# 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 $(i=1,2, \cdots, L)$. For the $j$-th character an unbiased estimate of the population mean, $\bar{Y}_{j}$, is $\bar{y}_{j s t}$ which has the sampling variance.

$$
\begin{equation*}
V_{j}=V\left(\bar{y}_{j s t}\right)=\sum_{i=1}^{L} W_{i}^{2} S_{i j}^{2} X_{i} \quad j=1,2, \cdots, p \tag{2.1}
\end{equation*}
$$

where

$$
W_{i}=\frac{N_{i}}{N}, \quad S_{i j}^{2}=\frac{1}{N_{i}-1} \sum_{h=1}^{N_{i}}\left(y_{i j t h}-\bar{Y}_{i j}\right)^{2}
$$

and

$$
X_{i}=\frac{1}{n_{i}}-\frac{1}{N_{i}}, \quad a_{i j}=W_{i}^{2} S_{i j}^{2}
$$

in usual notation.
Let $c_{i j}$ 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

$$
\begin{equation*}
\sum_{i=1}^{L} \sum_{j=1}^{p} c_{i j} n_{i} \leq k \tag{2.2}
\end{equation*}
$$

The multivariate allocation problem can be stated as:
Minimize $\quad z_{j}=\sum_{i=1}^{L} a_{i j} X_{i}-\sum_{i=1}^{L} \frac{a_{i j}}{N_{i}}, \quad j=1,2, \cdots, p$.

Subject to

$$
\begin{align*}
& \sum_{i=1}^{L} \sum_{j=1}^{p} \frac{c_{i j}}{X_{i}} \leq k \\
& \frac{1}{N_{i}} \leq X_{i} \leq 1, \quad i=1,2,3, \cdots, L \tag{2.3}
\end{align*}
$$

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

Minimize

$$
Z_{k}=\sum_{i=1}^{L} \frac{a_{i k}}{X_{i}}
$$

Subject to

$$
\begin{align*}
& \sum_{i=1}^{L} \sum_{j=1}^{p} c_{i j} X_{i} \leq k  \tag{2.4}\\
& 1 \leq X_{i} \leq N_{i}, \quad i=1,2, \cdots, L
\end{align*}
$$

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


The constraints in (2.5b) are convex (see Kokan and Khan [7]) and the constraint (2.5c) and the bounds ( 2.5 d ) 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 $k=1,2, \cdots, p$ 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 $X^{k(0)}=\left(X_{1}{ }^{k(0)}, \cdots, X_{L}^{k(0)}, X_{L+k}^{k(0)}\right)$ be the solution of LPP, which minimizes (2.5a) subject to the single constraint (2.5c) and the bounds (2.5d).

Compute $\quad g_{k}\left(X^{k(0)}\right)=\sum_{i=1}^{L} \frac{a_{i k}}{x_{i}^{k(0)}}-X_{L+k}^{k(0)}$
Define $\in$ to be a small tolerance limit for the convergence.
If $g_{k}\left(X^{k(0)}\right) \leq \in$, this means that (2.5b) as also satisfied to the tolerance limits and thus $X^{k(0)}$ solves the problem.

If $g_{k}\left(X^{k(0)}\right)>\in$, we linearize the convex constraint $g_{k}(X) \leq 0$ about the point $X^{k(0)}$ as:

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

Where

$$
\nabla g_{k}\left(X^{k(0)}\right)^{\prime}=\left[-\sum_{j=1}^{P} \frac{c_{1 j}}{X_{1}^{k(0)^{2}}}, \cdots,-\sum_{j=1}^{P} \frac{c_{L j}}{X_{L}^{k(0)^{2}}}, \quad-1\right]
$$

This give

$$
\begin{equation*}
2 \sum_{i=1}^{L} \frac{a_{i k}}{x_{i}^{k(0)}}-\sum_{j=1}^{p} \frac{a_{i k} x_{i}}{X_{i}^{k(0)^{2}}}-X_{L+k} \leq 0 \tag{3.1}
\end{equation*}
$$

The constraint (2.5b) is then replaced by this linearized constraint (3.1). The following LPP approximates the NLPP (2.5).
$\left.\begin{array}{ll}\text { Minimize } & Z_{k}=X_{L+k} \\ \text { Subject to } & 2 \sum_{i=1}^{L} \frac{a_{i k}}{X_{i}^{K(0)}}-\sum_{i=1}^{L} \frac{a_{i k} X_{i}}{X_{i}^{k(0)^{2}}}-X_{L+k} \leq 0 \\ & \sum_{i=1}^{L} \sum_{j=1}^{P} c_{i j} X_{i} \leq k \\ & 1 \leq X_{i} \leq N_{i}, \quad i=1,2, \cdots, L\end{array}\right\}$

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

$$
\begin{equation*}
g_{k}\left(X^{k(t)}\right)=\sum_{i=1}^{L} \frac{a_{i k}}{X_{i}^{k(t)}}-X_{L+k}^{k(t)} . \tag{3.3}
\end{equation*}
$$

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

$$
X^{k(t)}=X_{\text {opt }} \quad \text { and } \quad Z_{k}\left(X^{k(t)}\right)=Z_{\text {opt }} .
$$

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 $t_{k}^{*}-t h$ iteration. After getting $t_{k}^{*}$, we solve the following LPP's for $k=1,2, \cdots, p$. Here it is noted that each LPP consists of $1+p+p\left(\sum_{1}^{p} t_{k}^{*}\right)$ linear constraints and 2L lower and upper bounds.

Minimize $\quad Z_{k}=X_{L+k}$
Subject to $\quad 2 \sum_{i=1}^{L} \frac{a_{i k}}{x_{i}^{k(0)}}-\sum_{i=1}^{L} \frac{a_{i k} x_{i}}{x_{i}^{k(0)^{2}}}-x_{L+k} \leq 0, \quad k=1,2, \cdots, p$

$$
\begin{equation*}
2 \sum_{i=1}^{L} \frac{a_{i k}}{x_{i}^{k(l)}}-\sum_{i=1}^{L} \frac{a_{i k} x_{i}}{x_{i}^{k(l)^{2}}}-x_{L+k} \leq 0 \tag{3.4}
\end{equation*}
$$

$$
\begin{array}{ll}
\sum_{i=1}^{L} \sum_{j=1}^{p} c_{i j} X_{i} \leq k & l=1,2, \cdots, t_{k}^{*} \\
1 \leq X_{1} \leq N_{i}, & k=1,2, \cdots, p \\
i=1,2, \cdots, L
\end{array}
$$

Let the minimum values of $Z_{k}$ thus found be $Z_{k}^{o}, k=1,2, \cdots, p$ at the corresponding minimal points $X_{k}^{o}, k=1,2, \cdots, p$. The $p$ solutions $X_{1}^{o}, \cdots, X_{p}^{o}$ 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:

Minimize $\delta$
Subject to $2 \sum_{i=1}^{L} \frac{a_{i k}}{X_{i}^{k(0)}}-\sum_{i=1}^{L} \frac{a_{i k} X_{i}}{X_{i}^{k(0)^{2}}}-X_{L+K} \leq 0$

$$
\begin{array}{lr}
2 \sum_{i=1}^{L} \frac{a_{i k}}{X_{i}^{k(l)}}-\sum_{i=1}^{L} \frac{a_{i k} X_{i}}{X_{i}^{k(l)^{2}}}-X_{L+K} \leq 0 \\
\sum_{i=1}^{L} \sum_{j=1}^{p} c_{i j} X_{i} \leq k & l=1,2, \cdots, t_{k}^{*}  \tag{4.1}\\
X_{L+k}-\delta \leq Z_{k}^{0} & k=1,2, \cdots, p
\end{array}
$$

$$
1 \leq X_{i} \leq N_{i} \quad l=1,2, \cdots, L
$$

Where $\delta$ (dummy variable) represents the worst deviation level. Note that the aspiration levels are being taken as the minimum values of the objective functions at $Z_{k}^{o}, k=1,2, \cdots, p$. 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 $g_{k}\left(X^{k(t)}\right) \leq \in$, for some $t$, say $t_{k}^{*}$, where $\in$ is a suitable tolerance limit, then $X^{k\left(t_{k}^{*}\right)}=X_{k}^{o}$ and $Z_{k}^{o}=Z_{k}\left(X^{k\left(t_{k}^{*}\right)}\right)$ 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 $k=1,2, \cdots, p$ to obtain $X_{k}^{o}$, the approximate minimal points for the respective objective functions, with minimum corresponding values of $Z_{k}$ as $Z_{k}^{o}$.

Step IV. Solve the Chebyshev approximation model (4.1) of the multivariate allocation problem (2.5) to obtain $X_{c h}^{*}$ (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 $Z_{k}\left(X_{j}^{o}\right)=Z_{k}^{j o}, j=1,2, \cdots, p$. Clearly $Z_{k}^{k o}=Z_{k}^{o}=\min _{j} Z_{k}\left(x_{j}^{o}\right)$.
Denote $Z_{k}^{o}=L_{k}$ and $\max _{j} Z_{k}\left(X_{j}^{0}\right)=U_{k}$.
The differences of the maximum and minimum values of $Z_{k}$ are denoted by $d_{k}=U_{k}-L_{k}, \quad k=1,2, \cdots, p$.

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

Minimize $\delta$

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

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 $N_{i}, W_{i}$, $S_{i 1}, S_{i 2}$ and $S_{i 3}$ are given in the following table.

| Stratum <br> $\boldsymbol{i}$ | $\boldsymbol{N}_{\boldsymbol{i}}$ | $\boldsymbol{W}_{\boldsymbol{i}}$ | $\boldsymbol{S}_{\boldsymbol{i} 1}$ | $\boldsymbol{S}_{\boldsymbol{i} 2}$ | $\boldsymbol{S}_{\boldsymbol{i} 3}$ | $\boldsymbol{C}_{\boldsymbol{i} 1}$ | $\boldsymbol{C}_{\boldsymbol{i} 2}$ | $\boldsymbol{C}_{\boldsymbol{i} 3}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 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:

$$
\left(a_{i j}\right)=\left(\begin{array}{lll}
0.36 & 0.81 & 0.09 \\
3.24 & 0.2025 & 9.925
\end{array}\right)
$$

The multivariate allocation problem is stated as:
Minimize $\quad Z_{1}=\frac{0.36}{X_{1}}+\frac{3.24}{X_{2}}, \quad Z_{2}=\frac{0.81}{X_{1}}+\frac{0.2025}{X_{2}}, \quad Z_{3}=\frac{0.09}{X_{1}}+\frac{9.925}{X_{2}}$

Subject to $\quad 3 X_{1}+4 X_{2} \leq 100$

$$
\begin{align*}
& 2 \leq X_{1} \leq 18  \tag{7.1}\\
& 2 \leq X_{2} \leq 27
\end{align*}
$$

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:

$$
\begin{array}{lll}
X_{1}^{0}=(7.6923,19.2308) & \text { with } & Z_{1}^{0}=0.2153 \\
X_{2}^{0}=(14.8,13.9) & \text { with } & Z_{2}^{0}=0.0693 \\
X_{3}^{0}=(3.1983,22.6012) & \text { with } & Z_{3}^{0}=0.4672
\end{array}
$$

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 $\delta$
Subject to $\quad 3 X_{1}+4 X_{2} \leq 100$
$-0.0900 X_{1}-0.8100 X_{2}-X_{3} \leq-3.6000$
$-0.0715 X_{1}-0.1842 X_{2}-X_{3} \leq-1.8660$
$-0.0010 X_{1}-0.3291 X_{2}-X_{3} \leq-2.1050$
$-0.0038 X_{1}-0.0800 X_{2}-X_{3} \leq-1.0926$
$-0.0900 X_{1}-0.0176 X_{2}-X_{3} \leq-0.8382$
$-0.0135 X_{1}-0.0073 X_{2}-X_{3} \leq-0.4460$
$-0.0011 X_{1}-0.0245 X_{2}-X_{3} \leq-0.6034$
$-0.0031 X_{1}-0.0113 X_{2}-X_{3} \leq-0.4498$
$-0.0070 X_{1}-0.0084 X_{2}-X_{3} \leq-0.4306$
$-0.0047 X_{1}-0.0095 X_{2}-X_{3} \leq-0.4338$
$-0.0060 X_{1}-0.0088 X_{2}-X_{3} \leq-0.4306$
$-0.2025 X_{1}-0.0506 X_{2}-X_{4} \leq-1.0126$
$-0.0428 X_{1}-0.0303 X_{2}-X_{4} \leq-0.5290$
$-0.2025 X_{1}-0.00095 X_{2}-X_{4} \leq-0.8378$
$-0.0487 X_{1}-0.0015 X_{2}-X_{4} \leq-0.4320$
$-0.0107 X_{1}-0.0076 X_{2}-X_{4} \leq-0.2650$
$-0.0225 X_{1}-2.4813 X_{2}-X_{5} \leq-10.0150$
$-0.0224 X_{1}-0.6149 X_{2}-X_{5} \leq-5.0306$
$-0.0186 X_{1}-0.1513 X_{2}-X_{5} \leq-2.5328$
$-0.0005 X_{1}-0.0435 X_{2}-X_{5} \leq-1.3270$
$-0.0225 X_{1}-0.0180 X_{2}-X_{5} \leq-0.9346$
$-0.0225 X_{1}-0.0236 X_{2}-X_{5} \leq-0.9972$
$X_{3}-\delta \leq 0.2153$
$X_{4}-\delta \leq 0.0693$
$X_{5}-\delta \leq 0.4672$
$2 \leq X_{1} \leq 18$
$2 \leq X_{2} \leq 27$
$0.001 \leq X_{3} \leq 30$
$0.001 \leq X_{4} \leq 30$
$0.001 \leq X_{5} \leq 30$
$0.001 \leq X_{6} \leq 30$.

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

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,

$$
\begin{aligned}
& L_{1}=0.2153, U_{1}=0.2574 \\
& L_{2}=0.0693, U_{2}=0.2623 \\
& L_{3}=0.4672, U_{3}=0.7201 \\
& d_{1}=0.0421, \quad d_{2}=0.1930, \quad d_{3}=0.2529 .
\end{aligned}
$$

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

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.
5. 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.
7. 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.

