

Chebyshev Approximate Solution to Allocation Problem in Multiple Objective Surveys with Random Costs

Mohammed Faisal Khan^{1*}, Irfan Ali², Qazi Shoeb Ahmad¹

¹Department of Mathematics, Integral University, Lucknow, India

²Department of Statistics & Operations Research, Aligarh Muslim University, Aligarh, India

E-mail: *faisalkhan004@yahoo.com

Received July 23, 2011; revised August 29, 2011; accepted September 8, 2011

Abstract

In this paper, we consider an allocation problem in multivariate surveys as a convex programming problem with non-linear objective functions and a single stochastic cost constraint. The stochastic constraint is converted into an equivalent deterministic one by using chance constrained programming. The resulting multi-objective convex programming problem is then solved by Chebyshev approximation technique. A numerical example is presented to illustrate the computational procedure.

Keywords: Chance Constrained Programming, Multivariate Stratified Sampling, Optimum Allocation, Chebyshev Approximation

1. Introduction

Optimum allocation of sample sizes to various strata in univariate stratified random sampling is well defined in the literature. But usually in real life problems 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 various characters is generally much different. Further the strata variances for the various characters may not be distributed in the same way. Allocation based on one character may not be optimum for the others. One way to resolve this problem is to search for a compromise allocation, which is in some sense optimum for all the characters.

Kokan and Khan [1], Chatterjee [2], Huddleston [3], Bethel [4], Chromy [5] all discussed the use of convex programming in relation to multivariate optimal allocation problem. The above convex programming approaches give the optimal solution to the problem with given tolerance limits on variances but the resulting cost may not be acceptable so that a further search is usually required for an optimal solution which falls within the budgetary constraint limit.

The case when sampling variances are random in the constraints has been dealt with by Diaz-Garcia [6]. Javaid and Bakhshi [7] applied modified E-model for solving the multivariate allocation problem when the costs are con-

sidered random in the objective function. Bakhshi [8] find the optimal Sample Numbers with a Probabilistic Cost Constraint.

In this paper, we consider the problem of allocating the sample to various strata when several characters are under study and the budget is fixed. We minimize the variances of various characters subject to the condition of given budget. The problem is transformed to a convex programming problem (CPP) with several linear objective functions and single convex constraint. The resulting CPP is then solved by Chebyshev approximation approach.

2. Formulation of the Problem

We consider a multivariate population consisting of N units which is divided into L disjoint strata of sizes N_1, N_2, \dots, N_L such that $N = \sum_{i=1}^L N_i$. Suppose that p characteristics ($j = 1, \dots, p$) are measured on each unit of the population. We assume that the strata boundaries are fixed in advance. Let n_i units be drawn without replacement from the i^{th} stratum $i = 1, \dots, L$. For j^{th} character, an unbiased estimate of the population mean \bar{Y}_j ($j = 1, \dots, p$) denoted by \bar{y}_{jst} , has its sampling variance

$$V(\bar{y}_{jst}) = \sum_{i=1}^L \left(\frac{1}{n_i} - \frac{1}{N_i} \right) W_i^2 S_{ij}^2, j = 1, \dots, p \quad (2.1)$$

where $W_i = \frac{N_i}{N}$ is the stratum weight and

$S_{ij}^2 = \frac{1}{N_i - 1} \sum_{h=1}^{N_i} (y_{ijh} - \bar{Y}_{ij})^2$ is the variance for the j^{th}

character in the i^{th} stratum.

Let c_{ij} be the cost of enumerating the j^{th} character in the i^{th} stratum and let the overhead cost c_0 be constant. Let \bar{C} be the upper limit on the total cost of the survey. Then assuming linear cost function one should have

$$\sum_{i=1}^L \sum_{j=1}^p c_{ij} n_i + c_0 \leq \bar{C} \quad \text{or} \quad \sum_{i=1}^L c_i n_i \leq C, \quad (2.2)$$

where $c_i = \sum_{j=1}^p c_{ij}$ is the cost of enumeration of all the p character in the i^{th} stratum and $C = \bar{C} - c_0$.

The restrictions on the sample size from various strata are

$$2 \leq n_i \leq N_i, \quad i = 1, \dots, L \quad (2.3)$$

Let us assume that the survey is to be conducted in such a way that the variances for all the p characters are minimized for a fixed budget *i.e.*, we have to minimize all the variances together given by (2.1).

Ignoring the constant terms in (2.1), the NLP problems to be solved are

$$\left. \begin{aligned} \min_n \cdot V &= \sum_{i=1}^L \frac{W_i^2 S_{ij}^2}{n_i}, \quad j = 1, \dots, p \\ \text{Subject to} & \\ \sum_{i=1}^L c_i n_i &\leq C \\ \text{and } 2 \leq n_i &\leq N_i, \quad i = 1, \dots, L, \quad n_i \in N \end{aligned} \right\} \quad (2.4)$$

By making the transformation $n_i = \frac{1}{x_i}, i = 1, \dots, p$ and putting $a_{ij} = W_i^2 S_{ij}^2$, the problems (2.4) are transformed into

$$\left. \begin{aligned} \min \sum_{i=1}^L a_{ij} x_i, \quad j = 1, \dots, p, & \quad (1) \\ \text{Subject to} & \\ \sum_{i=1}^L \frac{c_i}{x_i} \leq C & \quad (2) \\ \text{and } \frac{1}{N} \leq x_i \leq \frac{1}{2}, \quad i = 1, \dots, L & \quad (3) \end{aligned} \right\} \quad (2.5)$$

In many practical situations the costs c_i in the various strata are not fixed and vary from one unit to the other. Let us assume that $c_i, i = 1, \dots, L$ are independently normally distributed random variables. So, we write the above problem in the following chance constrained programming form:

$$\left. \begin{aligned} \min \sum_{i=1}^L a_{ij} x_j, \quad j = 1, \dots, p & \quad (1) \\ \text{Subject to} & \\ P \left(\sum_{i=1}^L \frac{c_i}{x_i} \leq C \right) \geq p_0 & \quad (2) \\ 2 \leq n_i \leq N_i, \quad i = 1, \dots, L, \quad n_i \in N & \quad (3) \end{aligned} \right\} \quad (2.6)$$

where $p_0, 0 \leq p_0 \leq 1$ is a specified probability.

3. Deterministic Equivalent Using Chance Constrained Programming

We have assumed that the costs $c_i, i = 1, \dots, L$ in the constraint function 2.6 (2) are independently and normally distributed random variables. Then function $\sum_{i=1}^L \frac{c_i}{x_i}$, will

also be normally distributed with mean as

$$E \left(\frac{\sum_{i=1}^L c_i}{x_i} \right) = \frac{\sum_{i=1}^L E(c_i)}{x_i} = \sum_{i=1}^L \frac{\mu_i}{x_i}, \quad (3.1)$$

where $\mu_i = E(c_i), i = 1, \dots, L$, and variance as

$$V \left(\sum_{i=1}^L \frac{c_i}{x_i} \right) = \sum_{i=1}^L \frac{1}{x_i^2} V(c_i) = \sum_{i=1}^L \frac{\sigma_i^2}{x_i^2} \quad (3.2)$$

where $\sigma_i^2 = V(c_i)$.

Now let $f(c) = \sum_{i=1}^L \frac{c_i}{x_i}$, where $c = (c_1, \dots, c_n)$, then {2.6 (2)} is given by

$$P(f(c) \leq C) \geq p_0,$$

or

$$P \left\{ \frac{f(c) - E(f(c))}{\sqrt{V(f(c))}} \leq \frac{C - E(f(c))}{\sqrt{V(f(c))}} \right\} \geq p_0,$$

where $\left[\frac{f(c) - E\{f(c)\}}{\sqrt{V\{f(c)\}}} \right]$ is a standard normal variable

with mean zero and variance one. Thus the probability of realizing $\{f(c)\}$ less than or equal to C can be written as

$$P(f(c) \leq C) = \phi \left[\frac{C - E\{f(c)\}}{\sqrt{V\{f(c)\}}} \right], \quad (3.3)$$

where $\phi(z)$ represents the cumulative density function of the standard normal variable evaluated at z . If K_α represents the value of the standard normal variable at which $\phi(K_\alpha) = p_0$, then the constraint (3.3) can be writ-

ten as

$$\phi \left[\frac{C - E\{f(c)\}}{\sqrt{V\{f(c)\}}} \right] \geq \phi(K_\alpha). \tag{3.4}$$

The inequality (3.4) will be satisfied only if

$$\left[\frac{C - E\{f(c)\}}{\sqrt{V\{f(c)\}}} \right] \geq K_\alpha,$$

or equivalently,

$$E(f(c)) + K_\alpha \sqrt{V(f(c))} \leq C. \tag{3.5}$$

Substituting from (3.1) and (3.2) in (3.5), we get

$$\left(\sum_{i=1}^L \frac{\mu_i}{x_i} \right) + K_\alpha \sqrt{\sum_{i=1}^L \frac{\sigma_i^2}{x_i^2}} \leq C. \tag{3.6}$$

If the constants μ_i and σ_i in (3.6) are unknown then we use their estimators $\hat{\mu}_i$ and $\hat{\sigma}_i^2$.

Thus

$$\hat{E} \left(\sum \frac{c_i}{x_i} \right) = \sum \left[\frac{\hat{E}(c_i)}{x_i} \right] = \sum_{i=1}^L \frac{\bar{c}_i}{x_i}, \text{ say,}$$

$$\hat{V} \left(\sum \frac{c_i}{x_i} \right) = \sum_{i=1}^L \frac{\bar{\sigma}_{c_i}^2}{x_i^2}, \text{ say,}$$

where \bar{c}_i and $\bar{\sigma}_{c_i}^2$ are the estimated means and variances from the sample.

Thus, an equivalent deterministic constraint to the stochastic constraint 2.6 (2) is given by

$$\left(\sum_{i=1}^L \frac{\bar{c}_i}{x_i} + c_0 \right) + K_\alpha \sqrt{\sum_{i=1}^L \frac{\bar{\sigma}_{c_i}^2}{x_i^2}} \leq C.$$

The equivalent deterministic non-linear programming problem to the chance constrained programming problem (2.6) is obtained as

$$\left. \begin{aligned} \min \cdot V_j &= \sum_{i=1}^L a_{ij} x_j, \quad j = 1, \dots, p & (1) \\ \text{Subject to} & \\ \left(\sum_{i=1}^L \frac{\bar{c}_i}{x_i} \right) + K_\alpha \sqrt{\sum_{i=1}^L \frac{\bar{\sigma}_{c_i}^2}{x_i^2}} &\leq C & (2) \\ \frac{1}{N_i} \leq x_i \leq \frac{1}{2}, \quad i &= 1, \dots, L & (3). \end{aligned} \right\} \tag{3.7}$$

4. Convex Chebyshev Approximation Problem

Consider p convex smooth functions

$$g_j(\underline{x}) = g_j(x_1, \dots, x_n), \quad j = 1, \dots, p \tag{4.1}$$

and a region Ω defined by q inequalities

$$\Psi_i(\underline{x}) = \Psi_i(x_1, \dots, x_n) \leq 0, \quad i = 1, \dots, q \tag{4.2}$$

where Ψ_i are also convex smooth functions.

The Convex Chebyshev Approximation Problem (CCAP) for minimizing the system (4.1) under Constraints (4.2) consists in finding $\underline{x}^* \in \Omega$ for which

$$\max_j g_j(\underline{x}) = \min_{\underline{x} \in \Omega} \max_j f_j(\underline{x}). \tag{4.3}$$

Since $\max_j g_j(x)$ is convex as can be seen from the

Figure 1 below, the convex Chebyshev approximation problem is convex.

Corresponding to the points $(x_1, \dots, x_5) \in \Omega$, we have

$$\begin{aligned} \max_i f_i(x_k) &= \{f_4(x_1), f_3(x_2), f_2(x_3), \\ &= f_3(x_3), f_1(x_4) = f_2(x_4), f_1(x_5)\} \end{aligned}$$

In the general (CCAP) (4.1) & (4.2) we introduce an auxiliary variable x_{n+1} and the auxiliary constraints $a_j g_j(\underline{x}) \leq x_{n+1}$, where a_j are some constants. The problem (4.3) then is equivalent to

$$\left. \begin{aligned} \min Z &= x_{n+1} \\ \text{subject to} & \\ a_j g_j(x_1, \dots, x_n) - x_{n+1} &\leq 0, \quad j = 1, \dots, p \\ \text{and } \Psi_i(x_1, \dots, x_n) &\leq 0, \quad i = 1, \dots, q \end{aligned} \right\} \tag{4.4}$$

5. Solutions Using Chebyshev Approximation Technique

The p objective functions in 3.7 (1) are linear. The single constraint 3.7 (2) is convex (see Kokan and Khan [1]). So (3.7) represents p convex programming problems. Let us denote the feasible region defined by 3.7 (2) and (3) introduce an auxiliary variable $x_{L+1}, j = 1, \dots, p$. From

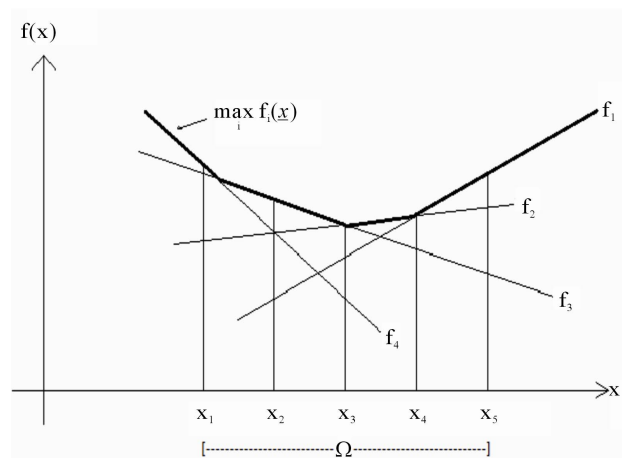


Figure 1. Convexity of the function $\max_i f_i(\underline{x})$.

by Ω . Suppose that the feasible region is not void. Let us (4.1) to (4.3) it follows that the problem (3.7) is equivalent to the convex Chabyshev's approximation problem of finding $\underline{x}^* \in \Omega$ such that

$$\max_j a_j V_j(\underline{x}) = \min_{\underline{x} \in \Omega} \max_j a_j V_j(\underline{x}), \tag{5.1}$$

where a_j are the weights assigned to the p variances according to their importance. The problem (5.1) is then equivalent to the following problem with a linear objective function:

$$\left. \begin{aligned} Z &= x_{L+1} & (1) \\ \text{subject to} & & \\ a_j V_j(\underline{x}) &\leq x_{L+1} \text{ or } a_j \sum_{i=1}^L a_{ij} x_i - x_{L+1} \leq 0, j=1, \dots, p & (2) \\ \sum_{i=1}^L \frac{\bar{c}_i}{x_i} + K_\alpha \sqrt{\sum_{i=1}^L \frac{\bar{\sigma}_i^2}{x_i^2}} &\leq C & (3) \\ \text{and } \frac{1}{N_i} &\leq x_i \leq \frac{1}{2}, i=1, \dots, L & (4) \end{aligned} \right\} (5.2)$$

The non-linear programming problem in (5.2) is convex as the objective functions in 5.2 (1) and the constraint 5.2 (2) are linear. Further the left hand side in 5.2 (2) is convex. So it is possible to solve the convex programming problem (5.2) by using any standard convex programming algorithm. The optimal sample numbers thus obtained may turn out to be fractional. However, it is known that the variance functions are flat at the optimum solution. So for large or even moderate sample size it is enough to round the fractional values to the nearest integers. However, for small $n_i = 1/x_i$ the branch and bound method should be applied for finding the optimal integer solution.

6. Numerical Illustration

The following numerical example demonstrates the solu-

$$\left. \begin{aligned} \min \cdot &= X_5 \\ \text{Subject to} & & \\ 0.75(552.640X_1 + 136.277X_2 + 274.114X_3 + 2588.343X_4) - X_5 &\leq 0 \\ 0.25(14926.197X_1 + 165.9747X_2 + 130.202X_3 + 3084.324X_4) - X_5 &\leq 0 \\ \left(\frac{3}{X_1} + \frac{4}{X_2} + \frac{5}{X_3} + \frac{7}{X_4} \right) + 2.33 \sqrt{\frac{0.6}{X_1^2} + \frac{0.5}{X_2^2} + \frac{0.7}{X_3^2} + \frac{0.8}{X_4^2}} &\leq 500 \\ \frac{1}{1419} &\leq X_1 \leq \frac{1}{2}, \frac{1}{619} \leq X_2 \leq \frac{1}{2}, \frac{1}{1253} \leq X_3 \leq \frac{1}{2}, \frac{1}{899} \leq X_4 \leq \frac{1}{2}. \end{aligned} \right\} (6.1)$$

The Chebyshev point by solving the convex programming problem (6.1) is

$$X_{ch}^* = (0.02159, 0.12169, 0.10866, 0.03882) \text{ with } X_5 = 119.0883$$

tion procedure. The data used in this example is from a stratified random sample survey conducted in Varanasi district of Uttar Pradesh (U.P), India to study the distribution of manurial resources among different crops and cultural practices (see Sukhatme[9]). Relevant data with respect to the two characteristics ‘‘area under rice’’ and ‘‘total cultivated area’’ are given in **Table 1**. The total number of villages in the district was 4190.

In order to demonstrate the procedure the following are also assumed. The per unit travel costs c_i , ($i=1, \dots, 4$) of measurement in various strata are independently normally distributed with the following means and variances $E(c_1)=3$, $E(c_2)=4$, $E(c_3)=5$, $E(c_4)=7$ and $V(c_1)=0.6$, $V(c_2)=0.5$, $V(c_3)=0.7$, $V(c_4)=0.8$

Let us assign the weights to the variances of the two characters in proportion to the inverse of the sums

$$\sum_{i=1}^4 S_{i1} \text{ and } \sum_{i=1}^4 S_{i2} \text{ which turn out to be } a_1 = 0.75 \text{ and } a_2 = 0.25.$$

The total amount available for the survey C is assumed as 600 units including an expected overhead cost $t_0 = 100$ units.

Let the chance constraint 2.6 (2) be required to be satisfied with 99% probability. Then k_α is such that $\phi(k_\alpha) = 0.99$. The value of standard normal variable K_α corresponding to 99% confidence limits is 2.33. Thus, the problem (5.2) is obtained as:

Table 1. Data for four strata and two characteristics.

Stratum i	N_i	W_i	S_{i1}^2	S_{i2}^2
1	1419	0.3387	4817.72	130121.15
2	619	0.1477	6251.26	7613.52
3	1253	0.2990	3066.16	1456.40
4	899	0.2146	56207.25	66977.72

The values of sample sizes rounded to nearest integers are $n_1 = 46$, $n_2 = 8$, $n_3 = 9$ and $n_4 = 26$, with a total of 89. Corresponding to this allocation the values of the variances for the two characters are obtained as $V_1 = 159.05$, $V_2 = 478.32$

Remark: We may compare these results with the compromise solution (Cochran [10]) which is obtained by solving the following NLP problem:

$$\begin{aligned} & \min_n V \\ & = 0.75 \left(\frac{552.640}{n_1} + \frac{136.277}{n_2} + \frac{274.114}{n_3} + \frac{2588.343}{n_4} \right) \\ & + 0.25 \left(\frac{14926.197}{n_1} + \frac{165.39747}{n_2} + \frac{130.202}{n_3} + \frac{3084.324}{n_4} \right) \end{aligned}$$

Subject to

$$\begin{aligned} & (3n_1 + 4n_2 + 5n_3 + 7n_4) \\ & + 2.33\sqrt{0.6n_1^2 + 0.5n_2^2 + 0.7n_3^2 + 0.8n_4^2} \leq 500 \\ & 2 \leq n_1 \leq 1419, 2 \leq n_2 \leq 619, 2 \leq n_3 \leq 1253, 2 \leq n_4 \leq 899 \end{aligned}$$

The integer solution is obtained as $n_1 = 44$, $n_2 = 9$, $n_3 = 10$ and $n_4 = 29$ with a total of 92. The values of the individual variances, corresponding to this allocation are obtained as $V_1 = 144.36$ and $V_2 = 476.90$.

7. Conclusions

The optimum allocation problem in multivariate stratified sampling with random costs has been formulated as a problem in multiple linear objectives under the single probabilistic constraint. An equivalent deterministic model of the stochastic programming problem is established by using Chance Constrained programming method. The problem is then solved by using the Chebyshev approximation technique. A comparative study of the increases in the variance using Chebyshev approximation as compared to the compromise allocation is also presented.

8. References

- [1] A. R. Kokan and S. Khan, "Optimum Allocation in Multivariate Surveys: An Analytical Solution," *Journal of the Royal Statistical Society. Series B*, Vol. 29, No. 1, 1967, pp. 115-125.
- [2] S. Chatterjee, "Multivariate Stratified Surveys," *Journal of the American Statistical Association*, Vol. 63, No. 322, 1968, pp. 530-534.
- [3] H. F. Huddleston, et al., "Optimal Sample Allocation to Strata Using Convex Programming," *Journal of the Royal Statistical Society. Series C*, Vol. 19, No. 3, 1970, pp. 273-278.
- [4] J. Bethel, "An Optimum Allocation Algorithm for Multivariate Survey," *Proceeding of the Survey Research Section*, American Statistical Association, 1985, pp. 204-212.
- [5] J. R. Chromy, "Design Optimization with Multiple Objectives," *Proceeding of the Survey Research Section*, American Statistical Association, 1987, pp. 194-199.
- [6] J. A. Diaz-Garcia and M. M. Garay-Tapia, "Optimum allocation in Stratified surveys: Stochastic Programming," *Computational Statistics and Data Analysis*, Vol. 51, No. 6, 2007, pp. 3016-3026. [doi:10.1016/j.csda.2006.01.016](https://doi.org/10.1016/j.csda.2006.01.016)
- [7] S. Javed, Z. H. Bakhshi and M. M. Khalid, "Optimum allocation in Stratified Sampling with Random Costs," *International Review of Pure and Applied Mathematics*, Vol. 5, No. 2, 2009, pp. 363-370.
- [8] Z. H. Bakhshi, M. F. Khan and Q. S. Ahmad, "Optimal Sample Numbers in Multivariate Stratified Sampling with a Probabilistic Cost Constraint," *International Journal of Mathematics and Applied Statistics*, Vol. 1, No. 2, 2010, pp. 111-120.
- [9] P. V. Sukhatme, B. V. Sukhatme, S. Sukhatme and C. Asok, "Sampling Theory of Surveys with Applications," 3rd Edition, Iowa State University Press, Ames, 1984.
- [10] W. G. Cochran, "Sampling Techniques," 3rd Edition, John Wiley and Sons, New York, 1977.