International Journal of Applied Mathematics

Volume 35 No. 1 2022, 173-180
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v35i1.13

COMPROMISE ALLOCATION FOR TWO-STAGE SAMPLING WITH QUADRATIC TRAVEL COST USING DYNAMIC PROGRAMMING TECHNIQUE

Abrar Ahmad[1], A.H. Ansari[2], Quazzafi Rabbani[1 §]

[1]Department of Mathematics and Statistics
Integral University, Lucknow - 226026, INDIA
[2]Kejriwal Institute of Manangement
and Development Studies
Namkum, Ranchi - 834010, INDIA

Abstract: In the stratified sampling literature the main problem is to determine the sample sizes that should be selected from each stratum under (i) Equal Allocation, (ii) Proportional Allocation and (iii) Optimum Allocation. The best method is optimum allocation but in real situation the implementation of optimum allocation is not possible. In this case it is of interest to find near optimal allocation or compromise allocation. In case of multivariate sampling problem (where p different characteristics are under study) the optimal allocation method does not give the optimal solution for each variable and then researcher have to adapt in solution up to some extent by which the solution gives the optimal allocation in some sense. The compromise allocation is advisable in this situation. The present paper discusses a real situation problem where the two stage sampling are under study for more than one characteristics with quadratic travel cost of survey, the problem can be formulated as Multivariate Non Linear Programming Problem (MNLPP). The MNLPP is then solved by Dynamic Programming Technique with a numerical example.

Received: October 26, 2021

§Correspondence author

AMS Subject Classification: 62D05
Key Words: cluster sampling; two stage sampling; dynamic programming techniques; compromise allocation and quadratic travel cost

1. Introduction

In a multivariate study (more than one characteristic) one may not able to find optimal solution or allocation because an optimal allocation for one variable not necessarily satisfies the optimal solution for other variables. Therefore, one cannot use the optimum allocation technique of Neyman [5] for multiple characteristics and to find such allocation which is an adjustment of conflict (compromise) for all characteristics. This technique is also known as 'compromised allocation technique' which is in some sense optimum for all the characteristics.

According to Ignizio [6] a multivariate optimal allocation having more than one objective problem can be solve by using different weights to the different objective functions according to their importance and combined as a single objective function then solve the problem as Linear Programming Problem (LPP) or as Non Linear Programming Problem (NLPP). Another way to solve the problem with more than one objective function to optimize the most important objective and put tolerance limits on the others and treat these as constraints of the problem.

In a special situation where the correlation among the characteristics may be high and the individual allocations may differ relatively less the average individual optimal allocation was suggested by Cochran [12]. Many other authors also proposed several criteria to work out compromise allocations.

Non Linear Programming Problem (NLPP) can be formulated for compromise allocation in multivariate cluster sampling for multiple characteristics. The NLPPs are then solved using Dynamic programming technique and explicit formulas are obtained for the optimum allocation of the first-stage and second-stage sampling units.

There are a number of solution procedures available for single variable non linear programming problem (NLPP) to get optimal solution in both the cases (i) for one stage as well as (ii) for two stage sampling. In case of two stage sampling design, the problem splits into each stage univariate problem and solved accordingly by any available method like (Cochran [12], Arnold [2], Sadooghi-Alvandi [11], Valliant and Gentle [10], Clark and Steel [9], Dever et al. [4]).

But for more than one variable (characteristics) are under study the procedures for determining optimum allocations are not well defined. The compromise allocation method is used in such situation, because in many surveys, the use of two-stage sampling designs often specifies two stages of selection- the first stage unit is known as primary sampling units (PSUs), and second stage unit is known as secondary sampling units (SSUs) which will be the sub-samples of PSUs. Waters and Chester [7] proposed a graphical approach to identify the possible optimum solution for multivariate case. Ansari [1] used linear cost to discuss the procedure but it is not possible that the cost function will always be linear, the cost function can be non-linear. So, in this paper authors consider a quadratic cost function.

In the present paper, a multivariate two-stage sampling design with quadratic travel cost has been developed as compromise allocation method. The problems of determining the compromise allocations have been formulated as Nonlinear Programming Problems (NLPP), which can be solved by Dynamic programming Technique.

2. The formulation

A two-stage sampling with multivariate as characteristics say p characteristics are under study, n units of PSU are drawn randomly from N units in first stage and m subunits of SSU are randomly drawn from and M units in the second stage within each selected n units of PSU. Let y_{ijk} denotes the value obtained from j^{th} subunit in the i^{th} primary unit, $\bar{y}_{ik} = \sum_{j=1}^{m} \frac{y_{ijk}}{m}$ the sample mean per subunit in the i^{th} primary unit, and $\bar{\bar{y}}_k = \sum_{i=1}^{n} \frac{\bar{y}_{ik}}{n}$ the overall sample mean per subunit for k^{th} characteristic. It could be shown that $\bar{\bar{y}}_k$ is an unbiased estimate of the overall population mean $\bar{\bar{Y}}_k$ of k^{th} characteristic with variance $V(\bar{\bar{y}}_k) = \left(\frac{N-n}{N}\right)\frac{S_{1k}^2}{n} + \left(\frac{M-m}{M}\right)\frac{S_{2k}^2}{nm}$ where the stratum standard deviations among the primary unit means be S_{1k}^2 and the stratum standard deviations among the subunit within primary unit for k^{th} characteristic be S_{2k}^2, respectively. If the above stratum standard deviations among PSU and SSU are unknown but their estimates are available then the variance can be written as

$$V(\bar{\bar{y}}_k) = \left(\frac{N-n}{N}\right)\frac{s_{1k}^2}{n} + \left(\frac{M-m}{M}\right)\frac{s_{2k}^2}{nm}. \tag{1}$$

Assuming a quadratic cost function of two-stage sampling procedure, the total cost of the survey may be expressed as

$$C = c_1 n^2 + c_2\,(n\,m)^2 + d_1 n + d_2\,(n\,m)\,, \tag{2}$$

where c_1, c_2 are the cost of measurement of all the characteristics survey for PSU and SSU and d_1 and d_2 is the travel cost of survey for PSU be the travel cost of all the characteristics survey for PSU and SSU.

The optimum value of n and m for an individual characteristic can thus be calculated by minimizing the variance in (1) for the given cost in (2), or by minimizing the cost for fixed variance.

Generally a compromise criterion is needed in multivariate sample survey to find out an acceptable number of PSU's and SSU's which is optimum, in some sense, for all characteristics. Khan *et al.* [8] suggested the compromise criterion for the predetermined total cost of the survey, in two-stage sampling, if we are interested to find the means of p characteristics, it may be a reasonable criterion to determine the optimal choice of n and m by minimizing the weighted sum of the variances of the two-stage sample means of all the p characteristics, that is

$$\sum_{k=1}^{p} W_k V(\bar{\bar{y}}_k),$$

where W_k is the weight assigned to the k^{th} characteristic in proportion to its importance as compared to other characteristics and $V(\bar{\bar{y}}_k)$ as given above. Ignoring the *fpc,* V will be equivalent to minimize

$$\sum_{k=1}^{p} W_k \left\{ \frac{1}{n}\left(s_{1k}^2 + \left(\frac{M-m}{mM}\right) s_{2K}^2 \right) \right\}.$$

If the cost is fixed then the problem of optimum allocation could be as under by following NLPP

$$\left.\begin{array}{ll} \text{Minimize} & V(m,n) = \sum_{k=1}^{p} W_k \left(\frac{s_{1k}^2}{n} + \frac{s_{2k}^2}{m\,n} - \frac{s_{2k}^2}{nM} \right) \\ \text{Subject to} & c_1 n^2 + c_2\,(n\,m)^2 + d_1 n + d_2\,(n\,m) \le C_0 \\ \text{and} & n,\, m \ge 0 \end{array}\right\}, \tag{3}$$

the restrictions *n, m>0* and *M>0* are obvious because negative values of the number of PSU's and SSU's are of no practical use, so NLPP (3) becomes

$$\left.\begin{array}{ll} \text{Minimize} & V(m,n) = \left(\frac{\sum_{k=1}^{p} W_k \left(s_{1k}^2 - \frac{s_{2k}^2}{M} \right)}{n} + \frac{\sum_{k=1}^{p} W_k s_{2k}^2}{m\,n} \right) \\ \text{Subject to} & c_1 n^2 + c_2\,(n\,m)^2 + d_1 n + d_2\,(n\,m) \le C_0 \\ \text{and} & n,\, m \ge 0 \end{array}\right\}, \tag{4}$$

$$\left.\begin{array}{ll}\text{Minimize} & V(\alpha)=\left(\frac{A_1}{\alpha_1}+\frac{A_2}{\alpha_2}\right)\\ \text{Subject to} & c_1{\alpha_1}^2+c_2{\alpha_2}^2+d_1\alpha_1+d_2\alpha_2\le C_0\\ \text{and} & \alpha_1,\alpha_2\ge 0\end{array}\right\},\tag{5}$$

where $A_1=\sum_{k=1}^{p}W_k\left(s_{1k}^2-\frac{s_{2k}^2}{M}\right)$, $A_2=\sum_{k=1}^{p}W_ks_{2k}^2$, $\alpha_1=n$ and $\alpha_2=mn$ and in general form we can write it as

$$\left.\begin{array}{l}V(\alpha_j)=\sum_{j=1}^{2}\frac{A_j}{\alpha_j}\\ \sum_{j=1}^{2}c_j\alpha_j^2+d_j\,\alpha_j\le Co\\ \alpha_j\ge 0\,;\ j=1,2\end{array}\right\}.\tag{6}$$

3. Solution procedure using dynamic programming techniques

In this section, solution procedures have been developed for solving MPP (6), using dynamic programming technique. Consider the r^{th} stage sub problem of MPP (6) for the first $r(<p)$ groups

$$\left.\begin{array}{ll}\text{Minimize} & \sum_{j=1}^{r}f_j(\alpha_j)\\ \text{subject to} & \sum_{j=1}^{r}g_j(\alpha_j)\le C_p\\ \text{and} & \alpha_j\ge 0\ ;\ j=1,2,...,r\end{array}\right\},\tag{7}$$

where $f_j(\alpha_j)=\frac{A_j}{\alpha_j}$, $g_j(\alpha_j)=c_j\alpha_j^2+d_j\,\alpha_j\le C_0$; $j=1,2,...,r$; $c_r\le C_0$ is the available budget for measurements of the selected units from the first r groups. With the above definition of C_r we have:
$C_r=C_0$ for $r=p$
also
$C_r=g_1\,(\alpha_1)+g_2\,(\alpha_2)+...+g_r\,(\alpha_r)$
$C_{r-1}=g_1\,(\alpha_1)+g_2\,(\alpha_2)+...+g_{r-1}\,(\alpha_{r-1})=C_r-g_r(\alpha_r)$
: :
$C_2=g_1\,(\alpha_1)+g_2\,(\alpha_2)=C_3-g_3(\alpha_3)$
and
$C_1=g_1\,(\alpha_1)=C_2-g_2\,(\alpha_2)$.

If $f\,(r,C_r)$ denotes the minimum value of the objective function of sub problem (7), then

$$f(r,C_r)=\underset{\text{feasible}\,\alpha_j}{\text{Min}}\{\sum_{j=1}^{r}f_j(\alpha_j):\sum_{j=1}^{r}g_j(\alpha_j)=C_r\ \text{ and }\ \alpha_j\ge 0\}.\tag{8}$$

For the first stage ($r = 1$):
$f(1, C_1) = \frac{A_1 c_1}{C_1}$ at $\alpha_1 = \frac{-d_1+\sqrt{d_1^2+4c_1 C_1}}{2c_1}$ (ignorance of negative α_1)

and for ($r \geq 2$),

$$f(r, C_r) = \left\{ \frac{A_r}{\alpha_r} + f(r-1, C_{r-1}) \right\}, \quad (9)$$

gives the required recurrence relation.

From $f(k, C_k)$ the optimum value of α_k is obtained from $f(k-1, C_{k-1})$ the optimum value of α_{k-1}, is obtained and so on until $\alpha_1$is determined. After obtaining α_j; $j = 1, 2, ..., k$ the value of n , m are obtained accordingly.

4. A numerical example

In the table given below the stratum sizes and stratum weights are given for five different characteristics under study in a population of size 200. The sampling cost for first stage is Rs.60 and Rs.50 and observation cost in second stage is Rs.20 and Rs.25 respectively, where the total allotted budget is Rs.2000 and M=5.

Characteristics	1	2	3	4	5
W_k	0.25	0.32	0.21	0.08	0.14
S_{1k}^2	58	44	42	64	77
S_{2k}^2	116	123	148	137	109

Table 1

The above problem can be solved by using Dynamic Programming Techniques (DPT) manually.

In Table 2 the value of $A_1$and A_2 are calculated.

Characteristics	W_k	S_{1k}^2	S_{2k}^2	$A_1 = W_k(s_{1k}^2 - \frac{s_{2k}^2}{M})$	$A_2 = W_k s_{2k}^2$
1	0.25	58	116	8.7	29
2	0.32	44	123	6.208	39.36
3	0.21	42	148	2.604	31.08
4	0.08	64	137	2.928	10.96
5	0.14	77	109	7.728	15.26
Total				**28.168**	**125.66**

Table 2

With the above values the problem can be written in the form of (7) as

$$\left.\begin{array}{ll} \text{Minimize} & \sum_{j=1}^{2} f_j(\alpha_j) \\ \text{subject to} & \sum_{j=1}^{2} g_j(\alpha_j) \le C_p \\ \text{and} & \alpha_j \ge 0 \ ; \ j = 1, 2 \end{array}\right\},$$

where $f_j(\alpha_j) = \frac{A_j}{\alpha_j}$, $g_j(\alpha_j) = c_j \alpha_j^2 + d_j \alpha_j \le C_0$; $j = 1, 2, ..., r$.

The above problem is solved by software Lingo 18.0 for convenience but it can be solved by other software as well manually, the optimal values of where α_j are $\alpha_1 = 2.912753$ and $\alpha_2 = 14.880593882$ then the values of $n = 2.912753$ and $m = 5.108773$ and the value of objective function (ignoring fpc) $\sum_{j=1}^{2} f_j(\alpha_j) = 34.26748$.

References

[1] A.H. Ansari, Compromise allocation for two stage cluster sampling using dynamic programming technique, *International J. of Mathematical Archive,* **6**, No 4 (2015), 10–15.

[2] B.F. Arnold, Procedures to determine optimum two-stage sampling plans by attributes, *Metrika,* **33** (1986), 93–109.

[3] H.W. Kuhn and A.W. Tucker, Nonlinear programming, *Proc. of the Second Berkley Symp. on Mathematical Statistics and Probability*, University of California Press, Berkley (1951), 481–492.

[4] J.A. Dever, J. Liu, G.V. Iannacchione and D.E. Kendrick, An optimal allocation method for two-stage sampling designs with stratification at the second stage, *ASA Proc. of the Joint Statistical Meeting, Amer. Stat. Association*, Alexandria, VA (2001).

[5] J. Neyman, On the two different aspects of the representative method: The method of purposive selection, *J. of Royal Stat. Soc.*, **97** (1934), 558–606.

[6] J.P. Ignizio and T.M. Cavalier, *Linear Programming*, Printice Hall, Englewood Cliff, New Jersey (1994).

[7] J.R. Waters and A.J. Chester, Optimal Allocation in multivariate, two-stage sampling designs, *The American Statistician*, **41** (1987), 46–50.

[8] M.G.M. Khan, E.A. Khan and M.J. Ahasn, An optimal multivariate stratified sampling design using dynamic programming, *Australian & New Zealand J. Statist.*, **45**, No 1 (2003), 107–113.

[9] R.G. Clark and D.G. Steel, Optimum allocation of sample to strata and stages with simple additional constraints, *J. of the Royal Stat. Soc., Ser. D: The Statistician*, **49** (2000), 197–207.

[10] R. Valliant and J.E. Gentle, An application of mathematical programming to sample allocation, *Computational Statistics & Data Analysis*, **25** (1997), 337–360.

[11] S.-A. Mohammad, The choice of subsample size in two-stage sampling, *J. of the Amer. Stat. Association*, **81** (1986), 555–558.

[12] W.G. Cochran, *Sampling Techniques*, Third E d., John Wiley & Sons, Inc., New York (1977).