A Separable Programming Approach To Optimum Allocation In Multivariate Response Errors

S. Maqbool , Nageena Nazir And S. A. Mir

Assistant Professor , Department of Statistics

Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar Srinagar

ABSTRACT.

One of the areas of statistics that is most commonly used in all fields of scientific investigation is that of probabilistic sampling. Obtaining good results in agricultural, biological and social research depends on successful, well-implemented adequate sampling plans. In this paper, we examine the problem of optimum allocation when dealing with response errors. The problem is formulated as a non linear programming problem with linear constraints and when non linearity occurs in the objective function. A solution procedure is developed using separable programming technique.

KEYWORDS:

Multivariate stratified sampling, response errors, cost function, separable programming.

1. INTRODUCTION:

The problem of sample allocation in multivariate stratified sampling has drawn the attention of researchers for a long time starting apparently with Neyman (1934). It is felt that unless the strata variances for various characters are distributed in the same way, the classical Neyman allocation based on the variances of a single character is of no use because an allocation which is optimum for one characteristics may not be acceptable for another. Due to this fact there is no unique or even widely accepted solution to the problem of optimum allocation in multivariate stratified sampling. Kokan and Khan (1967), Chatterjee (1968), Bethel (1985,1989), Chromy (1987), S. maqbool (2001) and Jose and Liliana (2008) all discuss the use of convex programming in relation to multivariate optimal allocation problem. The "convex programming" approach of Kokan and Khan (1967) gives the optimal solution to the defined problem but the procedure becomes involved if the number of characters to be estimated is large. Various procedures are available for dealing with separable programming problems. One of them is to use the simplex method with restricted basis entry rule (see Hadley (1964))**.** The convex separable objective functions are approximated by piece wise linear functions and then the restricted basis entry simplex method is used. The approximations enlarge the size of the problem as the number of variables increases rapidly if the number of strata increases. But good computer codes are available for solving even very large LPPs.

2. ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING:

We consider the situation where p characteristics are measured on each unit of a population, which is partitioned into L strata. Suppose that the sampling scheme within each stratum is that of simple random sampling without replacement (SRSWOR). Let n_{h} be the number of units to be drawn without replacement from the h^{th} stratum of size N_{h}, $h = (1,2,...,L)$. Let $\mathrm{y}_{\mathrm{hij}}$ be the value obtained for i^{th} unit on j^{th} character. For j^{th} character an unbiased estimate of the population mean $\bar{Y}_j$ is

$\bar{y}_{jst} = \sum_{h=1}^{L} W_h \bar{y}_{hj}$, where $\bar{y}_{hj} = \sum_{i=1}^{n_h} \frac{y_{hij}}{n_h}$ = Sample mean in the h^{th} stratum. Let the population mean in the h^{th} stratum be given by $\bar{Y}_{hj}$.

Define

$$S_{hj}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(y_{hij} - \bar{Y}_{hj} \right)^2 , W_h = \frac{N_h}{N} \quad and \ a_{hj} = W_h^2 S_{hj}^2$$

.

Then the variance of $\bar{y}_{jst}$ is given by

$$V(\bar{y}_{jst}) = \sum_{h=1}^{L} a_{hj} \left(\frac{1}{n_h} - \frac{1}{N_h} \right), \qquad j = 1,2,\ldots,P .$$

For large strata sizes it may be approximated by

$$V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{a_{hj}}{n_h} \tag{2.1}$$

Generally in multivariate surveys the upper tolerance limits on the variances of various characters are required to be achieved by the sample.

Let c_{hj} be the cost of enumerating j^{th} character on a unit in the h^{th} stratum and let $\sum_{j=1}^{p} c_{hj} = c_h$. Then the total cost of the survey is usually of the linear form

$$C = \sum_{h=1}^{L} c_h n_h \tag{2.2}$$

Now our aim is to determine the sample allocation $n_h, h = 1,\ldots,L$ so that the total cost of the survey is minimized subject to the conditions that the upper tolerance limits given on the variances for the various characters are satisfied. The sample allocation problem can be stated as to minimize the cost

$C = \sum_{h=1}^{L} c_h n_h$ Under the constants $\sum_{h=1}^{L} \frac{a_{hj}}{n_h} \le v_j$, $j = 1,\ldots,p$ and the limits $1 \le n_h \le N_h$, $h = 1,\ldots.,L$. Substituting x_h for $\frac{1}{n_h}$, the problem reduces to

$$\left.\begin{aligned} &Minimize \quad C = \sum_{h=1}^{L} \frac{c_h}{x_h} \\ &Subject \ to \quad \sum_{h=1}^{L} a_{hj} x_h \le v_j \ , j = 1,\ldots,p, \\ &\qquad \frac{1}{N_h} \le x_h \le 1 \ , h = 1,\ldots..,L. \end{aligned}\right\} \tag{2.3}$$

As both the objective function and the constraints are separable, it is possible to write the problem (2.3) in the form of a separable programming problem as follows,

$$\left.\begin{aligned} &Minimize \quad \sum_{h=1}^{L} f_h(x_h) \\ &Subject \ to \quad \sum_{h=1}^{L} g_{hj}(x_h) \le v_j , j = 1,2,\ldots,P \\ &\qquad \frac{1}{N_h} \le x_h \le 1 , h = 1,\ldots,L \end{aligned}\right\} \tag{2.4}$$

where $f_h(x_h) = \frac{c_h}{x_h} \quad and \quad g_{hj}(x_h) = a_{hj} x_h$.

3. RESPONSE ERRORS:

The category of non-sampling errors which arise from defective methods of data collection and from errors in tabulation Suppose a population of M interviewers is available to enumerate a population of N units on each of which p characters are defined. A SRS of m interviewers out of M is selected to interview a SRS of n units from N, so that the number of units allotted to each interview is $\frac{n}{m} = \bar{n}$ (assumed to be integer).

Let $\quad \bar{y}_j = \sum_{h=1}^{m} \sum_{I=1}^{\bar{n}} \frac{y_{hjI}}{n} \qquad , j = 1,2,\ldots,p$

be the mean value of the j^{th} character, where y_{hjI} is the value obtained for the I^{th} sample unit by h^{th} interviewer. Assuming N and M to be large relative to n and m, we have

$$V(\bar{y}_j) = \frac{(\sigma^2_{yj} - \sigma_{yjI})}{n} + \frac{\sigma_{yjI}}{m} \quad , j = 1,2,\ldots,p; I = 1,2,\ldots,\bar{n}$$
,

where σ^2_{yj} represents the "total variance" of individual responses around the mean of all individual responses in the population for j^{th} character and σ_{yjI} is the covariance between responses obtained from different individuals by the same interviewer for the j^{th} character (see Hansen and Hurwitz (1946)). Suppose that some upper limits v_j $j = 1,2,\ldots p$ are given on the variances of the mean values for various characters.

The cost function is of the form

$C = c_1 n + c_2 m$ Where $c_1 =$ Cost per individual and $c_2 =$ Cost per interviewer employed in the survey. The problem is to find the values of n and m which minimize the total sampling cost C subject to the bounds v_j on the variances.

Denoting by $a_{1j} = (\sigma^2_{yj} - \sigma_{yjI})$, $a_{2j} = \sigma_{yjI}$ and using the transformations $\frac{1}{n} = X_1$ *and* $\frac{1}{m} = X_2$, the problem becomes

$$\left.\begin{aligned} &Minimize \quad \sum_{h=1}^{2} \frac{C_h}{X_h} \\ &Subject \ \ to \quad \sum_{h=1}^{2} a_{hj} X_h \le v_j, \qquad j = 1,2,\ldots,p \\ &\qquad \frac{1}{N} \le X_1 \le X_2 \le 1. \end{aligned}\right\} \tag{3.1}$$

4.SOLUTION USING SEPARABLE PROGRAMM ING TECHNIQUE:

The non-linear functions $f_h(x_h)$ can be approximated by piece wise linear functions. Denote $H = \{1,2,..,L\}$; For each $h \in H$, let the feasible range of the variables x_h be given by the interval $[a_h, b_h]$ and choose a set of n_h grid points a_{hr} $(r = 1,2,\ldots,n_h)$ such that $a_h = a_{h1} < a_{h2} < \ldots < a_h n_h = b_h$.

Every point x_h in the grid $\lfloor a_{hr}, a_{h,r+1} \rfloor$ interval can be expressed as

$$X_h = \lambda_{hr} a_{hr} + \lambda_{h,r+1,} a_{h,r+1},$$

$$where \ \ \lambda_{hr} + \lambda_{h,r+1} = 1, \ \lambda_{hr} \ge 0 \ \ and \ \ \lambda_{h,r+1} \ge 0. \tag{4.1}$$

A linear approximation for a function $f_h(x_h), h \in H$ in the grid interval $\lfloor a_{hr}, a_{h,r+1} \rfloor$ is $\hat{f}_h(x_h) = \lambda_{hr} f_h(a_{hr}) + \lambda_{h,r+1} f_h(a_{h,r+1})$.

In general for the complete interval $x_h \in [a_h, b_h]$, the piecewise linear approximation $\hat{f}_h(x_h)$ can be written as

$$\hat{f}_h(x_h) = \sum_{r=1}^{n_h} \lambda_{hr} f_h(a_{hr}) \tag{4.2}$$

with $x_h = \sum_{r=1}^{n_h} \lambda_{hr} a_{hr}$, $\sum_{r=1}^{n_h} \lambda_{hr} = 1$; $\lambda_{hr} \ge 0$, for $r = 1,2,\ldots,n_h$,

provided for each h, at the most two adjacent λ_{hr} are positive.

An approximating linear program to the non-linear separable program (2.4) is thus obtained as

$$Minimize \ Z = \sum_{h \in H} \sum_{r=1}^{n_h} \lambda_{hr} f_h(a_{hr})$$

$$Subject \ \ to \quad \sum_{h \in H} \sum_{r=1}^{n_h} \lambda_{hrj} g_{hrj}(a_{hrj}) \le v_j, \qquad j = 1,2,\ldots,p \tag{4.3}$$

$$\sum_{r=1}^{n_h} \lambda_{hrj} = 1,\ \lambda_{hr} \geq 0,\ r = 1,\ldots,n_h,\ h \in H\ .$$

Problem (4.3) is a linear programming problem in $\sum_{h=1}^{L} n_h$ variables λ_{hr} and can be solved by simplex method using restricted basis entry rule for separable functions (see Hadley (1964)).

The optimal values $\lambda_{hr}^{*}\ \left(r = 1,2,\ldots,n_h, h \in H\right)$ obtained by solving the problem (4.3) yields an approximate optimal solution $\hat{x}_h$ to the original problem (2.4) by the substation.

$$\hat{x}_h = \sum_{r=1}^{n_h} \lambda_{hr}^{*} a_{hr} \quad ,h \in H$$

REFERANCES:

Bethel, J. (1985). An optimum allocation algorithm for multivariate survey ,*Proceedings of the Survey Research Section, Amer. Stat. Assoc.*, 204-212.

Bethel, J. (1989). Sample allocation in multivariate surveys. *Survey methodology*, **15**(1), 47-57.

Chatterjee, S.(1968). Multivariate stratified surveys. *Jour. Amer.Stat. Assoc.*,**63**,530-534.

Chromy, J.R. (1987). Design optimization with multiple objectives. *Proceedings of the survey research section, Amer.Stat.Assoc*, 194-199.

Hadley, G. (1964). *Non-linear and dynamic programming.* Addison-Wisely, Reading, Mass.

Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non- response in sampling surveys. *Jour. Amer. Stat. Assoc.*, **41**, 517-529.

Kokan, A.R. and Khan, S.U. (1967). Optimum allocation in multivariate surveys: an analytical solution. *Jour. Roy Stat. Soc.*, B, **29**, 115-125.

Neyman, J. (1934). On the two different aspects of representative method; The method of stratified sampling and the method of purposive selection. *Jour. Roy. Stat. Soc.*, **97**, 558-606.

S. Maqbool (2001). *Optimization techniques in sample surveys*. Ph. D thesis submitted to Aligarh Muslim University, Aligarh. India.

Jose A Diaz and Liliana Ulloa (2008). Multiobjective optimization for optimum allocation in multivariate stratified sampling. Survey Methodology, vol. 34,No.2,PP,215-222.