BULLETIN of the MALAYSIAN MATHEMATICAL SCIENCES SOCIETY

Optimal Allocation in Multivariate Sampling Through Chebyshev Approximation

SHARIEFUDDIN PIRZADA AND SHOWKAT MAQBOOL

P.G. Department of Mathematics and Statistics, University of Kashmir, Srinagar-190 006 India e-mail: pirzada@rediffmail.com

Abstract. In multivariate stratified sampling the problem of allocating the sample to various strata can be formulated as a programming problem with several linear objective functions and single convex constraint. The problem has been solved by finding the Chebyshev point for various conflicting objective functions. A comparison with the fuzzy programming solution has also been made.

1. Introduction

Usually in sample surveys more than one population characteristics are estimated. These characteristics may be of conflicting nature. When stratified sampling is used, an allocation that is optimum for one character is generally not so for others. A suitable overall optimality criterion is required for dealing with such problems.

Kokan and Khan [7] formulated the above problem as a non-linear multi objectiveprogramming problem. They also derived an analytical solution to the problem. Cochran [4] initially suggested the use of the average of individual optimum allocations for various characters. Chatterjee [1] gave a compromise allocation by minimizing the sum of the proportional increases in the variances due to the use of non-optimum allocation. Jahan *et al.*, [5] discussed the problem of obtaining the compromise allocation by minimizing the total relative increase in the variances as compared to the optimum allocation. Charnayak and Slarytsky [2], Charnayak and Chornous [3] suggested new criteria and explored further the already existing criteria.

In this paper, we consider the problem with fixed (given) budget. Also the tolerance limits are given on the variances for certain characters. The problem of allocating the sample to various strata may then be viewed as that of minimizing the variances of various characters subject to the conditions of given budget and tolerance limits on certain variances. The problem turns out to be non-linear programming problem with several linear objective functions and single convex constraint. The non-linearity of the convex constraint is handled through cutting plane technique. The resulting LPP is then solved by Chebyshev approximation approach. The criteria behind the Chebyshev approximation is to find a solution that minimizes the single worst. It is a minimax goal programming approach.

1. Allocation problem

Suppose that *p*-characteristics are measured on each unit of a population which is partitioned into *L* strata. Let n_i , be the number of units drawn without replacement from the *i*-th stratum ($i = 1, 2, \dots, L$). For the *j*-th character an unbiased estimate of the population mean, \overline{Y}_j , is \overline{y}_{jst} which has the sampling variance.

$$V_j = V(\bar{y}_{jst}) = \sum_{i=1}^{L} W_i^2 S_{ij}^2 X_i \qquad j = 1, 2, \cdots, p$$
(2.1)

where

$$W_i = \frac{N_i}{N}$$
, $S_{ij}^2 = \frac{1}{N_i - 1} \sum_{h=1}^{N_i} (y_{ijth} - \overline{Y}_{ij})^2$

and

$$X_i = \frac{1}{n_i} - \frac{1}{N_i}, \quad a_{ij} = W_i^2 S_{ij}^2.$$

in usual notation.

Let c_{ij} be the cost of enumerating the *j*-th characteristic in the *i*-th stratum and let *k* be the upper limit on total cost of the survey. Then one should have

$$\sum_{i=1}^{L} \sum_{j=1}^{p} c_{ij} n_i \le k.$$
(2.2)

The multivariate allocation problem can be stated as:

Minimize $z_j = \sum_{i=1}^{L} a_{ij} X_i - \sum_{i=1}^{L} \frac{a_{ij}}{N_i}, \quad j = 1, 2, \dots, p.$

Subject to

$$\sum_{i=1}^{L} \sum_{j=1}^{p} \frac{c_{ij}}{X_i} \le k$$
$$\frac{1}{N_i} \le X_i \le 1, \quad i = 1, 2, 3, \cdots, L$$
(2.3)

where $\frac{1}{X_i}$ is used for n_i . If we consider the Problem (2.3) separately for each character, then ignoring the constant term in the objective function, the problem for say, *k*-th character, becomes

222

Optimal Allocation in Multivariate Sampling Through Chebyshev Approximation

Minimize

$$Z_k = \sum_{i=1}^L \frac{a_{ik}}{X_i}$$

7 n

Subject to

$$\sum_{i=1}^{L} \sum_{j=1}^{P} c_{ij} X_i \le k$$

 $1 \le X_i \le N_i, \qquad i = 1, 2, \dots, L.$

By introducing a new variable x_{L+k} , the problem (2.4) transforms to

 $Z_k = x_{L+k}$

Minimize Subject to

$$g_k(X) = \sum_{i=1}^{L} \frac{a_{ik}}{X_i} - X_{L+k} \le 0$$
 (b) (2.5)

$$\sum_{i=1}^{L} \sum_{j=1}^{p} c_{ij} X_{i} \le k$$
(c)
$$1 \le X_{i} \le N_{i}, \quad i = 1, 2, \dots, L.$$
(d)

The constraints in (2.5b) are convex (see Kokan and Khan [7]) and the constraint (2.5c) and the bounds (2.5d) are linear. The problem (2.5a)-(2.5d) is therefore a convex programming problem with linear objective and can be solved by using any method of convex programming. However, we have *p* such problems for $k = 1, 2, \dots, p$ and each of these may have a different solution. A unique solution may be obtained by using the criteria of Chebyshev approximation. In order to be able to find a Chebyshev point for *p* problems, we first linearize the convex constraints (2.5b) by using the cutting plane technique of Kelly [6].

3. Linearizing the constraints

Let $X^{k(0)} = (X_1^{k(0)}, \dots, X_L^{k(0)}, X_{L+k}^{k(0)})$ be the solution of LPP, which minimizes (2.5a) subject to the single constraint (2.5c) and the bounds (2.5d).

Compute
$$g_k(X^{k(0)}) = \sum_{i=1}^L \frac{a_{ik}}{x_i^{k(0)}} - X_{L+k}^{k(0)}$$

Define \in to be a small tolerance limit for the convergence.

If $g_k(X^{k(0)}) \le \in$, this means that (2.5b) as also satisfied to the tolerance limits and thus $X^{k(0)}$ solves the problem.

223

(2.4)

(a)

If $g_k(X^{k(0)}) > \in$, we linearize the convex constraint $g_k(X) \le 0$ about the point $X^{k(0)}$ as:

$$g_{k}(X) \approx g\left(X^{k(0)}\right) + \nabla g_{k}\left(X^{k(0)}\right)'\left(X - X^{k(0)}\right) \leq 0$$

Where

$$\nabla g_k \left(X^{k(0)} \right)' = \left[-\sum_{j=1}^{P} \frac{c_{1j}}{X_1^{k(0)^2}}, \dots, -\sum_{j=1}^{P} \frac{c_{Lj}}{X_L^{k(0)^2}}, -1 \right]$$

This give

$$2\sum_{i=1}^{L} \frac{a_{ik}}{x_i^{k(0)}} - \sum_{j=1}^{p} \frac{a_{ik}x_i}{x_i^{k(0)^2}} - X_{L+k} \le 0$$
(3.1)

The constraint (2.5b) is then replaced by this linearized constraint (3.1). The following LPP approximates the NLPP (2.5).

Minimize
$$Z_k = X_{L+k}$$
 (a)

Subject to

$$2\sum_{i=1}^{L} \frac{a_{ik}}{X_{i}^{K(0)}} - \sum_{i=1}^{L} \frac{a_{ik}X_{i}}{X_{i}^{k(0)^{2}}} - X_{L+k} \le 0$$
 (b)

$$\sum_{i=1}^{L} \sum_{j=1}^{P} c_{ij} X_{i} \le k$$
(c)
$$1 \le X_{i} \le N_{i}, \qquad i = 1, 2, \dots, L$$
(d)
(3.2)

$$1 \le X_i \le N_i, \qquad i = 1, 2, \cdots, L \qquad (d)$$

Denote the solution of LPP (3.2) by

$$X^{k(1)} = \left(X_1^{k(1)}, \cdots, X_L^{k(1)}, X_{L+k}^{k(1)}\right).$$

At *t*-th iteration, we find

$$g_k \left(X^{k(t)} \right) = \sum_{i=1}^{L} \frac{a_{ik}}{X_i^{k(t)}} - X_{L+k}^{k(t)} .$$
(3.3)

If in 3.3, $g_k(X^{k(t)}) \leq \in$, the process terminates and we take

$$X^{k(t)} = X_{opt}$$
 and $Z_k(X^{k(t)}) = Z_{opt}$.

Otherwise we linearize the constraint $g_k(X)$ about the point $X^{k(t)}$.

The process is then repeated until (3.3) is satisfied say at $t_k^* - th$ iteration. After getting t_k^* , we solve the following LPP's for $k = 1, 2, \dots, p$. Here it is noted that each LPP consists of $1 + p + p\left(\sum_{i=1}^{p} t_k^*\right)$ linear constraints and 2L lower and upper bounds.

Let the minimum values of Z_k thus found be Z_k^o , $k = 1, 2, \dots, p$ at the corresponding minimal points X_k^o , $k = 1, 2, \dots, p$. The *p* solutions X_1^o, \dots, X_p^o have been obtained by minimizing the individual objective functions subject to the linearized constraints which will give us the aspiration levels being used in Chebyshev approximation.

4. Solution using Chebyshev approximation

For obtaining a unique solution suitable for all the p objective functions, we use the Chebyshev approximation technique. The Chebyshev approximation formulation of the multiple objective allocation problem (2.5) is the following LPP:

Subject to $2\sum_{i=1}^{L} \frac{a_{ik}}{X_{i}^{k(0)}} - \sum_{i=1}^{L} \frac{a_{ik}X_{i}}{X_{i}^{k(0)^{2}}} - X_{L+K} \leq 0$ $2\sum_{i=1}^{L} \frac{a_{ik}}{X_{i}^{k(1)}} - \sum_{i=1}^{L} \frac{a_{ik}X_{i}}{X_{i}^{k(1)^{2}}} - X_{L+K} \leq 0$ $\sum_{i=1}^{L} \sum_{j=1}^{p} c_{ij}X_{i} \leq k \qquad l = 1, 2, \cdots, t_{k}^{*}$ $k = 1, 2, \cdots, p$ $X_{L+k} - \delta \leq Z_{k}^{0}$ $1 \leq X_{i} \leq N_{i} \qquad t = 1, 2, \cdots, L$

Where δ (dummy variable) represents the worst deviation level. Note that the aspiration levels are being taken as the minimum values of the objective functions at Z_k^o , $k = 1, 2, \dots, p$. Since we are solving the non-linear problem by linearizing the objective function, the actual aspiration levels should be computed by substituting the point in the non-linear objective function instead of the linearized one.

5. Algorithm

Let us consider the problem (3.2). Set k = 1

- **Step I.** If k > p, go to Step III. Otherwise find the point $X^{k(0)}$ by minimizing (3.2a) subject to (3.2c) and (3.2d). At the first iteration we solve the LPP (3.2) to obtain the solution $X^{k(1)}$.
- **Step II.** If $g_k(X^{k(t)}) \le \in$, for some t, say t_k^* , where \in is a suitable tolerance limit, then $X^{k(t_k^*)} = X_k^o$ and $Z_k^o = Z_k(X^{k(t_k^*)})$ and go to Step I with k = k + 1. Otherwise relinearize the constraint $g_k(X)$ about the point $X^{k(t)}$ and add this constraint to the LPP (3.2) to obtain the LPP (3.4). Let the solution of (3.4) be $X^{k(t+1)}$. Repeat Step II with t = t + 1.
- **Step III.** Solve LPP (3.4) for $k = 1, 2, \dots, p$ to obtain X_k^o , the approximate minimal points for the respective objective functions, with minimum corresponding values of Z_k as Z_k^o .

226

Minimize

δ

Optimal Allocation in Multivariate Sampling Through Chebyshev Approximation

Step IV. Solve the Chebyshev approximation model (4.1) of the multivariate allocation problem (2.5) to obtain X_{ch}^* (Chebyshev point).

6. Fuzzy solution

Like Chebyshev approximation the basis of fuzzy programming is also to minimize the worst deviation from any goal. For obtaining a fuzzy solution, we first compute the maximum value U_k , and the minimum value L_k , for each k.

Let $Z_k(X_j^o) = z_k^{jo}$, $j = 1, 2, \dots, p$. Clearly $Z_k^{ko} = Z_k^o = \min_j Z_k(x_j^o)$.

Denote $Z_k^o = L_k$ and $\max_j Z_k(X_j^0) = U_k$.

The differences of the maximum and minimum values of Z_k are denoted by $d_k = U_k - L_k$, $k = 1, 2, \dots, p$.

The fuzzy programming formulation of the multivariate allocation problem (2.5) is the following LPP:

Minimize

δ

Subject to

$$\begin{array}{ll} \mathbf{0} & 2\sum_{i=1}^{L} \frac{a_{ik}}{x_{i}^{k(0)}} - \sum_{i=1}^{L} \frac{a_{ik}X_{i}}{x_{i}^{k(0)^{2}}} - X_{L+k} \leq 0, \quad k = 1, 2, \cdots, p \\ \\ & 2\sum_{i=1}^{L} \frac{a_{ik}}{x_{i}^{k(l)}} - \sum_{i=1}^{L} \frac{a_{ik}X_{i}}{x_{i}^{k(l)^{2}}} - X_{L+k} \leq 0, \quad l = 1, 2, \cdots, t_{k}^{*}, \quad k = 1, 2, \cdots, p \quad (6.1) \\ \\ & \sum_{i=1}^{L} \sum_{j=1}^{P} c_{ij}X_{i} \leq k \\ & X_{L+k} - d_{k}\delta \leq Z_{k}^{o} \qquad \qquad k = 1, 2, \cdots, p (*) \\ & 1 \leq X_{i} \leq N_{i} \qquad \qquad i = 1, 2, \cdots, L \end{array}$$

Comparing (4.1) and (6.1) it can be noted that the fuzzy programming solution is better than the Chebyshev solution if d_k , the difference between maximum and minimum values of the objective functions, are greater than 1 for all characteristics. The reason behind this is that the constraints (6.1.*) in fuzzy programming are less restrictive than the corresponding constraints in Chebyshev solution.

7. Example

In a stratified population with two strata and three characteristics, the values of N_i , W_i , S_{i1} , S_{i2} and S_{i3} are given in the following table.

Stratum i	N _i	W _i	S_{i1}	<i>S</i> _{<i>i</i>2}	<i>S</i> _{<i>i</i>3}	C_{i1}	C_{i2}	<i>C</i> _{<i>i</i>3}
1	18	0.30	2	3	1	0.6	0.9	1.5
2	27	0.45	4	1	7	0.8	1.2	2.0

The variance coefficients matrix is given by:

$$(a_{ij}) = \begin{pmatrix} 0.36 & 0.81 & 0.09 \\ 3.24 & 0.2025 & 9.925 \end{pmatrix}$$

The multivariate allocation problem is stated as:

Minimize $Z_1 = \frac{0.36}{X_1} + \frac{3.24}{X_2}, \quad Z_2 = \frac{0.81}{X_1} + \frac{0.2025}{X_2}, \quad Z_3 = \frac{0.09}{X_1} + \frac{9.925}{X_2}$ Subject to $3X_1 + 4X_2 \le 100$ $2 \le X_1 \le 18$ $2 \le X_2 \le 27$ (7.1)

The lower limits over the sample numbers in the two strata are taken at 2 as one would like to draw a sample of at least two units from each stratum.

The solutions X_1^0 , X_2^0 and X_3^0 corresponding to the three objective functions Z_1 , Z_2 and Z_3 by solving LPP (3.3) for k = 1, 2, 3 are obtained as:

$$X_1^0 = (7.6923, 19.2308)$$
 with $Z_1^0 = 0.2153$
 $X_2^0 = (14.8, 13.9)$ with $Z_2^0 = 0.0693$
 $X_3^0 = (3.1983, 22.6012)$ with $Z_3^0 = 0.4672$

These optimal values of Z_1^0 , Z_2^0 and Z_3^0 are used as aspiration levels in the Chebyshev approximation model.

The Chebyshev approximation model (4.1) yields the following LPP:

Minimize

Subject to $3X_1 + 4X_2 \le 100$

δ

 $-0.0900X_1 - 0.8100X_2 - X_3 \le -3.6000$ $-0.0715X_1 - 0.1842X_2 - X_3 \le -1.8660$ $-0.0010X_1 - 0.3291X_2 - X_3 \le -2.1050$ $-0.0038X_1 - 0.0800X_2 - X_3 \le -1.0926$ $-0.0900X_1 - 0.0176X_2 - X_3 \le -0.8382$ $-0.0135X_1 - 0.0073X_2 - X_3 \le -0.4460$ $-0.0011X_1 - 0.0245X_2 - X_3 \le -0.6034$ $-0.0031X_1 - 0.0113X_2 - X_3 \le -0.4498$ $-0.0070X_1 - 0.0084X_2 - X_3 \le -0.4306$ $-0.0047X_1 - 0.0095X_2 - X_3 \le -0.4338$ $-0.0060X_1 - 0.0088X_2 - X_3 \le -0.4306$ $-0.2025X_1 - 0.0506X_2 - X_4 \le -1.0126$ $-0.0428X_1 - 0.0303X_2 - X_4 \le -0.5290$ $-0.2025X_1 - 0.00095X_2 - X_4 \le -0.8378$ $-0.0487X_1 - 0.0015X_2 - X_4 \le -0.4320$ $-0.0107X_1 - 0.0076X_2 - X_4 \le -0.2650$ $-0.0225X_1 - 2.4813X_2 - X_5 \le -10.0150$ $-0.0224X_1 - 0.6149X_2 - X_5 \le -5.0306$ $-0.0186X_1 - 0.1513X_2 - X_5 \le -2.5328$ $-0.0005X_1 - 0.0435X_2 - X_5 \le -1.3270$ $-0.0225X_1 - 0.0180X_2 - X_5 \le -0.9346$ $-0.0225X_1 - 0.0236X_2 - X_5 \le -0.9972$ $X_3 - \delta \le 0.2153$ $X_4 - \delta \leq 0.0693$ $X_5 - \delta \le 0.4672$ $2 \leq X_1 \leq 18$ $2 \le X_2 \le 27$ $0.001 \le X_3 \le 30$ $0.001 \le X_4 \le 30$ $0.001 \le X_5 \le 30$ $0.001 \le X_6 \le 30$.

The Chebyshev point by solving the above problem is $X_{ch}^* = (6.1362, 20.3978)$ with $\delta = 0.0333$. The values of the three objective functions at this point are $Z_{ch}^1 = 0.2175$, $Z_{ch}^2 = 0.1419$ and $Z_{ch}^3 = 0.5013$.

Further the values of Z_1 at the points X_2^0 and X_3^0 are 0.2574 and 0.2560, the values of Z_2 at the points X_1^0 and X_3^0 are 0.1158 and 0.2623 and the values of Z_3 at the points X_1^0 and X_2^0 are 0.5298 and 0.7201. So,

$$L_1 = 0.2153, U_1 = 0.2574$$

$$L_2 = 0.0693, U_2 = 0.2623$$

$$L_3 = 0.4672, U_3 = 0.7201$$

$$d_1 = 0.0421, d_2 = 0.1930, d_3 = 0.2529.$$

The fuzzy point for the given problem by solving the LPP (6.1) is $X_{fz}^* = (6.2694, 20.2980)$ with $\delta = 0.1396$. The values of the objective functions at this point are $Z_{fz}^1 = 0.2170$, $Z_{fz}^2 = 0.1302$, $Z_{fz}^3 = 0.5034$.

The solution obtained by fuzzy programming gives a slight improvement over the Chebyshev solution in the value of the second objective function only.

References

- 1. S. Chatterjee, A note on Optimum allocation, Skand. Akt. 50 (1967), 40-44.
- 2. O.I. Charnayak and A. Slarytsky, Optimal allocation in stratified sampling with convex cost function, *Visnyk of Kviv. University, Economics* **39** (1998), 42–46 (in Ukrainian).
- 3. O.I. Charnayak and G. Chornous Optimal allocation in Stratified sampling with a non-linear cost function, *Theory of Stochastic Processes* 6(22) **3–4** (2000), 6–17.
- 4. W.G. Cochran, Sampling Techniques, 3rd Ed., John Wiley and Sons, New York, 1977.
- N. Jahan, M.I.M. Khan and M.J. Ahsan, A generalized compromise allocation, *Journal of the Indian Statistical Association* 32 (1994), 95–101.
- 6. J.E. Kelly, The cutting plane method for solving convex programs, *Jour. Soc. Indust. Appl. Math.* **8** (1960), 703–712.
- A.R. Kokan and S. Khan, Optimum allocation in multivariate surveys: An analytical solution, Journal of Royal Statistical Society, Ser. B 29 (1967), 115–125.

Keywords: Chebyshev approximation, cutting plane, fuzzy solution, multivariate stratified sampling.