.8), we get unbiased estimators of the biases of the estimators, namely,

$$\hat{B} = D\Lambda^{-1}e - D.$$

That is

$$\hat{B}_m = \sum_j D_j S_{j-1} - D_m, (j = 2, 3, \ldots, p-1, n), (m = 2, 3, \ldots, p-1, n), \ldots \quad (8.10)$$

$S_{n-1} = S_{p-1}$, where S_j is the sum of the elements in the j-th row of Λ^{-1}.

Particular Cases : (*i*) $p = 2$. This is the case considered earlier. In this case, the following 2 estimators of $f(0)$ may be considered.

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} f(t_s)$$

$$T_n = f(\bar{t}).$$

$$B_m = A/m, (m = 1, n).$$

Since $$\Lambda = \left(1 - \frac{1}{n}\right),\ \Lambda^{-1} = \frac{n}{n-1},\ S_1 = \frac{n}{n-1},$$

we get $$\hat{B}_n = \frac{n}{n-1} (T_1 - T_n) - (T_1 - T_n)$$

$$= (T_1 - T_n)/(n-1).$$

Case (*ii*) : $p = 3$.

Let us consider the following three estimators of $f(0)$.

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} f(t_s) \quad \ldots \quad (8.11)$$

$$T_2 = \frac{2}{n(n-1)} \sum f(\bar{t}(2)) \quad \ldots \quad (8.12)$$

$$T_n = f(\bar{t}). \quad \ldots \quad (8.13)$$

$$B_m = \frac{A_2}{m} + \frac{A_3}{m^2}, (m = 1, 2, n). \quad \ldots \quad (8.14)$$

Since $$\Lambda = \begin{pmatrix} 1 - \frac{1}{2} & 1 - \frac{1}{n} \\ 1 - \frac{1}{2^2} & 1 - \frac{1}{n^2} \end{pmatrix}$$

and $$\Lambda^{-1} = -\frac{4n^2}{(n-1)(n-2)} \begin{pmatrix} \frac{n^2-1}{n^2} & -\frac{n-1}{n} \\ -\frac{3}{4} & \frac{1}{2} \end{pmatrix},$$

we get after simplification

$$B_1 = \frac{n-2}{n-1}T_1 + \frac{4}{n-2}T_2 - \frac{n^2}{(n-1)(n-2)}T_n \quad \dots \quad (8.15)$$

$$B_2 = -\frac{1}{n-1}T_1 + \frac{n+2}{n-2}T_2 - \frac{n^2}{(n-1)(n-2)}T_n \quad \dots \quad (8.16)$$

$$B_n = -\frac{1}{n-1}T_1 + \frac{4}{n-2}T_2 - \frac{(3n-2)}{(n-1)(n-2)}T_n. \quad \dots \quad (8.17)$$

REFERENCES

GODAMBE, V. P. (1955): A unified theory of sampling from finite population. *J. Roy. Stat. Soc.*, 17(B), 269-278.

MIDZUNO, H. (1950): An outline of the theory of sampling systems. *Ann. Math. Stat.*, 1, 149-156.

MURTHY, M. N. and NANJAMMA, N. S. (1959): Almost unbiased estimates based on interpenetrating sub-sample estimates. *Sankhyā*, 21, 381-392.

MURTHY, M. N., NANJAMMA, N. S. and SETHI, V. K. (1959): Some sampling systems providing unbiased ratio estimators. *Sankhyā*, 21, 299-314.

MURTHY, M. N. (1962): Generalized unbiased estimation in sampling from finite populations. Submitted for publication.

Paper received : February, 1962.