Solution of a Multivariate Stratified Sampling Problem through Chebyshev's Goal Programming

Mohd. Vaseem Ismail
Faculty of Pharmacy
Jamia Hamdard, New Delhi
India

Kaynat Nasser
Department of Mathematics
Integral University, Lucknow
India

Qazi Shoeb Ahmad
Department of Mathematics
Integral University, Lucknow
India
qazishoeb@rediffmail.com

Abstract

We consider the problem of minimizing variances for the various characters for a fixed (given) budget. Each convex objective function is first linearised at its minimal point. The resulting multi-objective linear programming problem is then solved by Chebyshev's goal programming. A numerical example is given to illustrate the procedure.

Keywords Multivariate stratified sampling, Multi-objective linear programming, Chebyshev's goal programming.

1. Introduction

In sample surveys, estimation of more than one population characteristics may be required. When stratified sampling is to be used, then an allocation criterion among various strata that is uniformly optimum may not exist. A suitable overall optimality criterion is required for dealing with such situations.

Multi-objective optimization (or programming), also known as '''multi-criteria''' or '''multi-attribute''' optimization, is the process of simultaneously optimizing two or more conflicting objectives subject to certain constraints.

Various authors have to date suggested new criteria or improved existing ones. For a review of these works, see Neyman (1934), Peter and Bucher (1940), Geary (1949), Dalebius (1957), Ghosh (1958), Yates (1960), Aoyama (1963), Chatterjee (1968). The use of convex programming in relation to the multivariate optimal allocation problem has been discussed by Kokan and Khan (1967), Huddleston, et al (1970), Arvanitis and Afonja (1971), Chromy (1987), Bethel (1985, 1989) among others. Each approach has its advantages and disadvantages. The weighted average method is computationally simple, intuitively appealing and can be solved under a fixed cost assumption, but the choice of the weights is arbitrary and the optimality properties are not clear. The convex programming approach gives the optimal solution to the defined problem where the upper limits are given on the variances and the cost is to be minimized. But if the variances are to be minimized a further search is usually required for an optimal solution.

2. Multivariate Stratified Sampling

We consider a multivariate population partitioned into L strata. Suppose that p characteristics are measured on each unit of the population. We assume that the strata boundaries are fixed in advance. Let n_i be the number of units drawn without replacement from i^{th} stratum $(i = 1,2,...,L)$. Let N_i be the size of i^{th} stratum. For j^{th} character, an unbiased estimate of the population mean $\bar{Y}_j \; (j = 1,2,..., p)$ denoted by $\bar{y}_{jst}$, has its sampling variance

$$V(\bar{y}_{jst}) = \sum_{i=1}^{L}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)W_i^2 S_{ij}^2, \;\; (j = 1,2,..., p)$$

where

$$W_i = \frac{N_i}{N}, \; S_{ij}^2 = \frac{1}{N_i - 1}\sum_{k=1}^{N_i}\left(y_{ijk} - \bar{Y}_{ij}\right)^2$$

Substituting $a_{ij} = W_i^2 S_{ij}^2$, we get

$$V(\bar{y}_{jst}) = \sum_{i=1}^{L}\frac{a_{ij}}{n_i} - \sum_{i=1}^{L}\frac{a_{ij}}{N_i}, \; (j = 1,2,..., p) \quad (2.1)$$

Let C_{ij} be the cost of enumerating the j^{th} character in the i^{th} stratum and let C be the upper limit on the total cost of the survey. Then assuming linear cost function one should have

$$\sum_{i=1}^{L}\sum_{j=1}^{p} C_{ij} n_i \le C \;\text{ or }\; \sum_{i=1}^{L} C_i n_i \le C \quad (2.2)$$

where $C_i = \sum_{j=1}^{p} C_{ij}$, the cost of enumeration of all the p characters in the i^{th} stratum. Further one should have

$$1 \le n_i \le N_i, \; (i = 1,2,..., L). \quad (2.3)$$

We determine the optimum values of n_i by minimizing (in some sense) all the p variances (2.1) for a fixed budget (2.2) i.e. we have to

$$\text{Minimize } V_j = \sum_{i=1}^{L}\frac{a_{ij}}{n_i} - \sum_{i=1}^{L}\frac{a_{ij}}{N_i}, \; (j = 1,2,..., p)$$

$$\text{Subject to} \quad \sum_{i=1}^{L} C_i n_i \le C \tag{2.4}$$

and $1 \le n_i \le N_i$, $(i = 1,2,...,L)$

Since $N_i^{,s}$ are given, it is enough to minimize

$$V_j = \sum_{i=1}^{L} \frac{a_{ij}}{n_i}, \; (j = 1,2,...,p)$$

Using X_i for n_i, the problem (2.4) can be written as the following multi-objective non-linear programming problem:

$$\left.\begin{aligned} &\text{Minimize } V_j = \sum_{i=1}^{L} \frac{a_{ij}}{X_i}, && (j = 1,2,...,p) && (a) \\ &\text{Subject to} \quad \sum_{i=1}^{L} C_i X_i \le C && && (b) \\ &\text{and} \quad 1 \le n_i \le N_i, \; (i = 1,2,...,L) && && (c) \end{aligned}\right\} \tag{2.5}$$

The objective functions in (2.5) are convex [see Kokan and Khan (1967)], the single constraint is linear and the bounds are also linear. The problem (2.5) is, therefore a multi-objective convex programming problem.

Each objective function in (2.5) is convex and the single constraints as well as the upper and lower bounds are linear. The problem (2.5) for $j = k$ is, therefore, a convex programming problem which can be solved by using any method of convex programming. Each of the p problems for $k = 1,2,...,p$ may have a different solution. A unique solution, suitable for all the p problems is obtained here by using the criterion of Chebyshev's goal programming. In order to be able to apply the Chebyshev's goal programming approach we approximate the convex objective functions in (2.5) by linear ones and then solve the resulting linear programming problems. The criterion behind the Chebyshev's goal programming is to find a solution that minimizes the single worst unwanted deviation from any (soft) goal. In other words, it is a minimax goal programming approach.

3. Transformation into a Multi-objective Linear Programming Problem

In the multi-objective allocation problem (2.5) there are p non-linear objective functions which are later turning into soft goals with a single linear constraint (hard goal). To apply Chebyshev's goal programming approach, all the hard and soft goals must be in linear form so the worst deviation from the approximated linear goals is minimized. We thus approximate the non-linear soft goals by linear ones.

It may be noted that an analytic solution of the problem (2.5) for single character, say,

$j = k$ is given (see Kokan and Khan(1967)) as

$$x'_{ik} = \frac{C\sqrt{a_{ik}C_i}}{C_i\left\{\sum_{i=1}^{L}\sqrt{a_{ik}C_i}\right\}}, \qquad (i = 1,2,...,L) \tag{3.1}$$

provided that $1 \le x'_{ik} \le N_i$, $(i = 1,2,...,L)$.

In case the lower and/or upper bounds are violated for some i (which is a very extreme case and rarely occurs in practice), some extra efforts are needed as explained in the above reference. However, since at this stage we need only approximate points, we may fix such x'_{ik} at the corresponding bounds.

Our strategy will be to approximate the convex objective surface V_k by the tangent hyperplane at the point (3.1).

This is obtained as

$$V_k = V_{k(x'_{ik})} + \nabla V'_{k(x'_{ik})}\left(X_i - x'_{ik}\right), \qquad (i = 1,2,...,L)$$

where $\nabla V'_{k(x'_{ik})}$ is the vector of partial derivatives,

$$\nabla V'_{k(x'_{ik})} = \left[-\frac{a_{1k}}{\left(x'_{1k}\right)^2}, -\frac{a_{2k}}{\left(x'_{2k}\right)^2}, ..., -\frac{a_{Lk}}{\left(x'_{Lk}\right)^2}\right]$$

Then $$\nabla V'_{k(x'_{ik})}\left(X_i - x'_{ik}\right) = \sum_{i=1}^{L}\frac{a_{ik}}{x'_{ik}} - \sum_{i=1}^{L}\frac{a_{ik}X_i}{\left(x'_{ik}\right)^2}$$

This gives $$V_k \approx 2\sum_{i=1}^{L}\frac{a_{ik}}{x'_{ik}} - \sum_{i=1}^{L}\frac{a_{ik}X_i}{\left(x'_{ik}\right)^2} = v_k \quad \text{(say)}$$

Then the multi-objective convex programming problem (2.5) reduces to the following approximate multi-objective linear programming problem:

$$\begin{aligned}
&\text{Minimize } v_j = 2\sum_{i=1}^{L}\frac{a_{ik}}{x'_{ik}} - \sum_{i=1}^{L}\frac{a_{ik}X_i}{\left(x'_{ik}\right)^2} \qquad (j = 1,2,...,p)\\
&\text{Subject to } \sum_{i=1}^{L} C_i X_i \le C\\
&\text{and } \quad 1 \le X_i \le N_i, \quad (i = 1,2,...,L)
\end{aligned} \tag{3.2}$$

4. Solution Using Chebyshev's Goal Programming

It can be noted that for individual objective functions the solutions of the respective problems in (2.5) and those in (3.2) coincide for $j = 1, 2, ..., p$ and are given by (3.1).

To solve the multi-objective linear programming problem (3.2), we use the Chebyshev's goal programming approach in which the p objective functions are put in the form of constraints, termed as soft goals, with upper bounds called aspiration level. Aspiration level L_k is nothing but the minimum value of V_k obtained by solving convex programming problem (2.5) individually for the k^{th} objective function. The explicit solutions for these p problems can again be obtained by using (3.1).

The Chebyshev's goal programming model first solving 3.2 is given as

$$\begin{aligned} &\text{Minimize } \delta \\ &\text{Subject to} \quad \sum_{i=1}^{L} C_i X_i \le C \\ &\qquad 2\sum_{i=1}^{L} \frac{a_{ij}}{x_{ij}} - \sum_{i=1}^{L} \frac{a_{ij} X_i}{(x'_{ij})^2} - \delta \le L_j, \qquad (j = 1, 2, ..., p) \\ &\text{and} \quad 1 \le X_i \le N_i, \quad (i = 1, 2, ..., L) \end{aligned} \tag{4.1}$$

where δ (dummy variable) represents the worst deviation level.

Our practical experience shows that the solution X^*_{ch} by transforming the multi objective convex programming to the multi objective linear programming problem and using the Chebyshev's approach for its solution, provides us a satisfactory point in the sense that the values of the various objective functions at this point remain very close to the optimal values obtained by individually solving the convex programming problems (2.5) for various $j = 1, 2, ..., p$.

This observation is evident also from the numerical example given below

5. Numerical Example

Consider a population divided in two strata with three characteristics under study for which the values of N_i, W_i, S_{i1}, S_{i2} and S_{i3} are given in the following table:

Table 1.1

Stratum *i*	N_i	W_i	S_{i1}	S_{i2}	S_{i3}	C_{i1}	C_{i2}	C_{i3}
1	180	0.40	1.50	2.25	0.75	0.60	0.90	1.50
2	270	0.60	3.00	4.75	5.25	0.80	1.20	2.00

The variance coefficient matrix is obtained by $a_{ij} = W_i^2 S_{ij}^2$ as

$$\left(a_{ij}\right) = \begin{pmatrix} 0.36 & 0.81 & 0.09 \\ 3.24 & 8.12 & 9.92 \end{pmatrix}$$

Let us fix the budget at 100 units

The above problem is transformed to the multi-objective convex programming problem as

$$\begin{aligned} &\text{Minimize} && V_1 = \frac{0.36}{X_1} + \frac{3.24}{X_2},\ V_2 = \frac{0.81}{X_1} + \frac{8.12}{X_2} \text{ and } V_3 = \frac{0.09}{X_1} + \frac{9.92}{X_2} \\ &\text{Subject to} && 3X_1 + 4X_2 \le 100 \\ &\text{and} && 1 \le X_1 \le 180 \\ & && 1 \le X_2 \le 270 \end{aligned} \qquad 5.1$$

First we find out the solutions to the problem of minimizing V_1, V_2 and V_3 individually, subject to the linear constraints $3X_1 + 4X_2 \le 100$ by using 3.1.

For V_1 the solution is

$$x'_{11} = \frac{100\sqrt{0.36 \times 3}}{3\{\sqrt{0.36 \times 3} + \sqrt{3.24 \times 4}\}} = 7.47$$

$$x'_{21} = \frac{100\sqrt{3.24 \times 4}}{4\{\sqrt{0.36 \times 3} + \sqrt{3.24 \times 4}\}} = 19.40$$

Similarly the solutions of V_2 and V_3 are given by (7.16, 19.63) and (2.54, 23.10) respectively.

Now, linearized form of the objective function V_1 at the point (7.47, 19.40) is obtained as

$$v_1 \approx -0.0065X_1 - 0.0086X_2 + 0.4304$$

Similarly the linearized forms of the objective functions V_2 and V_3 at the respective points are obtained as

$$v_2 \approx -0.0158X_1 - 0.0211X_2 + 1.0540$$

$$v_3 \approx -0.0140X_1 - 0.0186X_2 + 0.9300$$

The values of L_1, L_2 and L_3 (aspiration levels) at the points (7.47, 19.40), (7.16, 19.63) and (2.54, 23.10) are obtained as 0.2512, 0.5270 and 0.4650 respectively.

Now, the approximated multi-objective linear programming problem to the multi-objective convex programming problem (5.1) is

$$\begin{aligned} \text{Minimize} \quad & v_1 \approx -0.0065X_1 - 0.0086X_2 + 0.4304, \\ & v_2 \approx -0.0158X_1 - 0.0211X_2 + 1.0540 \\ \text{and} \quad & v_3 \approx -0.0140X_1 - 0.0186X_2 + 0.9300 \\ \text{Subject to} \quad & 3X_1 + 4X_2 \le 100 \\ \text{and} \quad & 1 \le X_1 \le 180 \\ & 1 \le X_2 \le 270 \end{aligned} \tag{5.2}$$

The Chebyshev's model of the problem (5.2), becomes as to

$$\begin{aligned} \text{Minimize} \quad & \delta \\ \text{Subject to} \quad & -0.0065X_1 - 0.0086X_2 - \delta \le -0.2152 \\ & -0.0158X_1 - 0.0211X_2 - \delta \le -0.5270 \\ & -0.0140X_1 - 0.0186X_2 - \delta \le -0.4650 \\ & 3X_1 + 4X_2 \le 100 \\ \text{and} \quad & 1 \le X_1 \le 180 \\ & 1 \le X_2 \le 270 \\ & \delta \ge 0 \end{aligned} \tag{5.3}$$

The Chebyshev's point by solving the LPP (5.3) is $X_{ch}^{*} = (12.15, 15.89)$ with $\delta = 0$. The values of sample sizes n_1 and n_2 rounded to the nearest integers, are 12 and 16 respectively.

The solution print out of the problem through MATLAB is:

$X =$

12.1442
15.8825
0

$Lambda =$

0
0
0
0
0
0

$How =$

ok

$Z =$

0

This solution is being summarized in Table 1.2

The percent increases in the three variances for the Chebyshev's point as compared to the respective individual variance minimization points as 104.78%, 110.23% and 136.04%.

Table 1.2: Values of V_j at the individual optimal points and at the Chebyshev's point

	Optimization w.r.t. V_1	Optimization w.r.t. V_2	Optimization w.r.t. V_3	Chebyshev's point
Rounded n_1 and n_2	(7, 19)	(7, 20)	(3, 23)	(12, 16)
Value of V_1	0.2219	0.2234	0.2609	0.2325
Value of V_2	0.5432	0.5218	0.6232	0.5752
Value of V_3	0.5351	0.5090	0.4614	0.6277

References

1. Aoyama, H. (1963): Stratified random sampling with optimum allocation for multivariate populations. *Ann. Inst. Stat. Math*., 14, 251-258.
2. Chatterjee, S. (1968): Multivariate Stratified Surveys. *Jour. Amer. Stat. Assoc.* 63, 530-534.
3. Chromy, J.R. (1987): Design optimization with multiple objectives. *Proceedings of the Survey Research Section, ASA*, 194-199.
4. Cochran, W.G. (1977): Sampling Techniques. *John Wiley & Sons, New York.*
5. Ghosh, S. P. (1958): A note on stratified random sampling with multiple characters. *Cal. Stat. Assoc. Bull.,* 8, 81-89.
6. Jahan, N., Khan, M.I.M., and Ahsan, M.J. (1994): A generalized compromise allocation, *Journal of the Indian Statistical Association.* 32, 95-101.
7. Kelly, J.E. (1960): The cutting Plane Method for Solving Convex Programs. *Jour. Soc. Indust. Appl. Math.* 8, 703-712.
8. Kokan, A.R and Khan, S.U. (1967): Optimum allocation in multivariate surveys-An analytical solution. *Jour. Roy. Stat. Soc. Ser.* B.29, 115-125.
9. Neyman, J. (1934): On the two different aspects of representative method: The method of stratified sampling and the method of purposive selection. *Jour Roy. Stas. Soc*., 97, 558-606.
10. Yates, F. (1960): Sampling methods for censuses and surveys (2nd ed). *Charles Griffin and Co. Ltd. London.*