Biometrika (1989), 76, 4, pp. 789–94
Printed in Great Britain

Optimal estimation of finite population total under a general correlated model

BY RAHUL MUKERJEE
Stat-Math Division, Indian Statistical Institute, Calcutta- 700 035, India
AND S. SENGUPTA
Department of Statistics, University of Calcutta, Calcutta- 700 019, India

SUMMARY

Restricting attention to fixed size sampling designs and linear unbiased estimators of a finite population total, we give methods for finding estimators with minimum model expected variance and the optimal strategy under a general correlated superpopulation model. Some techniques popular in the theory of optimal experiments help in the derivation. Several earlier optimality results are deduced as special cases.

Some key words: Directional derivative; Finite population; Fixed size sampling design; Linear unbiased estimator; Optimal strategy; Superpopulation model.

1. INTRODUCTION

Let U be a finite population of N units labelled $i = 1, \ldots, N$, and y be a real variable assuming value Y_i on unit i. The problem is to estimate the population total $Y = \Sigma Y_i$ on the basis of a sample, i.e. a subset s of U, drawn according to a sampling design p with positive inclusion probability π_i for every unit i. We consider a superpopulation model consisting of prior distributions α such that

$$E_\alpha(Y_i) = \mu_i, \quad E_\alpha\{(Y_i - \mu_i)(Y_j - \mu_j)\} = v_{ij}, \tag{1·1}$$

where E_α and E_p denote expectations with respect to α and p respectively. Let P_n denote the class of designs p with fixed sample size n, and let L_n denote the class of linear unbiased estimators

$$e = a_s + \textstyle\sum_{i \in s} b_{si} Y_i \tag{1·2}$$

based on p, where the a_s and b_{si}'s are real constants satisfying

$$E_p(a_s) = \textstyle\sum_s a_s p(s) = 0, \quad \sum_{s \ni i} b_{si} p(s) = 1 \quad (i = 1, \ldots, N), \tag{1·3}$$

Σ_s denoting the sum over all s. Writing H_n as the class of strategies (p, e) with $p \in P_n$ and $e \in L_n$, we derive the optimal strategy in H_n under the model (1·1), in the sense of rendering the model expected variance $E_\alpha E_p\{(e - Y)^2\}$ a minimum for every α. The optimal strategy is generally found to depend on $\mu = (\mu_1, \ldots, \mu_N)'$ and $V = ((v_{ij}))$, which is assumed to be positive-definite.

It may be noted that (1·1) is a generalization of the models considered by Godambe (1955), Hájek (1959), Cassel, Särndal & Wretman (1976) and Tam (1984) and that the earlier optimality results obtained by these authors can be deduced as special cases. The results in this paper also give a method for finding the minimum model expected variance, under the general model (1·1), and hence may be found useful in studying the robustness of a strategy in H_n.

2. OPTIMALITY RESULTS

Consider $(p, e) \in H_n$ and let b_s be a $n\times 1$ vector with its elements b_{si} $(i \in s)$; let V_s be a $n \times n$ submatrix of V obtained by considering the units $i \in s$ and 1 be a $N \times 1$ vector with all elements unity. By (1·3), it is easy to verify that

$$E_\alpha E_p\{(e-Y)^2\} = \sum_s \left(a_s - \sum_{i=1}^N \mu_i + \sum_{i\in s} b_{si}\mu_i\right)^2 p(s) + \sum_s b_s' V_s b_s p(s) - 1'V1$$

$$\geq \textstyle\sum_s b_s' V_s b_s p(s) - 1'V1 \qquad (2\cdot 1)$$

with equality if and only if

$$a_s = \sum_{i=1}^N \mu_i - \sum_{i\in s} b_{si}\mu_i, \qquad (2\cdot 2)$$

for every s with $p(s) > 0$.

Let $V_s^{-1} = ((v_s^{ij}))$. Define for $i, j = 1, \ldots, N$,

$$\phi_{ij} = \textstyle\sum_{s \ni ij} v_s^{ij} p(s) \qquad (2\cdot 3)$$

and Φ as the $N \times N$ matrix with its elements ϕ_{ij}. Since $\pi_i > 0$ for every i, it can be seen that Φ is nonsingular. This is because for any $w = (w_1, \ldots, w_N)'$, $w'\Phi w \geq 0$ with equality if and only if $w_i = 0$ for all $i \in s$, this being true for every s such that $p(s) > 0$; compare with (3·1). Let

$$\lambda = (\lambda_1, \ldots, \lambda_N)' = \Phi^{-1}1, \qquad (2\cdot 4)$$

λ_s being a $n \times 1$ subvector of λ given by the elements $i \in s$ and

$$b_s^* = V_s^{-1}\lambda_s \qquad (2\cdot 5)$$

with its elements b_{si}^* $(i \in s)$. From (1·3), (2·3)–(2·5),

$$\textstyle\sum_s b_s^{*\prime} V_s b_s^* p(s) = \sum_s \lambda_s' V_s^{-1} \lambda_s p(s) = \lambda'\Phi\lambda = 1'\Phi^{-1}1, \qquad (2\cdot 6)$$

$$\textstyle\sum_s b_s' V_s b_s^* p(s) = \sum_s b_s' \lambda_s p(s) = 1'\lambda = 1'\Phi^{-1}1. \qquad (2\cdot 7)$$

In view of (2·1), (2·2), (2·6), (2·7), we obtain

$$E_\alpha E_p(e-Y)^2 \geq \textstyle\sum_s (b_s - b_s^*)' V_s (b_s - b_s^*) p(s) + 1'\Phi^{-1}1 - 1'V1 \geq 1'\Phi^{-1}1 - 1'V1 \qquad (2\cdot 8)$$

with equality if and only if (2·2) holds and further

$$b_s = b_s^* \qquad (2\cdot 9)$$

for every s with $p(s) > 0$. Note that the choice given by (2·2) and (2·9) is consistent with (1·3) since, by (2·3)–(2·5),

$$\sum_{s \ni i} b_{si}^* p(s) = \sum_{s \ni i} \sum_{j \in s} v_s^{ij} \lambda_j p(s) = \sum_{j=1}^N \lambda_j \phi_{ij} = 1.$$

Thus for a given p, the optimal estimator in L_n, under the model (1·1), is given by (2·2) and (2·9). The optimal design can now be obtained by minimizing the right-hand side of (2·8), or equivalently $1'\Phi^{-1}1$ with respect to $p \in P_n$. This is considered in § 3. The results so far obtained can be summarized as follows.

THEOREM 1. *For a given $p \in P_n$, under the superpopulation model* (1·1),

$$E_\alpha E_p(e-Y)^2 \geq 1'\Phi^{-1}1 - 1'V1$$

for every $e \in L_n$, with equality if and only if $e = e^*$, where e^* is specified by (2·2) and (2·9). Further, a strategy (p, e) is optimal in H_n provided $(p, e) = (p^*, e^*)$, where p^* is a sampling design that minimizes $1'\Phi^{-1}1$ with respect to $p \in P_n$.

Consider now a special case of (1·1) where, for $1 \leq i \neq j \leq N$, $v_{ij} = \rho(v_{ii}v_{jj})^{\frac{1}{2}}$, with the constant ρ free from i and j, $-1/(N-1) < \rho < 1$. By (2·3),

$$\phi_{ii} = g_1 v_{ii}^{-1}\pi_i \quad (1 \leq i \leq N), \quad \phi_{ij} = g_2(v_{ii}v_{jj})^{-\frac{1}{2}}\pi_{ij} \quad (1 \leq i \neq j \leq N).$$

Here

$$g_1 = \frac{1+(n-2)\rho}{(1-\rho)\{1+(n-1)\rho\}}, \quad g_2 = \frac{-\rho}{(1-\rho)\{1+(n-1)\rho\}},$$

and π_{ij} is the joint inclusion probability of units i and j. Define $l = (v_{11}^{\frac{1}{2}}, \ldots, v_{NN}^{\frac{1}{2}})'$. Observe that, by well-known relations on π_i's and π_{ij}'s, $l'\Phi l = g_1 n + g_2 n(n-1)$ and that, by the Cauchy–Schwarz inequality, $1'\Phi^{-1}1 \geq (1'l)^2/(l'\Phi l)$. Hence

$$1'\Phi^{-1}1 - 1'V1 \geq (1-\rho)\left\{n^{-1}\left(\sum_{i=1}^{N} v_{ii}^{\frac{1}{2}}\right)^2 - \sum_{i=1}^{N} v_{ii}\right\}$$

with equality if and only if Φl is proportional to 1 or equivalently

$$\pi_i = n v_{ii}^{\frac{1}{2}} \bigg/ \left(\sum_{i=1}^{N} v_{ii}^{\frac{1}{2}}\right) = \pi_{i0}$$

say for every i $(i = 1, \ldots, N)$. Further, for any p with $\pi_i = \pi_{i0}$ $(i = 1, \ldots, N)$, it is easy to verify that $b_{si}^* = \pi_{i0}^{-1}$. We thus have the following result.

COROLLARY 1. *Under the superpopulation model* (1·1) *with $v_{ij} = \rho(v_{ii}v_{jj})^{\frac{1}{2}}$ $(1 \leq i \neq j \leq N)$, a strategy (p, e) is optimal in H_n if and only if $\pi_i = \pi_{i0}$ for every i $(i = 1, \ldots, N)$ and e is given by the generalized difference estimator*

$$e = \sum_{i \in s} (Y_i - \mu_i)/\pi_{i0} + \sum_{i=1}^{N} \mu_i$$

for every s with $p(s) > 0$.

The earlier optimality results obtained by Godambe (1955), Hájek (1959), Cassel, Särndal & Wretman (1976) and Tam (1984) follow immediately from Corollary 1. Note, however, that in general the optimal estimator, as specified by (2·2) and (2·9), will not be a generalized difference estimator since the optimal coefficients b_{si}^* may depend on s. The following example serves as an illustration.

Example 1. Let $N = 4$, $n = 2$, $v_{ii} = \sigma^2$ $(i = 1, \ldots, 4)$, $v_{ij} = 0{\cdot}5\sigma^2$ $(1 \leq i \neq j \leq 3)$ and $v_{ij} = 0$ otherwise. As shown in Example 2 in § 3 then the optimal design is given by p^*, where $p^*(1, 2) = p^*(1, 3) = p^*(2, 3) = 0{\cdot}1181$, $p^*(1, 4) = p^*(2, 4) = p^*(3, 4) = 0{\cdot}2152$. Hence by (2·2), (2·9), the optimal strategy in H_n is (p^*, e^*), where

$$e^*(s) = \begin{cases} 1{\cdot}7889 \sum_{i \in s} (Y_i - \mu_i) + \sum_{i=1}^{4} \mu_i & \text{if } s = (i, j),\ 1 \leq i < j \leq 3; \\ 2{\cdot}6834(Y_i - \mu_i) + 1{\cdot}5489(Y_4 - \mu_4) + \sum_{i=1}^{4} \mu_i & \text{if } s = (i, 4),\ 1 \leq i \leq 3. \end{cases}$$

Note that e^* is different from e_1, the generalized difference estimator under the design p^*. It can be checked that $E_\alpha E_{p^*}\{(e^* - Y)^2\} = 2{\cdot}598\sigma^2$, while $E_\alpha E_{p^*}\{(e_1 - Y)^2\} = 2{\cdot}932\sigma^2$, so that the use of e^* rather than e_1 ensures a gain of over 10% in efficiency. Similarly, if one considers simple random sampling without replacement, say $\bar{p}$, then by (2·2), (2·9) it can be seen that $\bar{e}$, the corresponding optimal estimator, is different from e_2, the corresponding generalized difference estimator. Furthermore, $E_\alpha E_{\bar{p}}\{(\bar{e}-Y)^2\} = 2{\cdot}714\sigma^2$, $E_\alpha E_{\bar{p}}\{(e_2 - Y)^2\} = 3\sigma^2$, so that the gain in efficiency through the use of $\bar{e}$ rather than of e_2 is again about 10%.

3. Optimal sampling design

As noted in § 2, the derivation of the optimal design requires the minimization of $1'\Phi^{-1}1$ with respect to $p \in P_n$. Although in general an analytic solution to this nonlinear programming problem is not available, the algorithms popular in the theory of optimal experiments (Fedorov, 1972; Silvey, 1980) are useful.

Since we are considering unordered estimators, a design p in P_n may be conveniently represented by nonnegative quantities $\{p(s), s \in \mathscr{S}\}$, where

$$\mathscr{S} = \{(i_1, \ldots, i_n): 1 \leq i_1 < \ldots < i_n \leq N\}.$$

Clearly, $\Sigma' p(s) = 1$, where Σ' represents summation over $\mathscr{S}$. Then by (2·3),

$$\Phi = \sum' p(s) T(s), \tag{3·1}$$

where, for example with $s = (1, \ldots, n)$, the $N \times N$ matrix $T(1, \ldots, n)$ is defined as

$$T(1, \ldots, n) = \begin{pmatrix} V_{12\ldots n}^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

$V_{12\ldots n}$ being the $n \times n$ submatrix of V given by its first n rows and columns. Similarly, for each $s \in \mathscr{S}$ the matrix $T(s)$ of order $N \times N$ is defined. Note that $T(s)$ is nonnegative-definite for each s. Then analogously to Silvey (1980, pp. 19–20) one obtains the following theorem which involves the use of directional derivatives.

THEOREM 2. *A design* $\{p^*(s), s \in \mathscr{S}\}$ *is optimal in the sense of minimizing* $1'\Phi^{-1}1$, *that is maximizing* $-1'\Phi^{-1}1$, *in* P_n *if and only if*

$$F(\Phi^*, s) = \lim_{c \to 0+} c^{-1}[1'(\Phi^*)^{-1}1 - 1'\{(1-c)\Phi^* + cT(s)\}^{-1}1] \leq 0 \tag{3·2}$$

for every $s \in \mathscr{S}$, *where* $\Phi^* = \Sigma' p^*(s) T(s)$.

Since $T(s)$ is nonnegative-definite for each s, an explicit evaluation of the left-hand side of (3·2) shows that a design $\{p^*(s), s \in \mathscr{S}\}$ is optimal in P_n if and only if

$$F(\Phi^*, s) = 1'(\Phi^*)^{-1} T(s) (\Phi^*)^{-1} 1 - 1'(\Phi^*)^{-1} 1 \leq 0 \tag{3·3}$$

for every $s \in \mathscr{S}$. If the optimal design can somehow be guessed then (3·3) may be employed for a formal verification. In general, such a guess seems to be extremely difficult. Anyway, one may employ (3·3) to develop algorithms leading to a numerical determination of the optimal design. For example, a version of the W-algorithm (Silvey, 1980, pp. 29–30), as briefly outlined below, will be appropriate in the present context.

Let δ be a pre-assigned positive quantity and $\{c_k\}$ be a real sequence such that $0 < c_k < 1$ for each k, $\lim c_k = 0$ and Σc_k is divergent. At the first stage of iteration one may start with the design

$$p_1(s) = \binom{N}{n}^{-1}$$

for each $s \in \mathscr{S}$. For $k = 1, 2, \ldots$, let $\{p_k(s), s \in \mathscr{S}\}$ be the design at the kth stage of iteration and $\Phi_k = \Sigma' p_k(s) T(s)$. Let $F(\Phi_k, s)$ be defined as in (3·3). The iteration stops at the kth stage if $\max_{s \in \mathscr{S}} F(\Phi_k, s) < \delta$. Otherwise, one moves on to the $(k+1)$th stage of iteration and considers the design

$$p_{k+1}(s) = \begin{cases} (1 - c_{k+1}) p_k(s) & (s \neq s_{(k+1)}), \\ (1 - c_{k+1}) p_k(s_{(k+1)}) + c_{k+1} & (s = s_{(k+1)}), \end{cases}$$

where $s_{(k+1)}$ maximizes $F(\Phi_k, s)$ over $s \in \mathscr{S}$. Clearly

$$\Phi_{k+1} = (1 - c_{k+1})\Phi_k + c_{k+1} T(s_{(k+1)}),$$

and iteration is continued as before. Exactly as Silvey (1980, pp. 35–6), we can show that the above algorithm necessarily terminates and that if it terminates at the k'th stage then $1'(\Phi_{k'})^{-1}1 < 1'(\Phi^*)^{-1}1 + \delta$, where as before Φ^* corresponds to the optimal design. Thus the algorithm guarantees arbitrary close approach to the minimum possible value of $1'\Phi^{-1}1$.

Example 2. Let $N = 4$, $n = 2$ and suppose the v_{ij}'s are as in Example 1. From intuitive considerations one hopes that for the optimal design $p(1, 2) = p(1, 3) = p(2, 3) = q_1$, say, and $p(1, 4) = p(2, 4) = p(3, 4) = q_2$, say, where $3(q_1 + q_2) = 1$. It is easy to see that the choice of q_1, q_2 that minimizes $1'\Phi^{-1}1$ is $q_1 = 0{\cdot}1181$, $q_2 = 0{\cdot}2152$. Finally, it can be checked that the resulting design satisfies (3·3) and is, therefore, optimal.

Example 3. Let $N = 4$, $n = 2$, $v_{11} = 1{\cdot}0$, $v_{22} = 4{\cdot}0$, $v_{33} = 9{\cdot}0$, $v_{44} = 16{\cdot}0$, $v_{12} = v_{21} = 0{\cdot}4$, $v_{23} = v_{32} = 1{\cdot}2$, $v_{34} = v_{43} = 2{\cdot}4$, and $v_{ij} = 0$ otherwise. It is easy to obtain $T(s)$, for $s \in \mathscr{S}$. For example $T(1, 3)$ will be a 4×4 matrix, with 1 and $\frac{1}{9}$ in its (1, 1)th and (3, 3)th positions respectively, and zeros elsewhere. Here it is difficult to guess the optimal design but an application of the W-algorithm yields the optimal design p^* as $p^*(1, 3) = 0{\cdot}2213$, $p^*(2, 4) = 0{\cdot}4220$, $p^*(3, 4) = 0{\cdot}3567$, $p^*(1, 2) = p^*(1, 4) = p^*(2, 3) = 0$.

The optimal strategy discussed here generally involves the model parameters which may be unknown. In order to tackle this problem, asymptotic studies, along the lines of Särndal (1980) and Isaki & Fuller (1982) among others, may be appropriate. Consider a sequence of populations $\{U_t\}$ $(t = 1, 2, \ldots)$ such that U_t contains N_t units, where $N_t \to \infty$ as $t \to \infty$. Let $\mu_{(t)}$ and $V_{(t)}$ denote respectively the model mean vector and the model covariance matrix corresponding to U_t. Furthermore, as happens in many practical situations, let there exist a parameterization of $\mu_{(t)}$, $V_{(t)}$ as $\mu_{(t)} = X_t\gamma$, $V_{(t)} = V_{(t)}(\theta)$, where X_t is a $N_t \times h_1$ known matrix of values of regressor variables, the functional form $V_{(t)}(.)$ is known, γ and θ are $h_1 \times 1$ and $h_2 \times 1$ vectors of unknown parameters, and h_1, h_2 are known positive integers free from t. Let $Y_{(t)}$ be the population total, corresponding to U_t, of the variable of interest y. For $t = 1, 2, \ldots$, a sample s_t of n_t distinct units is considered from U_t, where $n_t \to \infty$ as $t \to \infty$. For $t = 1$, a sample s_1 is drawn from U_1 by simple random sampling without replacement and on the basis of the y-values ascertained from s_1, estimates $\hat{\gamma}_1$ and $\hat{\theta}_1$ of γ and θ may be obtained employing, for example, the method of two-stage least-squares; compare Malinvaud (1980, pp. 282–3). For $t = 2, 3, \ldots$, with reference to the population U_t, one may consider the strategy $(\hat{p}_t^*, \hat{e}_t^*)$ which is the optimal strategy corresponding to $\mu_{(t)} = X_t\hat{\gamma}_{t-1}$, $V_{(t)} = V_{(t)}(\hat{\theta}_{t-1})$, where $\hat{\gamma}_{t-1}$ and $\hat{\theta}_{t-1}$ are estimates of γ and θ obtained from s_{t-1} using two-stage least-squares. The results presented earlier may be employed to find $(\hat{p}_t^*, \hat{e}_t^*)$. Let (p_t^*, e_t^*) be the optimal strategy, with reference to U_t, when the model parameters γ and θ are known. Then under appropriate assumptions it is believed that for large t, the strategy $(\hat{p}_t^*, \hat{e}_t^*)$ should serve as a good approximation to (p_t^*, e_t^*) in the sense that the difference

$$n_t\{E_\alpha E_{\hat{p}_t^*}(\hat{e}_t^* - Y_{(t)})^2 - E_\alpha E_{p_t^*}(e_t^* - Y_{(t)})^2\}/N_t^2$$

should tend to zero as $t \to \infty$.

Acknowledgement

The authors are grateful to the referees for their very constructive suggestions.

References

CASSEL, C. M., SÄRNDAL, C. E. & WRETMAN, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63**, 615–20.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. New York: Academic Press.

GODAMBE, V. P. (1955). A unified theory of sampling from finite populations. *J. R. Statist. Soc.* B **17**, 269–78.

HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Časopis Pro Pěstování Matematiky*, Univ. Prague **84**, 387–423.

ISAKI, C. T. & FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Am. Statist. Assoc.* **77**, 89–96.

MALINVAUD, E. (1980). *Statistical Methods of Econometrics*. Amsterdam: North-Holland.

SÄRNDAL, C. E. (1980). On π-inverse weighting versus best linear weighting in probability sampling. *Biometrika* **67**, 639–50.

SILVEY, S. D. (1980). *Optimal Design*. London: Chapman and Hall.

TAM, S. M. (1984). Optimal estimation in survey sampling under a regression superpopulation model. *Biometrika* **71**, 645–7.

[*Received September* 1988. *Revised February* 1989]