GEOMETRIC PROGRAMMING APPROACH IN THREE - STAGE SAMPLING DESIGN

Shafiullah, Irfan Ali and Abdul Bari

ABSTRACT
In this paper we have formulated the problem of allocation of sample sizes in three-stage sampling design as a convex programming problem with linear objective function and non linear constraints. A Geometric Programming technique is developed for the solution of the resulting mathematical programming problem. A numerical example is given for the illustration of computational details of the procedure.

Key Words: Non-linear programming, convex programming, geometric programming, three-stage sampling.

1. INTRODUCTION
In the three–stage sampling design the process of sub sampling can be carried to a third stage by sampling the subunits instead of enumerating them completely. For instance, in surveys to estimate crop production in India (see Sukhatme, 1947), the village is a convenient sampling unit. Within a village, only some of the fields growing the crop in question are selected, so that the field is a sub-unit. When a field is selected, only certain parts of it are cut for the determination of yield per acre; thus the sub unit itself is sampled. Here we have to find the optimal sample sizes \( n, m \) and \( k \) for all the three stages with the minimum cost.

The use of three stage sampling designs generally specifies three stages of selection: primary sampling units (PSUs) at the first stage, sub samples from PSUs at second stage as a secondary sampling units (SSUs) units and again sub samples from SSUs at third stage as a tertiary sampling units (TSUs).The three stage sampling designs are well analyzed when two variable is measured. Different methods are available for obtaining the optimum allocation of sampling units to each stage. Geometric programming (GP) is very much connected with geometrical concepts because this method based on geometric inequality .The sums and products of positive numbers are important properties of GP. The degree of difficulties in GP plays very important roles in the solution of mathematical programming problems. The degree of difficulty of a GP problem is defined as:

\[ \text{Degree of difficulty} = \text{total no. of terms} - \text{total no. of decision variables} - 1 \]

If the degree of difficulty of primal problem is zero, then unique dual feasible solution exists. If the problem has positive degree of difficulty, then the objective function can be maximized by finding the dual feasible region, and if there is negative degree of difficulty then inconsistency of the dual constraints may occur. Geometric programming (GP), a smooth, systematic and an effective non-linear programming method used for solving problems of sample surveys, engineering design that takes the form of convex programming. The convex programming problems occurring in Geometric Programming are generally represented by an exponential or power function. Duffin and Zener has done the work in the field of engineering design problems in the early 1960s, and further extended by Duffin et al. [2]. Engineering design problems was also solved by Shiang [5] and Shaojian et.al [3] with the help of GP. Davis and Rudolph [7] applied GP to optimal allocation of integrated samples in quality control. Ahmed and Charles [15] applied geometric programming to optimum allocation problems in multivariate double sampling. Recently many authors have done the work on the Geometric Programming and multi-objective Geometric Programming in different directions. Some of them are: Ojha and Biswal [11] has worked on Posynomial Geometric Programming Problems with Multiple Parameter. Ojha and Das [13] as has done the work on multi-objective Geometric Programming problem being cost coefficients as continuous function with weighted mean. Verma [9], Islam [12] developed fuzzy geometric programming

The presentation of the paper is as follows: The formulation of an allocation problem in a three-stage sampling design is discussed in section 2 and the solution procedure for solving above formulated problem with geometric programming approach is discussed in section 3. The illustrative numerical example with hypothetical data is then presented in section 4 and finally some comments and conclusions which are drawn from the discussion are given in section 5.

2. STATEMENT OF THE PROBLEM:

Let us consider the population consists of NMK elements grouped into N first-stage units of M second-stage units and K third stage units each. Let n, m and k is the corresponding sample sizes selected with equal probability and without replacement at each stage. Let \( y_{iju} \) be the value obtained for \( u^{th} \) third-stage unit in the \( j^{th} \) second-stage unit drawn from the \( i^{th} \) primary unit.

The relevant population means per third-stage unit are as follows:

\[
\bar{Y}_{ij} = \frac{\sum_{u=1}^{k} y_{iju}}{K} = \text{Sample mean per } j^{th} \text{ second-stage unit at the } i^{th} \text{ primary stage unit.}
\]

\[
\bar{Y}_{i} = \frac{\sum_{j=1}^{M} \sum_{u=1}^{k} y_{iju}}{MK} = \text{Mean per element at the primary stage unit.}
\]

\[
\bar{Y} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{u=1}^{k} y_{iju}}{NMK} = \text{Mean per element in the population.}
\]

The following population variances are required:

\[
S_i^2 = \frac{\sum_{j=1}^{N} (Y_j - \bar{Y})^2}{N-1} = \text{variance within primary stage unit means.}
\]

\[
S_2^2 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (\bar{Y}_{ij} - \bar{Y}_i)^2}{N(M-1)} = \text{variance among secondary-stage units within primary stage unit means.}
\]

\[
S_3^2 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{u=1}^{k} (y_{iju} - \bar{Y}_{ij})^2}{NM(K-1)} = \text{variance for tertiary stage-units among secondary stage-units within primary stage unit means.}
\]

In case of equal third-stage units an unbiased estimate of \( \bar{Y}_i \) is \( \bar{Y}_i \) with its sampling variance as,

\[
V(y) = \frac{1}{n} S_1^2 + \frac{1}{nm} S_2^2 + \frac{1}{nmk} S_3^2
\]

In three-stage sampling the total cost function may be given by:

\[
C = C_1 n + C_2 nm + C_3 nmk
\]

\[
c_1 = \text{The cost for } i^{th} \text{ primary stage unit in the survey.}
\]

\[
c_2 = \text{The cost of enumerating the } j^{th} \text{ character per element in the secondary sample units.}
\]

\[
c_3 = \text{The cost for finding the } u^{th} \text{ character per element in the tertiary sample units}
\]

\[
C = \text{the total cost of enumerating all the } p \text{ characters per TSu.}
\]

Suppose that it is required to find the values of n, m and k so that the cost C is minimized, subject to the upper limits on the variances. If N, M and K are large, then from (1), the limits on the variances may be expressed as:

\[
\frac{S_1^2}{n} + \frac{S_2^2}{nm} + \frac{S_3^2}{nmk} \leq v_j \quad j=1,\ldots,p
\]

Where, \( v_j \) are the upper limits on the variances of various characters. Here \( S_{by}^2 \) is the variance among
primary stage units means, $s_{wji}^2$ is the variance of secondary stage units means, and $s_{wji}^2$ is the variance of tertiary stage units means for $j$ characteristic respectively. The problem therefore reduces to find the optimal values of sample sizes $n, m$ and $k$ which are expressed as:

$$\begin{align*}
\text{Min} & \quad C = C_1n + C_2 nm + C_3 nk \\
\text{Subject to} & \quad \frac{s_{wji}^2}{n} + \frac{s_{wji}^2}{nm} + \frac{s_{wji}^2}{nmk} \\
& \quad n \geq 1, \quad m \geq 1 \text{ and } k \geq 1
\end{align*}$$

(4)

3. FORMULATION OF THE PROBLEM BY USING GEOMETRIC PROGRAMMING APPROACH

Posynomial functions are minimized in the Geometric programming (GP) technique subject to several constraints. Posynomial functions can be defined as polynomials in several variables with positive coefficients in all terms and the power to which the variables are raised can be any real numbers. The cost function and the variance constraint functions are in the form of posynomials. Geometric programming always transforms the primal problem of minimizing a “posynomial” subject to “posynomial” constraints to a dual problem of maximizing a function of the weights on each constraint. Generally constraints are less than strata, so the transformation simplifies the procedure.

The mathematical form of problem (4) with the help of information given below can be expressed in the following way in equation (5) as:

$$\begin{align*}
\text{Find the vector } & \quad X = (x_1 + x_2 + x_3) \\
& \quad (x_1 = n, x_2 = nm \text{ and } x_3 = nmk) \text{ and } \\
& \quad S_{wq}^2 = a_{wq}, \quad S_{wq}^2 = a_{wq}, \quad S_{wq}^2 = a_{wq} \text{ for } q = 1, ..., p
\end{align*}$$

Min $C(X) = \sum_{i=1}^{3} C_i x_i = C_1n + C_2 nm + C_3 nk$ (i)

Subject to

$$\begin{align*}
g(X) = \sum_{i=1}^{3} \frac{a_{wq}}{x_q} \leq v_q, \quad q = 1, ..., p \quad (ii)
\end{align*}$$

and

$$x_i \geq 0, \quad i = 1, 2, 3 \quad (iii)$$

(5)

In the above equations we have noticed that the objective function (5i) is linear and the constraints (5ii) are nonlinear and the reduced two subscripts which in the standard GP (Primal) problem can be stated as:

$$\begin{align*}
\text{Min} & \quad f_q(x)
\end{align*}$$

Subject to

$$\begin{align*}
f_q(x) & \leq 1, \quad q = 1, ..., p \quad (6)
\end{align*}$$

The posynomial $q$ is given as:

$$f_q(x) = \sum_{i=q}^{p} d_i [\prod_{j=1}^{n} x_j^p], \quad d_i > 0, \quad x_j > 0, \quad q = 0, 1, ..., p$$

(7)

The number of posynomial terms in the function can be denoted by $k$, the number of variables is denoted by $n$ and the exponents $p\_q$ are real constants. The objective function $C(X)$ and the constraint function $g(X)$ for our allocation problem that is given respectively in equation (5i) and (5ii) have $k = 3 \quad n = 3$

$$p_1 = p_{22} = p_{33} = 1, \quad p_{12} = p_{21} = p_{13} = p_{31} = 0$$

and $d_i = e_i, \quad i = 1, 2, 3$

and the $q$th constraint has $k = 3, n = 3$

$$p_1 = p_{22} = p_{33} = -1, \quad p_{12} = p_{21} = p_{13} = p_{31} = 0$$

and $d_i = \frac{a_{wq}}{v_q}, \quad i = 1, 2, 3$

(see Maqbool and Pirzada [8]).

The dual form of Geometric Programming problem which is stated in (6) can be given as:

$$\begin{align*}
\text{Max} & \quad \prod_{q=0}^{p} \prod_{i\in q} \left( \frac{d_i}{w_i} \right)^{w_i} \prod_{i=1}^{n} \left( \sum_{i\in q} w_i \right)^{\sum_{i\in q} w_i} \\
\text{Subject to} & \quad \sum_{i\in\{0\}} w_i = 1 \quad (ii)
\end{align*}$$

$$\begin{align*}
\sum_{q=0}^{p} \sum_{i\in q} p_{ij} w_i = 0 \quad (iii)
\end{align*}$$

$$\begin{align*}
w_i \geq 0, \quad q = 0, ..., p \quad (iv)
\end{align*}$$

$$i = 1, ..., k_p$$

Where, $w\_i$’s are weights.
Following Duffin et al. [2] and Woolsey and Swanson [4], the allocation problem 5(i) & 5(ii) will be solved in four steps as follows:

Step 1: For the optimum value of the objective function, the objective function always takes the form:

\[
C_0(x) = \left( \frac{\text{Coeffi. of first term}}{w_1} \right)^{w_1} \times \left( \frac{\text{Coeffi. of Second term}}{w_2} \right)^{w_2} \times \ldots \times \left( \frac{\text{Coeffi. of last term}}{w_k} \right)^{w_k} \times \sum w's \text{ in the } \text{first constraints} \sum w's \text{ in the last constraints}
\]

The objective function (i.e. Cost) for our problem is:

\[
C = \left( \frac{C_1}{w_1} \right)^{w_1} \left( \frac{C_2}{w_2} \right)^{w_2} \left( \frac{C_3}{w_3} \right)^{w_3} (k_1)^{w_4} (k_2)^{w_5} (k_3)^{w_6}
\]

From equation 5(ii), we have

\[
k_i = \frac{a_1}{v_1}, k_2 = \frac{a_2}{v_2}, k_3 = \frac{a_3}{v_3}
\]

Where \( a_1, a_2 \) and \( a_3 \) are the constants of three terms in the \( i^{th} \) constraints, \( v_1, v_2 \) and \( v_3 \) are normalizing variables and \( k_1, k_2 \) and \( k_3 \) are normalized constants in the \( i^{th} \) constraints.

Step 2: The equations that can be used for geometric program for the weights are given below:

\[
\sum w's \text{ in the objective function} = 1 \quad \text{(10)}
\]

and for each primal variable \( x_j \) given \( n \) variables and \( k \) terms

\[
\sum_{i=1}^{n} (w_i \text{ for each term}) \times (\text{exponent on } x_j \text{ in that term}) = 0 \quad \text{(11)}
\]

In our case:

\[
w_1 + w_2 + w_3 = 1 \quad \text{(Normalizing condition, from equation 8(ii))}
\]

\[
(1)w_1 + (0)w_2 + (0)w_3 + (-1)w_4 + (0)w_5 + (0)w_6 = 0
\]

\[
(0)w_1 + (1)w_2 + (0)w_3 + (0)w_4 + (-1)w_5 + (0)w_6 = 0
\]

Orthogonality conditions are represented in equation (13). Combinedly, these conditions are referred to as dual constraints. For more details see Duffin et al. [4]. Now after solving equation (13) we get:

\[
w_1 - w_4 = 0 \Rightarrow w_1 = w_4
\]

\[
w_2 - w_5 = 0 \Rightarrow w_2 = w_5
\]

\[
w_3 - w_6 = 0 \Rightarrow w_3 = w_6
\]

Step 3: The terms which are used in the constraints to the optimal solution are always proportional to their weights. This can be expressed as:

\[
\begin{align*}
\begin{vmatrix}
C_1 & C_2 & C_3 \\
\frac{w_1}{w_1} & \frac{w_2}{w_2} & \frac{w_3}{w_3} \\
\end{vmatrix} & = k_1 \left( \frac{w_4}{w_1} \right) \left( \frac{w_5}{w_1} \right) \left( \frac{w_6}{w_1} \right) \\
\begin{vmatrix}
C_1 & C_2 & C_3 \\
\frac{w_1}{w_1} & \frac{w_2}{w_2} & \frac{w_3}{w_3} \\
\end{vmatrix} & = k_2 \left( \frac{w_4}{w_2} \right) \left( \frac{w_5}{w_2} \right) \left( \frac{w_6}{w_2} \right) \\
\begin{vmatrix}
C_1 & C_2 & C_3 \\
\frac{w_1}{w_1} & \frac{w_2}{w_2} & \frac{w_3}{w_3} \\
\end{vmatrix} & = k_3 \left( \frac{w_4}{w_3} \right) \left( \frac{w_5}{w_3} \right) \left( \frac{w_6}{w_3} \right)
\end{align*}
\]

Step 4: The primal variables can be obtained as:

\[
C_0(x^*) = \frac{\text{first term in the objective function}}{w_1}, \frac{\text{second term in the objective function}}{w_2}, \ldots, \frac{\text{last term in the objective function}}{w_k}
\]

In this case:

\[
\begin{align*}
\begin{vmatrix}
C_1 x_1 & C_2 x_2 & C_3 x_3 \\
\frac{w_1}{w_1} & \frac{w_2}{w_2} & \frac{w_3}{w_3} \\
\end{vmatrix} & = \frac{w_1}{x_1}, \frac{w_2}{x_2}, \frac{w_3}{x_3} \\
\begin{vmatrix}
C_1 x_1 & C_2 x_2 & C_3 x_3 \\
\frac{w_1}{w_1} & \frac{w_2}{w_2} & \frac{w_3}{w_3} \\
\end{vmatrix} & = \frac{w_1}{x_1}, \frac{w_2}{x_2}, \frac{w_3}{x_3}, \frac{k_1}{x_1}, \frac{k_2}{x_2}, \frac{k_3}{x_3}
\end{align*}
\]

Now from equation 15 (ii), the normalization condition is solved using above values of \( w_1, w_2 \) and \( w_3 \) in
equation (12), for obtaining the values of the variables, we have

\[ x_1 = \frac{k_1 + k_2 + k_3}{x_1 + x_2 + x_3} = 1 \]

\[ x_2 = \frac{k_2}{x_2} = 1 - \frac{k_1 - k_3}{x_1 + x_3} \]

\[ x_2 = \frac{k_2 x_2 x_3}{x_1 x_3 - k_1 x_3 - k_3 x_1} \quad (16) \]

\[ \frac{c_1 x_1}{w_1} = \frac{c_2 x_2}{w_2} \Rightarrow x_1 = x_2 \frac{c_2 k_1}{c_1 k_2} \quad (17) \]

\[ \text{Let } A = \sqrt{\frac{k_1 c_2}{c_1 k_2}} \quad (*) \]

\[ \frac{c_1 x_1}{w_1} = \frac{c_3 x_2}{w_3} \Rightarrow x_1 = x_3 \frac{c_4 k_1}{c_3 k_3} \]

\[ x_1 = B x_3 \quad (18) \]

\[ \text{let } B = \sqrt{\frac{c_3 k_3}{c_1 k_3}} \quad (**) \]

Then by putting the values of \( x_2 \) from equation (16) in equation (17), we have

\[ x_1 = \frac{k_2 A + k_1}{x_2 - k_3} x_3 \quad (19) \]

Now the value of \( x_2 \) is obtained from the above equation (16).

\[ x_2 = \frac{k_2 x_2 x_3}{x_1 x_3 - k_1 x_3 - k_3 x_1} \quad (20) \]

4. NUMERICAL ILLUSTRATION

For the illustration of the potential use of the proposed geometric programming procedure, we have considered the following hypothetical data:

\[ S_{b_1}^2 = 0.4560, \quad S_{w_1}^2 = 0.8878, \quad S_{k_1}^2 = 0.9040 \]

\[ S_{b_2}^2 = 0.5234, \quad S_{w_2}^2 = 0.4410, \quad S_{k_2}^2 = 0.5503 \]

\[ S_{b_3}^2 = 0.4085, \quad S_{w_3}^2 = 0.1128, \quad S_{k_3}^2 = 0.2013 \]

\[ v_1 = 0.03110, \quad v_2 = 0.02820, \quad v_3 = 0.02013 \]

\[ C_1 = 10, \quad C_2 = 3, \quad C_3 = 1.5 \]

\[ \text{Min } = C_1 n + C_2 nm + C_3 nmk \]

Subject to

\[ S_{b_1}^2 \frac{x_1}{x_1} + S_{w_1}^2 \frac{x_2}{x_2} + S_{k_1}^2 \frac{x_3}{x_3} \leq v_1 \]

\[ S_{b_2}^2 \frac{x_1}{x_1} + S_{w_2}^2 \frac{x_2}{x_2} + S_{k_2}^2 \frac{x_3}{x_3} \leq v_2 \]

\[ S_{b_3}^2 \frac{x_1}{x_1} + S_{w_3}^2 \frac{x_2}{x_2} + S_{k_3}^2 \frac{x_3}{x_3} \leq v_3 \]

\[ x_1 \geq 1, \quad x_2 \geq 1, \quad x_3 \geq 1 \quad (22) \]

Now by using the above values in equation (22) we get:

\[ \text{Min } C = 10 x_1 + 3 x_2 + 1.5 x_3 \quad (i) \]

Subject to

\[ \frac{0.4560}{x_1} + \frac{0.8878}{x_2} + \frac{0.9040}{x_3} \leq 0.03110 \quad (ii) \]

\[ \frac{0.5234}{x_1} + \frac{0.4410}{x_2} + \frac{0.5503}{x_3} \leq 0.02820 \quad (iii) \]

\[ \frac{0.4085}{x_1} + \frac{0.1128}{x_2} + \frac{0.2013}{x_3} \leq 0.02110 \quad (iv) \]

\[ x_1, \quad x_2, \quad x_3 \geq 0 \quad (v) \]

The normalized constraints for our problem are:
The above equation will give the following:

\[ \frac{0.4560}{X_1} + \frac{0.8878}{X_2} + \frac{0.9040}{X_3} \leq 1 \quad (i) \]

\[ \frac{0.0311}{0.03110} + \frac{0.0311}{0.03110} + \frac{0.0311}{0.03110} \leq 1 \quad (ii) \]

\[ \frac{0.5234}{0.5503} + \frac{0.4410}{0.5503} + \frac{0.5503}{0.5503} \leq 1 \quad (iii) \]

The above equation will give the following:

\[ \frac{14.6624}{X_1} + \frac{28.5466}{X_2} + \frac{29.0675}{X_3} \leq 1 \quad (i) \]

\[ \frac{16.8296}{X_1} + \frac{14.18}{X_2} + \frac{17.6945}{X_3} \leq 1 \quad (ii) \]

\[ \frac{19.3601}{X_1} + \frac{5.3459}{X_2} + \frac{9.5403}{X_3} \leq 1 \quad (iii) \]

The constraint 24 (ii) is assumed to be active (if all the three constraints were active, then two of them will not be able for finding an optimal dual solution nor an optimal solution to the original problem).

(Conditions for active and inactive constraints): At any feasible point \( x \) the \( h^{th} \) constraint is said to be active if \( \delta_h (x) = 0 \) and inactive if \( \delta_h (x) > 0 \). In our case the constraint 24 (ii) is active because it satisfies the condition of active constraint. This can be explained as:

After putting the values of \( x_1^*, x_2^* \) and \( x_3^* \) in the equation 24 (ii) we get:

\[ \frac{16.8296}{31.9740} + \frac{14.18}{53.5847} + \frac{17.6945}{84.6543} \leq 1 \Rightarrow .9998-1=-.0002 \leq 0 \]

Then \( K_1 = 16.8296 \), \( K_2 = 14.18 \) and \( K_3 = 17.6945 \)

On substituting the values of \( K_1, K_2, K_3, C_1, C_2 \) and \( C_3 \) in equations (16), (17), (**), (***), (18), (19) and (20), we get the values of \( x_1, x_2 \) and \( x_3 \) as:

\[ x_1 = 31.9740, \ x_2 = 53.5847 \] and \( x_3 = 84.6543 \).

By rounding the above values we get:

\[ x_1^* = 32, \ x_2^* = 54 \] and \( x_3^* = 85 \)

The optimum values of the sample sizes can be obtained as:

\[ x_1 = n = 31.9740 \cong 32 \]

\[ x_2 = nm = 53.5847 \]

\[ \Rightarrow m = \frac{53.5847}{31.9740} = 1.6759 \cong 2 \]

\[ x_3 = nmk = 84.6543 \]

\[ \Rightarrow k = \frac{84.6543}{53.5847} = 1.5798 \cong 2 \]

After putting the values of \( x_1^*, x_2^* \) and \( x_3^* \) in equation 23(i), we get the total cost as:

\[ C = 10 \times 32 + 3 \times 54 + 1.5 \times 85 = 609.5 \]

The feasibility of the solution is shown with the help of above example. Thus the requirement of sample for primary stage units is 32, the total of secondary stage units in each primary stage units \( nm = 54 \) and the tertiary stage unit within each secondary stage units giving us total of \( nmk = 85 \), elementary units for the sample.

6. CONCLUSION

In this paper we have discussed the optimum allocation in Three-stage sample surveys and provided an effective manual algorithm for solving an optimum allocation in multi-stage sample surveys by using Geometric programming. The algorithm of the solution procedure of Geometric Programming is very simple in comparison to the complex analytical techniques used in statistical literature. There may not be precise knowledge of parameters in the Geometric programming in real worlds due to insufficient information. The feasibility and effectiveness of the present approach has been illustrated by a hypothetical numerical example.

REFERENCES


BIOGRAPHICAL NOTES

• **Shafiullah** received M.Sc. in Operations Research from Aligarh Muslim University, Aligarh, India. He is pursuing Ph.D. from Department of Statistics and Operations Research Aligarh Muslim University, Aligarh, India.

• **Irfan Ali** received M.Sc. and M. Phil. in Statistics from Aligarh Muslim University, Aligarh, India. Presently he is pursuing Ph.D. from Department of Statistics and Operations Research Aligarh Muslim University, Aligarh, India. He has published 22 research papers in national and international journals of repute. He has also presented 6 research articles in national and international conferences.

• **Abdul Bari** received M.Sc. in Mathematics from Meerut University and M.Sc. and Ph.D. in Statistics from Aligarh Muslim University, Aligarh, India. He is a professor in the Department of Statistics and Operations Research Aligarh Muslim University, Aligarh, India. Prof. Abdul Bari has more than 30 years of experience in research and teaching. Prof. Abdul Bari current area of research includes mathematical programming, sample survey and reliability theory. Prof. Abdul Bari has published more than 35 research papers in national and international journals of repute. Prof. Abdul Bari has successfully supervised 6 Ph.D. Students and 9 M. Phil. Students.