View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by International Institute for Science, Technology and Education (IISTE): E-Journals

Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.6, No.4, 2016
www.iiste.org

# A Branch and Bound Approach to Optimal Allocation in Stratified Sampling

N. A. Sofi [1], Aquil Ahmad [2], S. Maqbool [3] and Bhat Bilal Ahmad [4]
[1] Department of IT & Statistics, SKUAST- Kashmir.
[2] Department of Statistics and Operations Research, AMU, Aligarh
[3] Division of AE & Statistics, FOA, SKUAST- Kashmir.
[4] Division of Social Science, Faculty of Fisheries, SKUAST- Kashmir.

**Abstract:-**
For practical applications of any allocations, integer values of the sample sizes are required. This could be done by simply rounding off the non-integer sample sizes to the nearest integral values. When the sample sizes are large enough or the measurement cost in various strata are not too high, the rounded off sample allocation may work well. However for small samples in some situations the rounding off allocations may become infeasible and non-optimal. This means that rounded off values may violate some of the constraints of the problem or there may exist other sets of integer sample allocations with a lesser value of the objective function. In such situations we have to use some integer programming technique to obtain an optimum integer solution.

**Keywords**: Stratified sampling, Non-linear Integer Programming, Allocation Problem, Langrangian Multiplier, Branch & Bound Technique

## 1. Introduction

Optimization is the science of selecting the best of many possible decisions in a complex real life situation. The ultimate target of all such decisions is to either maximize the desired benefit or to minimize the effort required, incurred in a certain course of action. The development of the concepts of linear and non-linear optimization model presumes that all of the data for the optimization model are known with certainty. However, uncertainty and inexactness of data and outcomes pervade many aspects of most optimization problem. As it turns out, when the uncertainty in the problem is of a particular (fairly general) form, it is relatively easy to incorporate the uncertainty into the optimization model

The precision of an estimator of the population parameters depends on the size of the sample and the variability or heterogeneity among the units of the population. If the population is very heterogeneous and considerations of the cost limit the size of the sample, it may be found impossible to get a sufficiently precise estimate by taking a simple random sample from the entire population. In stratified sampling the population of size N is divided into none overlapping and exhaustive groups called strata each of which is relatively more homogeneous as compared to the population as a whole. Independent simple random samples of predetermined sizes from each stratum are drawn and the required estimators of the population parameters are constructed.

## 2. Review of literature

Optimum allocation of sample sizes to various strata in univariate stratified random sampling is well defined in the literature. But usually in real life situations more than one population characteristics are to be estimated which may be of conflicting nature. There are situations where the cost of measurement varies from stratum to stratum. Also the cost of enumerating varies characters is generally much different. What is therefore optimum for one characteristic may not be optimum for the others.

The author to give a convex programming (CP) formulation to the allocation problem in multivariate stratified sampling was Kokan (1963). An analytical solution through this CPP model was provided by Kokan and Khan (1967). They also showed how the sample allocation problem in other designs, such as two-stage sampling or double sampling can be viewed as a CPP. In the presence of prior information and an overhead cost in each stratum, the cost function to be minimized becomes concave. A solution to this allocation problem in this situation was provided by Ahsan and Khan (1977). The problem of determining strata boundaries in multivariate surveys was considered by Ahsan et al. (1983).

Chadha et al. (1971) used dynamic programming technique to find the optimum allocation in univariate case. Later Omule (1985) used the same technique for the multivariate sampling. Khan (1995) and Khan (1997)

 

have also minimized the weighted sum of the variances of the estimates of different characteristics using dynamic programming.

Bethal (1989) expresses the optimal multi-character stratified sample allocation as a closed expression in terms of normalized Lagrangian multipliers whereas Rahim (1994) proposed an alternative procedure based on distance function of the sampling errors of all the estimates. Various authors like Armstrong and Mu (1992), Kreienbrock (1993), Nandi and Aich (1995), Chernyak and Starytskyy (1998), Chernyak (1999), Chernyak and Chornous (2000) either suggested new criteria or explored further the already existing criteria.

Garcia Diaz et al. (2005) give a detailed study of the problem under a number of stochastic optimization techniques. Javid et al. (2009), considered the case of random costs and used modified E-model for solving this problem. Bakhshi et.al. (2010) find the optimal sample number in Multivariate Stratified Sampling with probabilistic cost constraints. Khan et al. (2011, 2012) gave thorough study on Chebyshev approximate solution to allocation Problem in multivariate objective surveys with random costs and allocation in multivariate stratified survey with non-linear random cost function. Gupta et.al (2013) studied on optimal chance constraint multivariate stratified sampling design and also Fuzzy Goal Programming approach in stochastic multivariate stratified sampling surveys.

## 3. Integer Programming Problem

The enumeration techniques are designed in such a way that all the integer feasible points are enumerated either explicitly or implicitly in systematic manner so as to get finally the optimal integer point. The enumeration of integer points is possible as the feasible region of a bounded integer program always contains a finite set of feasible points. The enumeration techniques such as branch-and-bound enumeration and implicit enumeration which require to enumerate only a small subset of feasible integer points to arrive at the optimal solution. This technique Quasi-enumerative approach to problem solving that has been applied to a wide variety of combinatorial. It is fairly efficient for modest size problems and the general methodology forms an important part of the set of (exact) solution methods for the general class of integer linear programming problems (LPP's).

Any decision problem (with an objective to be maximized or minimized) in which the decision variables must assume non-fractional or discrete values may be classified as an inter optimization problem. In general, an integer problem may be constrained or un-constrained and the functions representing the objective and constraints may be linear or non-linear. An integer problem is classified as linear if by relaxing the integer restrictions on the variables, the resulting functions are strictly linear.

The general mathematical model of an integer programming problem can be stated as:

Maximize (or Minimize) $Z = f(x)$

Subject to

$$g_i(X) \leq \text{ or } = \text{ or } \geq b_i, \; i = 1,2,....,m$$

$$x_j \geq 0, \quad j = 1, 2, .....,n$$

$x_j$ is an integer

where $X = (x_1, x_2, ....., x_n)$ is the n–component vector of dimension variables an N = {1, 2, ……, n}.

If I = N that is, all variables are restricted to be integer, we have an all (or pure) integer programming problem, otherwise

If I $\neq$ N i.e. not all variables are restricted to be integer, we have a mixed integer programming problem (MIPP)

In this paper, we consider the allocation problem in stratified sampling with linear sampling costs. We use the branch and bound technique for obtaining the integer solution to the formulated non-linear integer programming problem. The basic idea of branch and bound is to partition a given problem into a number of sub problems. This process of partitioning is usually called branching and its purpose is to establish sub problems that are easier to solve than the original problem because of their smaller size or amenable structure. Numerical illustrations is presented here to support the theoretical results.

 

## 4. Problem Formulation

The following notations will be used to define the sample allocation problem. The decision variable of interest is the sample size of each stratum. The suffix $j$ stands for $j^{th}$ stratum, $j = 1, 2, \ldots, L$, where $L$ denotes the total number of strata into which the population has been divided.

$N_j$ = Total number of units in the stratum $j$, $j = 1,2, \ldots, L$

$n_j$ = Number of units selected in the sample from the stratum $j$

$W_j = \frac{N_j}{N}$ = Proportion of population units falling in the stratum $j$

$\bar{y}_j$ = $j^{th}$ Stratum mean

$S_j^2$ = $j^{th}$ Stratum variance

$C_j$ = Cost of surveying one unit in stratum $j$, ($C_j > 0, j = 1, 2, \ldots, L$)

$\bar{y}_{st} = \sum_{j-1}^{L} \frac{N_j}{N} \bar{y}_j$ Stratified sample mean

The variance of the $\bar{y}_{st}$ is given by

$$V(\bar{y}_{st}) = \sum_{j=1}^{L} \frac{W_j^2 S_j^2}{n_j} - \sum_{j=1}^{L} \frac{W_j^2 S_j^2}{N_j}$$

For large strata sizes the second term on the right may be ignored and we get

$$V(\bar{y}_{st}) \approx \sum_{j=1}^{L} \frac{W_j^2 S_j^2}{n_j}$$

The problem of optimal sample allocation involves determining the sample size $n_1, n_2, \ldots\ldots, n_j$ that minimize the variance $V(\bar{y}_{st})$ subject to a given sampling budget C, or determining $n_1, n_2, \ldots\ldots, n_j$ that minimizes sampling cost subject to an upper bound on the variance. The simplest cost function is of the form $\sum_{j=1}^{L} C_j n_j$. Within any stratum the cost is proportional to the size of sample, but the cost per unit $C_j$ may vary from stratum to stratum. This cost function is appropriate when the major item of cost is that of taking the measurements on each unit.

We consider the integer allocation problems for linear cost function and fixed budget.

For linear cost function and fixed budget the problem is formulated as

$$\text{Minimize} \quad \sum_{j=1}^{L} \frac{W_j^2 S_j^2}{n_j} \tag{4. 1}$$

$$\text{Subject to} \quad \sum_{j=1}^{L} C_j n_j \le C \tag{4. 2}$$

$$1 \le n_j \le N_j \tag{4. 3}$$

$$n_j \text{ integers,} \tag{4. 4}$$

## 5. Solution Procedure

The solution of problem (4.1) to (4.2) ignoring the upper and lower bound restricting and the integer requirements.

 

$$\phi = \sum_{j=1}^{L} \frac{W_j^2 S_j^2}{n_j} + \lambda \left[ \sum_{j=1}^{L} C_j n_j - C \right]$$

Differentiation with respect to $n_j$ and $\lambda$, we get

$$\frac{\partial \phi}{\partial n_j} = - \sum_{j=1}^{L} \frac{W_j^2 S_j^2}{n_j^2} + \lambda C_j = 0$$

$$\frac{\partial \phi}{\partial \lambda} = \sum_{j=1}^{L} C_j n_j - C = 0$$

which on simplification gives the initial solution as

$$n_j = \frac{C W_j S_J \Big/ \sqrt{C_j}}{\sum_{j=1}^{L} W_j S_j \sqrt{C_j}}, \qquad j = 1,2,.....L \tag{5.1}$$

Now the Land and Doig approach of the branch and bound technique will require the solution of sub-problems in which some of the $n_j$ are fixed. Suppose that at $K^{th}$ node, the fixed values of $n_j$ are for $j \in I_k$. Then the corresponding lagrangian is

$$\phi = \sum_{j \neq I_k}^{L} \frac{W_j^2 S_j^2}{n_j} + \lambda \left[ \sum_{j \neq I_k}^{L} C_j n_j - C \right]$$

Equating to zero the differentials of $\phi$ with respect to $n_j$ and $\lambda$, we obtain the solution at node *K* as

$$n_j = \frac{\left( C - \sum_{i \in I_K} C_i n_i \right) W_j S_j \Big/ \sqrt{C_j}}{\sum_{j \neq I_K}^{L} W_j S_j \sqrt{C_j}}, \quad j = 1,2,.....L \tag{5.2}$$

For branching from each node of the tree, we will choose an $n_j$ at the current node which either violates the integer requirements or which violates the upper and lower bounds. Whenever the branching is done on the bounds then one branch will fix the corresponding $n_j$ on the violated bound and the other on the next feasible integer value.

## 6. Numerical Illustration

The data in table below is related to the farmers of 64 villages in the south Kashmir of J&K (in thousands) for the year 2012. The villages are grouped into three strata. There are 16, 20 and 28 villages respectively in the first, second and third stratum and the data is condensed in the tabular form as below:

 

**Table - 6.1**

| Stratum | $N_j$ | $S_j^2$ | $W_j$ | $C_j$ |
|---|---|---|---|---|
| 1. | 16 | 540.0625 | 0.25 | 04 |
| 2. | 20 | 14.6737 | 0.3125 | 1.5 |
| 3. | 28 | 7.2540 | 0.4375 | 01 |

For the above data, the allocation problem may be stated as follows:

Minimize $$\frac{33.7539}{n_1} + \frac{1.4330}{n_2} + \frac{1.3885}{n_3}$$

Subject to

$$4n_1 + 1.5\,n_2 + n_3 \le 70$$
$$1 \le n_1 \le 16$$
$$1 \le n_2 \le 20$$
$$1 \le n_3 \le 28$$
$$n_1, n_2, n_3 \text{ integers}$$

Using (5.1), we get the optimal values $n_1, n_2$ *and* $n_3$ as

$$n_1 = 14.253 \qquad n_2 = 4.809 \qquad n_3 = 5.783$$

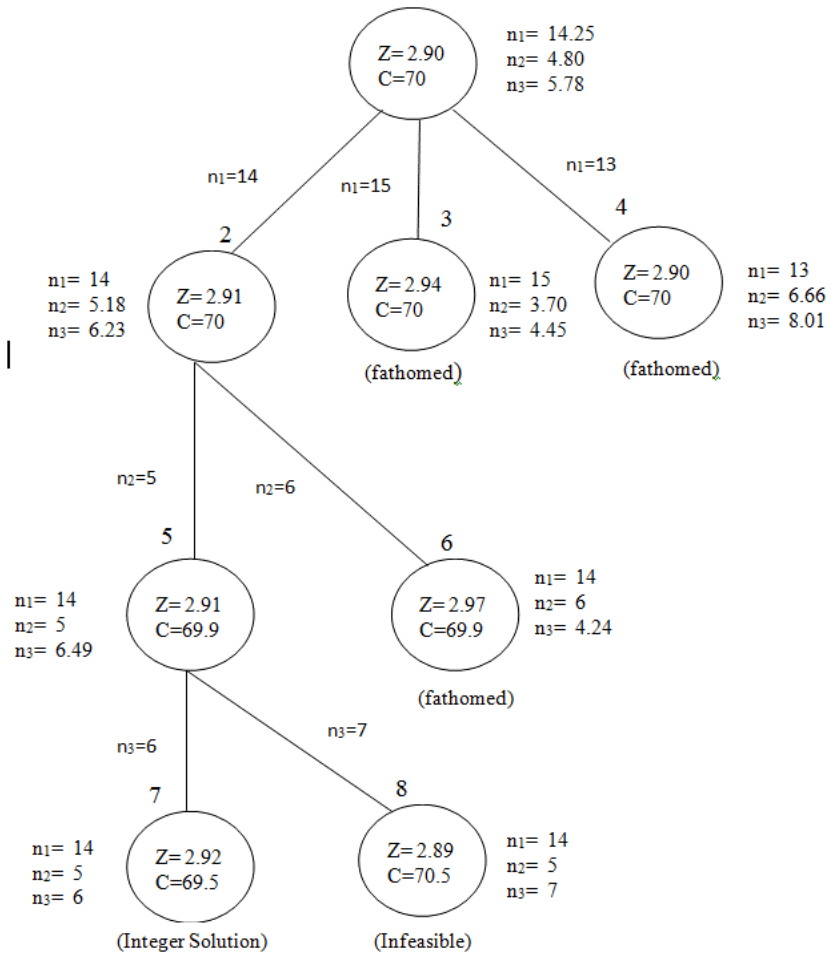


**Figure 6.1: The various nodes of branch and bound method while solving the problem**

 

## 7. Conclusions

In this paper, we use branch and bound techniques for obtaining the integer solution to the allocation problem in stratified sampling with linear cost function. The integer solution obtained by using branch and bound comes out to be at node. 7 as $n_1$=14, $n_2$=5 and $n_3$=6 with Z*=2.92. This solution happens to be our optimal solution. The main advantage of using the branch and bound to allocation problem is that, we can easily tackle the situation of over sampling – i.e. the optimal allocation requires sampling more than 100% in certain strata.

## 8. References

Ahsan, M. J. & Khan, S. U. (1977). Optimum allocation in multivariate stratified random sampling using prior information. *Journal of Indian Statistical Association, 15*, 57-67.

Ahsan, M. J., Khan, S. U. & Arshad, M. (1983). Minimising a non-linear function arising in stratification through approximation by quadratic function, *Journal of Indian Society of Statistics & Operations Research, 4*, 9-16.

Armstrong, J. B., & Mu, C. F. J. (1992). A sample allocation method for two-phase survey designs. *Survey Methodology, 18*(2), 253-262.

Bakhshi, Z.H., Khan, M. F., & Ahmad, Q. S. (2010). Optimal sample numbers in multivariate stratified sampling with probabilistic cost constraints. *International Journal of Mathematics and Applied Statistics, 1*(2),111-120.

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

Chaddha, R. L., Hardgrave, W. W., Hudson, D. J., Segal, M. and Suurballe, J. W. (1971). Allocation of total sample size when only the stratum means are of interest. *Technometrics, 13*, 817-831.

Chernayak, O. I. & Starytskyy, A. (1998). Optimal allocation in stratified sampling with convex cost function. *Visnyk of Kyiv University Economics (In Ukrainian), 39*, 42-46

Chernayak, O. I. (1999). Allocation problem in Bayesian stratified sampling with a non-linear cost function. *Bulletin of the international statistical institute, 52nd session contributed papers, LVIII*, 169-170.

Chernayak, O. I. & Chornous, G. (2000). Optimal allocation in stratified sampling with a non-linear cost function. *Theory of Stochastic Process, 6*, 6-17.

Garcia, Diaz, J. A., Ramos-Quirogh, R, & Casrera-Vicenico, E. (2005). Stochastic programming methods in the response surface methodology. *Comm. Statist. Data Analysis, 49*, 837-848.

Gupta, N., Shafiullah, Ali, I. & Bari A. (2013). A fuzzy goal programming approach in stochastic multivariate stratified sampling surveys. *The South Pacific of Natural and Applied Sciences*, CISRO, Fijji.

Javid, S., Bakhshi, Z. H. & Khalid M. M. (2009). Optimum allocation in stratified sampling with random costs. *Int. Review of Pure and Applied Mathematics, 5*(2), 363-370.

Khan, M. G. M. (1995). *Mathematical programming in sampling* (Doctoral thesis). Aligarh Muslim University, Aligarh, India.

Khan, E. A. (1997). *On use of mathematical programming techniques in some optimization problems arising in stratified sample surveys* (Doctoral thesis). Aligarh Muslim University, Aligarh, India.

Khan, M. F., Ali, I., and Ahmad, Q. S. (2011). Chebyshev approximate solution to allocation problem in multivariate objective surveys with random costs. *American Journal of Computational Mathematics, 1*(4), 247-251.

Khan, M. F., Ali, I., Raghan, Y. S. & Bari, A. (2012). Allocation in multivariate stratified survey with non-linear

 

random cost function. *American Journal of Operation Research, 2*(1), 122-125.

Kokan, A. R. (1963). Optimum allocation in multivariate surveys. *Jour. Roy. Stat. Soc., A,*126,557-565.

Kokan, A. R. & Khan, S. U. (1967). Optimum Allocation in Multivariate Surveys: An Analytical Solution. Journal of the Royal Statistical Society, Ser. B, 29, 115-125

Kreienbrock, L. (1993). Generalized measure of dispersion to solve the allocation problem in multivariate stratified random sampling. *Comm.Stat.Theo.Meth. 22*(1), 219-239.

Land, A. H. & Doig, A.G. (1960). An automatic method for solving discrete programming problems. *Econometrica, 28*, 497-520.

Nandi, S. B. & Aich, A. B. (1995). Optimal stratified sampling: An information theoretic approach. *Second International Triennial Calcutta Symposium on probability and Statistics.*

Omule, S. A. Y. (1985). Optimum design in multivariate stratified sampling. *Biometrical Journal, 27*(8), 907-912.

Rahim, M. A. (1994). Sample allocation in multivariate stratified design: *An alternative to convex programming.* Abstracts for survey Research Methods joint statistical meetings.