Journal of the Indian Statistical Association
Vol. 26, 91—101

A GENERALIZED METHOD OF ESTIMATION IN DOUBLE SAMPLING

T. P. Tripathi
Indian Statistical Institute, Calcutta, India.

H. P. Singh
J. N. Agricultural University, Jabalpur, India.

L. N. Upadhyaya
Indian School of Mines, Dhanbad, India.

Abstract

A general theory of double sampling for estimating a parameter $\varphi(y)$, utilizing information about a supplementary parameter $\varphi(x)$ obtained inexpensively through a preliminary large sample, has been presented. The general results are then used to deal with the problem of estimating variance σ_y^2 of a character y, in case of bivariate populations, where the information on the other character x, collected through the first phase sample, is expected to improve the usual unbiased estimator of σ_y^2. A wide class of estimators is discussed, an asymptotically optimum subclass is identified and the estimators based on two phase sampling are compared with the usual unbiased estimator of σ_y^2 in case of single phase sampling, under a linear cost function. In particular the results are derived for bivariate normal populations.

1. Introduction

Positive role of auxiliary information in obtaining sampling strategies with some optimal properties is now undisputed in the sphere of sample surveys. In the case of bivariate populations, the use of information on an auxiliary character x for estimating the mean of other (principal) character y has been discussed widely by various authors. Recently the problem of estimating population variance σ_y^2 using information on x has been discussed by Das and Tripathi (1978), and Srivastava and Jhaji (1980). Further, problem of estimating coefficient of variation C_y using information on x has been discussed by

Das and Tripathi (1981), and Tripathi, Upadhyaya and Singh (1987). Dealing with the general problem of estimating a parameter $\varphi(y)$ based on character y in case an ancillary parameter $\varphi(x)$ based on x is known, one may define, following Das and Tripathi (1980), a class of estimators

$$d^* = \frac{\hat{\varphi}(y) - t_1(\hat{\varphi}(x) - \varphi(x))}{[\hat{\varphi}(x) - t_2(\hat{\varphi}(x) - \varphi(x))]^{\alpha}} [\varphi(x)]^{\alpha} \quad \cdots \quad (1 \cdot 1)$$

for $\varphi(y)$ in case of general sampling designs, where $\hat{\varphi}(y)$ and $\hat{\varphi}(x)$ are estimators of $\varphi(y)$ and $\varphi(x)$ respectively, α is a suitably chosen constant and t_i $(i = 1, 2)$ are suitably chosen statistics (which in particular may be constants). The class of estimators defined in $(1 \cdot 1)$ requires apriori knowledge of $\varphi(x)$. In the situation where $\varphi(x)$ is not known, the above class of estimators cannot be used in practice. However if collection of information on x is not very costly two phase sampling or double sampling procedure may be used to extend the class in $(1 \cdot 1)$ for estimating $\varphi(y)$ even when $\varphi(x)$ is unknown.

In this paper a general double sampling procedure is developed for estimating $\varphi(y)$ in case $\varphi(x)$ is unknown but information on x is expected to improve usual estimator $\hat{\varphi}(y)$ of $\varphi(y)$. The peneral properties of the class are studied and the results are then in particular. obtained for estimating σ_y^2.

2. A general double sampling scheme for estimating any Parameter using supplementry information

We confine ourselves to fixed sample size designs. Let a preliminary large sample of size n be selected at moderate cost according to a specified sampling procedure and observations only on the auxiliary character x be collected. Further let a subsequent small sample of say m units be selected according to another specified sampling procedure and the observations on principal character y be made. The second phase sample may either be a subsample of the first or independent of the first in which case both x and y are to be observed. Let $E_1(\cdot)$, $V_1(\cdot)$, $C_1(\cdot)$ denote the unconditional and $E_2(\cdot)$, $V_2(\cdot)$, $C_2(\cdot)$ denote the conditional (given the first sample) mean, variance and convariance. Let $\varphi_{(1)}(x)$ be an unbiased estimator of $\varphi(x)$ based on the observations of the first phase sample only and $\varphi_{(2)}(x)$ and $\varphi_{(2)}(y)$ be the unbiased estimators of $\varphi(x)$ and $\varphi(y)$ respectively based mainly on the observations of the second phase sample. For the case of subsample, we shall further assume that

$$E_2(\varphi_{(2)}(x)) = \varphi_{(1)}(x).$$

We note that in both the cases, of subsample and independent samples,

$$C_2(\varphi_{(2)}(x), \varphi_{(1)}(x)) = C_2(\varphi_{(2)}(y), \varphi_{(1)}(x)) = 0. \quad \cdots \quad (2 \cdot 1)$$

Further, in case of independent samples if $\varphi_{(2)}(x)$ and $\varphi_{(2)}(y)$ do not use any information obtained on the first phase sample, then

$$\text{Cov}\,(\varphi_{(2)}(x), \varphi_{(1)}(x)) = \text{Cov}\,(\varphi_{(2)}(y), \varphi_{(1)}(x)) = 0.$$

But in general $\varphi_{(2)}(y)$ and $\varphi_{(2)}(x)$ may utilize the information on the first phase sample in the case of independent samples too, thus

$$\text{Cov}\,(\varphi_{(2)}(x), \varphi_{(1)}(x)) \quad \text{and} \quad \text{Cov}\,(\varphi_{(2)}(y), \varphi_{(1)}(x))$$

would not be zero for some sampling schemes.

We extend the generalized method of estimation in (1·1) to the class of estimators for $\varphi(y)$ defined by

$$d = \frac{\varphi_{(2)}(y) - t(\varphi_{(2)}(x) - \varphi_{(1)}(x))}{[t^{*}\varphi_{(1)}(x) + (1 - t^{*})\,\varphi_{(2)}(x)]^{\alpha}}[\varphi_{(1)}(x)]^{\alpha}, \quad \cdots \quad (2\cdot2)$$

where α is a suitably chosen constant and t and t^{*} are suitably chosen statistics such that their means exist. Regression type estimator (e. g. $\alpha = 0$), ratio and product type estimator (e. g. $t = 0$, $t^{*} = 0$ and $\alpha = +1$ or -1), regression cum ratio estimators (e. g. $t = 0$, $\alpha = 1$) and large number of other interesting estimators may be derived as particular members of the class.

In order to investigate properties of the class d in (2·2) we impose the following restrictions on the choises of t and t^{*} :

(*i*) in case t and t^{*} are not constants, they are such that

$$Et = T + O(m^{-r}), \quad r > 0,$$

$$Et^{*} = T^{*} + O(m^{-s}), \quad s > 0,$$

where T and T* are the constants (parameters) not depending on m.

(ii) $V(t)$, $V(t^{*})$ and $\text{Cov}\,(t, t^{*})$ are $O(m^{-1})$, covariance between t (or t^{*}) and any of the $\varphi_{(1)}(x)$, $\varphi_{(2)}(x)$ and $\varphi_{(2)}(y)$ are $O(m^{-1})$, and all the higher order moments are of the order $m^{-(1+r)}$ where $r > 0$.

(iii) t^{*} is such that

$$\left|\frac{(1 - t^{*})\,\varphi_{(2)}(x) + t^{*}\,\varphi_{(1)}(x) - \varphi(x)}{\varphi(x)}\right| < 1$$

Under the conditions (i), (ii) and (iii), the biases and mean squared errors (MSEs) of the class of estimators in d to the terms of order $O(n^{-1})$, are given by

$$B(d) = \varphi(y)\Big[\frac{\alpha(\alpha+1)}{2}\Big\{(1-T^*)^2 C^2(\varphi_{(2)}(x)) + T^{*2} C^2(\varphi_{(1)}(x))$$
$$+ 2(T^*)(1-T^*)\, C(\varphi_{(1)}(x), \varphi_{(2)}(x))\Big\} + \frac{\alpha(\alpha-1)}{2} C^2(\varphi_{(1)}(x))$$
$$- \alpha^2\big\{(1-T^*)\, C(\varphi_{(1)}(x), \varphi_{(2)}(x)) + (T^*)\, C^2(\varphi_{(1)}(x))\big\}$$
$$- (\varphi(x)/\varphi(y))\, \alpha(1-T^*)\big\{C(\varphi_{(2)}(y), \varphi_{(2)}(x)) - C(\varphi_{(2)}(y), \varphi_{(1)}(x))\big\}$$
$$+ \alpha(T)(1-T^*)(\varphi(x)/\varphi(y))\big\{C^2(\varphi_{(1)}(x)) + C^2(\varphi_{(2)}(x))$$
$$- 2\, C(\varphi_{(1)}(x), \varphi_{(2)}(x))\big\} - (\varphi(x)/\varphi(y))T\big\{C(\varphi_2(x), t) - C(\varphi_1(x), t)$$
$$+ \alpha T^*\big\{C(\varphi_2(x), t^*) - C(\varphi_1(x), t^*)\big\}\Big] \qquad \cdots \quad (2{\cdot}3)$$

and
$$M(d) = \varphi^2(y)\big[C^2(\varphi_{(2)}(y)) + Q^2\big\{C^2(\varphi_{(1)}(x)) + C^2(\varphi_{(2)}(x))$$
$$- 2C(\varphi_{(1)}(x), \varphi_{(2)}(x))\big\} - 2Q\big[C(\varphi_{(2)}(y), \varphi_{(2)}(x)) - C_{(2)}(y), \varphi_{(1)}(x))\big\}\big] \qquad \cdots \quad (2{\cdot}4)$$

respectively, where $C^2(\varphi_{(i)}(x)) = V(\varphi_{(i)}(x)) / \varphi^2(x), \quad i = 1, 2$

$$C(\varphi_{(i)}(x), t^*) = \text{Cov}(\varphi_{(i)}(x), t^*) / \varphi(x)\, T^*, \quad (i = 1, 2)$$

$$C(\varphi_{(i)}(x), t) = \text{Cov}(\varphi_{(i)}(x), t)/(T)\, \varphi(x);\ C\,\varphi_{(2)}(y), \varphi_{(1)}(x))$$
$$= \frac{\text{Cov}(\varphi_{(2)}(y), \varphi_{(1)}(x))}{\varphi(y)\, \varphi(x)} \quad (i = 1, 2)$$

$$C(\varphi_{(1)}(x), \varphi_{(2)}(x)) = \text{Cov}(\varphi_{(1)}(x), \varphi_{(2)}(x))/\varphi^2(x);$$
$$C^2(\varphi_{(2)}(y)) = V(\varphi_{(2)}(y)) / \varphi^2(y),$$
$$\text{and} \quad Q = \alpha(1 - T^*) + \{\varphi(x)/\varphi(y)\}\, T.$$

The optimum value of (α, T, T^*) which minimizes $M(d)$ in (2·4) is given by

$$Q = \frac{\varphi(x)}{\varphi(y)}\, \frac{[\text{Cov}(\varphi_{(2)}(y), \varphi_{(2)}(x)) - \text{Cov}(\varphi_{(2)}(y), \varphi_{(1)}(x))]}{[V(\varphi_{(1)}(x)) + V(\varphi_{(2)}(x)) - 2\,\text{Cov}(\varphi_{(1)}(x), \varphi_{(2)}(x))]}$$
$$= \frac{\varphi(x)}{\varphi(y)}\, \beta_{\varphi_{(2)}, u(y)} \qquad = Q_0 \text{ (say)} \qquad \cdots \quad (2{\cdot}5)$$

where $\beta_{\varphi_{(2)}, u(y)}$ is the regression coefficient of $\varphi_{(2)}(y)$ on $u = \varphi_{(2)}(x) - \varphi_{(1)}(x)$. Hence, the resulting (optimum) MSE of d is given by

$$M_0(d) = V(\varphi_{(2)}(y)) - Q_0^2\, (\varphi(y)/\varphi(x))^2 \Big\{V(\varphi_{(1)}(x))$$
$$+ V(\varphi_{(2)}(x)) - 2\,\text{Cov}(\varphi_{(1)}(x), \varphi_{(2)}(x))\Big\} \qquad \cdots \quad (2{\cdot}6)$$

Remarks

(1) In case of subsamples,

$$M(d) = V_1(E_2\varphi_{(2)}(y)) + E_1 V_2(\varphi_{(2)}(y)) + Q^2(\varphi(y)/\varphi(x))^2[E_1 V_2(\varphi_{(2)}(x))]$$
$$- 2Q(\varphi(y)/\varphi(x))\,[E_1 C_2(\varphi_{(2)}(x))] \quad \cdots \quad (2\cdot7)$$

$$Q_0 = \left[E_1 C_2(\varphi_{(2)}(y), \varphi_{(2)}(x)) / E_1 V_2(\varphi_{(2)}(x))\right]\frac{\varphi(x)}{\varphi(y)} \quad \cdots \quad (2\cdot8)$$

$$M_0(d) = V_1(E_2\varphi_{(2)}(y) + E_1 V_2(\varphi_{(2)}(y))$$
$$- Q_0^2(\varphi(y)/\varphi(x))^2\left\{E_1 V_2(\varphi_{(2)}(x))\right\} \quad \cdots \quad (2\cdot9)$$

(ii) In case of independent samples, the expression in (2·4) to (2·5) may be used directly in conjugation with (2·1).

(iii) Let $\alpha = 0$ then in case of subsamples, the regression type estimator

$$d_0 = \varphi_{(2)}(y) - t(\varphi_{(2)}(x) - \varphi_{(1)}(x)) \quad \cdots \quad (2\cdot10)$$

with

$$T = E_1 C_2(\varphi_{(2)}(y), \varphi_{(2)}(x)) / E_1 V_2(\varphi_{(2)}(x))$$
$$= \beta_{\varphi_{(2)}(y),\ \varphi_{(2)}(x)}$$

would be an asymptotically optimum estimator (AOE) in the class d with $M(d_0) = M_0(d)$.

In particular one may choose

$$t = \hat{\beta}_{\varphi_{(2)}(y),\, \varphi_{(2)}(x)} = \hat{C}_2\,\varphi_{(2)}(y)\,\varphi_{(2)}(x)) / \hat{V}_2(\varphi_{(2)}(x))$$

in d_0 provided coefficients of variation of $\hat{V}_2(\varphi_{(2)}(x))$ and $\hat{C}_2(\varphi_{(2)}(y), \varphi_{(2)}(x))$, unbiased estimators of $V_2(\varphi_{(2)}(x))$ and $C_2(\varphi_{(2)}(y), \varphi_{(2)}(x))$ respectively, are of the order m^{-r}, $r > 0$.

(iv) One may define another class of estimators for φ_0 as

$$e = g(\varphi_{(2)}(y), \varphi_{(2)}(x), \varphi_{(1)}(x))$$

where $g(\cdot)$ is function satisfying certain conditions similar to those given by Srivastava and Jhajii (1980).

It may be shown that none of the estimators in e have MSE smaller than $M_0(d)$ in (2·6). Thus AOE in the class d, in particular d_0 in (2·10), would also be AOE in the class obtained by the union of classes d and e.

(v) Results for regression and ratio estimators, for population mean $\varphi(y) = \bar{Y}$ in case $\varphi(x) = \bar{X}$ is unknown, based on double sampling [(Cochran (1977) p. 339. p. 343] and numerous other estimators may be easily derived from above general results.

3. Application of general theory to the estimation of population variance using double sampling

In bivariate populations, the problem of estimating the population variance σ_y^2 of character y in case σ_x^2 the population variance of another (auxiliary) character x, is known has been discussed by Das and Tripathi (1978), and Srinivastava and Jhajj (1980). However in many situations of practical importance σ_x^2 may not be known while the use of auxiliary information on x may be expected to improve the usual unbiased estimator s_y^2 of σ_y^2. In such situations, one may obviously resort to double sampling provided collection of information on x is not very costly. One may define estimators for σ_y^2 as

$$d_1 = s^2_{y(2)} - \lambda_1 \left(s^2_{x(2)} - s^2_{x(1)} \right) > 0,$$

$$d_2 = s^2_{y(2)} \left(s^2_{x(1)} / s^2_{x(2)} \right)^{\alpha}$$

$$d_3 = s^2_{y(2)} \; s^2_{x(1)} \left[s^2_{x(1)} \lambda_2 + (1 - \lambda_2) \, s^2_{x(2)} \right]^{-1},$$

$$d_4 = (1 - \lambda_3) \, s^2_{y(2)} + \lambda_3 s^2_{y(2)} \left(s^2_{x(2)} / s^2_{x(1)} \right),$$

$$d_5 = (1 + \lambda_4) \, s^2_{y(2)} - \lambda_4 \, s^2_{y(2)} \left(s^2_{x(2)} / s^2_{x(1)} \right),$$

$$d_6 = s^2_{y(2)} \left(s^2_{x(1)} / s^2_{x(2)} \right).$$

$$d_7 = s^2_{y(2)} \left(s^2_{x(2)} / s^2_{x(1)} \right),$$

in case of general double sampling scheme, where λ_i ($i = 1$ to 4) and α are suitably chosen constants. $s^2_{x(1)}$ is an unbiased estimator of σ_x^2 based on first phase sample of size n and $s^2_{y(2)}$ and $s^2_{x(2)}$ are unbiased estimators of σ^2_y and σ^2_x respectively based on second phase sample of size m. Obviously all these estimators are particular members of a very wide class of estimators

$$\hat{\sigma}^2_y = \frac{s^2_{y(2)} - t\left(s^2_{x(2)} - s^2_{x(1)}\right)}{\left[t^* s^2_{x(1)} + (1-t^*)\, s^2_{x(2)}\right]^{\alpha}} \cdot \left[s^2_{x(1)}\right]^{\alpha} \quad \cdots \quad (3\cdot 1)$$

The results about $\hat{\sigma}^2_y$ may be derived on the line of discussion in section 2 and the results about d^1 to d^4 may in turn be derived from the results about $\hat{\sigma}^2_y$. However to bring out the salient features of discussion we shall confine ourselves to SRSWOR at both the phases in which case

$$s^2_{x(1)} = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 ;\; s^2_{y(2)} = (m-1)^{-1} \sum_{i=1}^{m} (y_i - \bar{y}_m)^2 ,$$

$$s^2_{x(2)} = \sum_{i=1}^{m} (x_i - \bar{x}_m)^2 \Big/ (m-1)$$

$\bar{x}_n$, $\bar{y}_m$, $\bar{x}_m$ being sample means. In the following discussion we shall retain terms upto the order n^{-1}, and neglect sampling fractions, n/N and m/N, where N is the population size in case of finite populations.

Using the general results in (2·4) to (2·9) we have

Theorem 3·1. If both, the first phase and second phase, samples are simple random samples, the large sample approximations to the MSE of $\hat{\sigma}^2_y$ in case of subsamples and independent samples are given by

$$M\left(\hat{\sigma}^2_y\right) = (M_1/n) + (M_2/m), \quad \cdots \quad (3\cdot 2)$$

and

$$M^*\left(\hat{\sigma}^2_y\right) = \left(M_1^* / n\right) + (M_2/m) \quad \cdots \quad (3\cdot 3)$$

respectively, where $M_1 = [\,2(C_{22}(y, x) - 1) - Q(\beta_2(x) - 1)\,]\,Q\,\sigma^4_y$,

$$M_2 = \left[\beta_2(y) - 1)\sigma^4_y - M_1\right], \quad M_1^* = (\beta_2(x) - 1\; Q^2\, \sigma^4_y\, .$$

$Q = \alpha(1 - T^{*}) + \left(\sigma_x^2 / \sigma_y^2\right) T$, T and T* being defined in Section 2;

$$\beta_2(y) = \mu_4(y)/\sigma_y^4,\ \beta_2(x) = \mu_4(x)/\sigma_x^4,\ C_{22}(y, x) = \frac{\mu_{22}(x, y)}{\sigma_y^2\ \sigma_x^2} \text{ and}$$

$\mu_{rq}(y, x)$ being the (r, q)-th central product moment and $\mu_r(y)$ the rth central moment. The MSEs in (3·2) and (3·3) are minimized for $Q = Q_0$ and $Q = Q_0^{*}$ respectively, where

$$Q_0 = \frac{(C_{22}(y, x) - 1)}{(\beta_2(x) - 1)} = \frac{\rho^{*}\sqrt{\beta_2(y) - 1}}{\sqrt{\beta_2(x) - 1}} \qquad \cdots \quad (3\cdot4)$$

and $$Q_0^{*} = \frac{n}{(m + n)} \cdot \frac{\rho^{*}\sqrt{\beta_2(y) - 1}}{\sqrt{\beta_2(x) - 1}} \qquad \cdots \quad (3\cdot5)$$

ρ^{*} being the correlation coefficient between $(y - \bar{y})^2$ and $(x - \bar{X})^2$. Hence the resulting (optimum) MSEs of $\hat{\sigma}_y^2$ in case of subsamples and independent samples would be

$$M_0\left(\hat{\sigma}_y^2\right) = \left(\sigma_y^4 / m\right)\left[\beta_2(y) - 1 - \frac{(n - m)}{n} \cdot \frac{(C_{22}(y, x) - 1)^2}{(\beta_2(x) - 1}\right], \qquad \cdots \quad (3\cdot6)$$

and

$$M_0^{*}\left(\hat{\sigma}_y^2\right) = \left(\sigma_y^4 / m\right)\left[(\beta_2(y) - 1) - \frac{n}{(m + n)} \cdot \frac{(C_{22}(y, x) - 1)^2}{(\beta_2(x) - 1)}\right], \qquad \cdots \quad (3\cdot7)$$

respectively.

It is found that in case of bivariate normal population

$$M_1 = 2\sigma_y^4\, Q\,(2\rho^2 - Q),\ M_2 = \left(2\sigma_y^4 - M_1\right),\ M_1^{*} = 2\sigma_y^4\, Q^2$$

$$Q_0 = \rho^2,\ Q_0^{*} = \frac{0}{(m + n)}\rho^2,$$

$$M_0\left(\sigma_y^2\right) = \left(2\sigma_y^4 / m\right)\left[1 - \left(\frac{n - m}{n}\right)\rho^4\right]. \qquad \cdots \quad (3\cdot8)$$

$$M_0^{*}\left(\sigma_y^2\right) = \left(2\sigma_y^4 / m\right)\left[1 - \left(\frac{n}{m + n}\right)\rho^4\right]. \qquad \cdots \quad (3\cdot9)$$

where ρ is the correlation coefficient between y and x.

Remarks : (1) In case of subsamples, to our order of approximation, the estimator

$$d_6 = s^2_{y(2)} - t_0 \left(s^2_{x(2)} - s^2_{x(1)} \right)$$

$$\text{with} \quad t_0 = \frac{\left(m_{22}(y, x) - s^2_{y(2)} s^2_{x(2)} \right)}{\left(m_4(x) - s^4_{x(2)} \right)}$$

in general and with $t_0 = r^2 \, s^2_{y(2)} \Big/ s^2_{x(2)}$ in particular for bivariate normal populations would be AOE in the class (3·1), where

$$m_{22}(y, x) = \sum_{i=1}^{m} (y_i - \bar{y}_m)^2 \, (x_i - \bar{x}_m)^2 / m,$$

$$m_4(x) = \sum_{i=1}^{m} (x_i - \bar{x} \;)^4 / m \quad \text{and}$$

r is the sample correlation coefficient between y and x based on second sample.

(ii) In case of bivariate normal population with second phase sample as a subsample of the first phase, the optimum values of α and λ_i ($i = 1, 2, 3, 4$) in the estimators d_1 to d_5 would be given by

$$\lambda_{01} = \rho^2 \left(\sigma^2_y \Big/ \sigma^2_x \right) \text{ and } \alpha_0 = \lambda_{0i} = \rho^2, \; (i = 2, 3, 4).$$

4. Two Phase Sampling Versus Single Phase Sampling

Let the cost function be of the form

$$C = a_0 + n\, C_1 + m\, C_2 \,. \qquad \cdots \quad (4\cdot1)$$

where C is the total cost, a_0 is overhead cost, and C_1 and C_2 are the cost per unit for first and second phase samples respectively.

Following Cochran ([1977) p. 341] it is easily found that the value of n and m, which minimize MSE in (3·2) for fixed budget C in (4·1), are given by

$$n_0 = (C - a_0) \sqrt{M_1} \,/\, [\sqrt{C_1} \, (\sqrt{M_2 C_2} + \sqrt{M_1 C_1} \,)]$$

and $\cdots \quad (4\cdot2)$

$$m_0 = (C - a_0) \sqrt{M_2} \,/\, [\sqrt{C_2} \, (\sqrt{M_2 C_2} + \sqrt{M_1 C_1} \,)]$$

The resulting MSE of $\hat{\sigma}_y^2$ would be

$$M_0\left(\hat{\sigma}_y^2\right) = \left(\frac{y}{\sqrt{M_1C_1}} + \sqrt{M_2 C_2}\right)^2 \Big/ (C - a_0 \qquad \cdots \quad (4\cdot3)$$

Instead of double sampling for collecting information on x if all resources are devoted to observe y alone an unbiased estimator for σ_y^2 would be

$$s_y^2 = \sum_{i=1}^{n'} (y_i - \bar{y}_{n'})^2 \Big/ (n' - 1), \qquad \cdots \quad (4\cdot4)$$

where size n' of the single sample is given by $C = a_0 + n' C_2$.

Obviously

$$V\left(s_y^2\right) = \frac{\sigma_y^4\ (\beta_2(y) - 1)\ C_2}{(C - a_0)}. \qquad \cdots \quad (4\cdot5)$$

From (4·3) and (4·5) the relative efficiency of $\hat{\sigma}_y^2$ over s_y^2 is given by

$$RE\left(\hat{\sigma}_y^2 \cdot s_y^2\right) = 1/\left[\,1 - \delta + \delta(C_1/C_2)\,\right]^2 \qquad \cdots \quad (4\cdot5)$$

where

$$\delta = M_1 / (\beta_2(y) - 1)\ \sigma_y^4\ .$$

Hence, double sampling will lead to gain in precision if

$$C_1/C_2 < \delta / [\,1 + \sqrt{1 - \delta}\,]^2 = [\,1 - \sqrt{1 - \delta}\,]^2\, \delta. \qquad \cdots \quad (4\cdot7)$$

In case of bivariate normal population, with optimum choice of Q, viz. $Q_0 = \rho^2$, the inequality in (4·7) reduces to

$$C_1 / C_2 < \frac{\rho^4}{[1 + \sqrt{(1-\rho^4)}]^2} = \frac{[\,1 - \sqrt{(1-\rho^4)}\,]^2}{\rho^4} \qquad \cdots \quad (4\cdot8)$$

From (4·7) we infer that to obtain gain in precision by using double sampling over single phase sampling the ratio of the cost per unit in the first phase sample to the cost per unit in the second sample must must not exceed a critical value.

The form of R. E. $\left(\hat{\sigma}_y^2,\ s_y^2\right)$ in (4·6) is similar to that of relative efficiency of usual regression estimator for population mean in double sampling over sample mean and hence following Cochran (1977, pp. 342) the relative efficiency curves can be drawn easily with varying values of δ for certain values of (C_1/C_2).

References

Cochran, W. G. (1977). *Sampling Techniques*. Third edition. John Wiley and Sons, Inc New York.

Das A. K. and Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance, *Sankhya*, *C.* **40**, 139–148

Das, A. K. and Tripathi, T. P. (1980). Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known. *Sankhya*, **42**, C, pp. 76–86.

Das, A. K. and Tripathi, T. P. (1981). A class of estimators for coefficient of variation using knowledge on the C. V. of an auxiliary character. Abstract. Jour. Ind. Soc. Agri. Stat., 33, 1981. Tech. Report No. 12/81, Stat-Math., ISI, Calcutta.

Srivastava, S. K. and Jhajj, H. S. (1980). A class of estimators using auxilary information for estimating finite population variance. *Sankhya*, *C.* **42**, 87–96

Tripathi, T. P., Singh, H. P. and Upadhyaya, L. N. (1987). A general method of estimation and its application to estimation of coefficient of variation.