ON SOME MATHEMATICAL PROGRAMMING PROBLEMS IN SAMPLE SURVEYS AND RELIABILITY

THESIS

SUBMITTED TO

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

(A CENTRAL UNIVERSITY)

LUCKNOW

ESTABLISHED 1996

FOR THE AWARD OF DEGREE OF

Doctor of Philosophy

IN

APPLIED STATISTICS

Submitted by

Mradula

Under the Supervision of

Dr. Rahul Varshney

Assistant Professor

DEPARTMENT OF APPLIED STATISTICS
SCHOOL FOR PHYSICAL SCIENCES
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
Vidya Vihar, Raebareli Road Lucknow-226 025 (U.P.), India

ENROLMENT NO: 1151/15 **2021**

DECLARATION

I, **Mradula**, Enrolment No. 1151/15, hereby declare that the work which is being presented in the thesis entitled **“On some mathematical programming problems in sample surveys and reliability”** for the award of the degree of Doctor of Philosophy and submitted in the Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow (U.P.), India, is an original work carried out by me under the supervision of Dr. Rahul Varshney, Assistant Professor, Department of Applied Statistics, School for Physical Sciences, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow (U.P.), India.

The matter presented in this thesis has not been submitted by me for the award of any other degree or diploma of this or any other Institute. I also declare that the thesis is essentially free from all kinds of plagiarism.

Date:

Place:

(Mradula)
Research Scholar
Department of Applied Statistics
School for Physical Sciences,
Babasaheb Bhimrao Ambedkar
University,
Lucknow-226 025, India.

CERTIFICATE

This is to certify that the thesis titled “**On some mathematical programming problems in sample surveys and reliability”** submitted by **Mradula**, is an original research work and has not been previously submitted in part or full for the award of any other degree or diploma to this or any other University.

The thesis submitted to Babasaheb Bhimrao Ambedkar University, Lucknow satisfies all the requirements as stipulated in the *Doctor of Philosophy (Ph.D.) regulations -1999 as amended in 2013* and it is fit for submission and evaluation for the award of the degree of Doctor of Philosophy of the University.

Date: Supervisor

Head of the Department

Acknowledgment

Throughout the writing of this thesis, I have received a great deal of support and assistance. I take this opportunity to express my sincere gratitude to all of them.

First and foremost, I would like to express the deepest thank to my supervisor Dr. Rahul Varshney, Assistant Professor, Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, for supporting me during these past four years. He gently corrected me at every stage: coming up with the relevant ideas, implementing them efficiently, presenting them with less ambiguity, and writing concisely. I am deeply obliged to him for the help he extended to me in this academic venture.

I want to express my sincerest gratitude to our Head of the Department, Prof. Madhulika Dube, for her continuous encouragement, guidance, and motivation during my research work.

I am also thankful to Prof. Surinder Kumar, Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, for providing his valuable comments and insightful suggestions on parts of my thesis, which helped improve my research work.

I wish to express my profound gratitude to Dr. Subhash Kumar Yadav, Associate Professor, Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, for his motivation and constructive criticism for the refinement of my work, throughout the development. I want to express my sincere gratitude to Dr. Irfan Ali, Assistant Professor, Department of Statistics & Operations Research, AMU, Aligarh for his timely advice in suggesting the solutions to the problems and continually inspiring me to carry out my research work.

I am also thankful to the members of my Ph.D. committee for their helpful career advice and valuable suggestions in general and the faculty members of the Department Dr.

Amit Kumar Misra and Dr. Meenakshi Mishra for their encouragement, suggestions, and valuable comments. They have always been there to guide and motivate me whenever I encountered any difficulty.

I would like to thanks my seniors, Dr. Abhay Pratap Panday, Dr. Mayank Vaish, Dr. Ajay Kumar and Dr. Sumit Kumar. I would like to express my heartwarming thanks to my research colleagues Mrs. Vaishali Gupta, Ms. Ruby Chanchal and Mr. Ankit for their encouragement and moral support. I also wish to thanks my juniors, Mr. Arun Pal, Mr. Ravi Shankar Sharma, Ms. Taru Bari, Ms. Aditi and Mr. Vishwajeet for their support during my research work. I also appreciate the Department's non-teaching staff, Dr. Amrendra Pratap Bahadur Singh, Mrs. Saumya Trivedi and Mr. Nirmal Singh, colleagues at the university, wishing them every direct or indirect assistance and help during my research work. I would also like to express my sincere gratitude to the University Grants Commission for providing a fellowship during Ph.D.

I would like to acknowledge my all friends for their moral support and motivation, which drives me to give my best and all. I would like to acknowledge the people who mean the world to me, my parents, *Smt. Kamlesh Sharma and Sh. Satya Prakash Sharma*, my sisters *Er. Deepika Shukla, Er. Shailja Sharma* and brother, *Er. Nitin Sharma*, for supporting me emotionally, financially, opportunities and experiences that have made me who I am.

Finally, I want to express my gratitude to my husband, *Sulabh* has supported me throughout this process and constantly encouraged me when the tasks seemed arduous. You greatly inspired me to pursue a new path in life.

Date: **(Mradula)**

DEDICATED
TO
MY BELOVED PARENTS

Mrs. Kamlesh Sharma and Mr. Satya Prakash Sharma

List of published/communicated research papers

1. Varshney, R. and Mradula. (2019). Optimum allocation in multivariate stratified sampling design in the presence of nonresponse with Gamma cost function. *Journal of Statistical Computation and Simulation.* **89(13)**, 2454–2467.
2. Mradula., Yadav, S.K., Varshney, R., and Dube, M. (2019). Efficient estimation of population mean under stratified random sampling with linear cost function. *Communications in Statistics - Simulation and Computation.* DOI: https://doi.org/10.1080/03610918.2019.1643479.
3. Ahmadini, A .A. H., Mradula., Varshney, R., and Ali, I. On multivariate-multiobjective stratified sampling design under probabilistic environment: a fuzzy programming approach. *Journal of King Saud University - Science.* (Accepted)
4. Yadav, S.K., Mradula., and Varshney, R. Efficient estimation of population mean under stratified random sampling using non linear cost function. (Communicated)
5. Mradula., Varshney, R., and Dube, M. Determining optimum strata boundaries using auxiliary variable with nonlinear cost function. (Communicated)
6. Mradula., Varshney. R., and Ali, I. On availability and cost function for maintenance time of continuous operating series-parallel systems using particle swarm optimization. (Communicated)

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Introduction to sampling

In this material world, the information are available in its large volume and frequently obtained according to surveyor's specifications. A sampling method is a procedure of choosing a part of the population in a manner that the selected sample represents the population characteristics, including the variability between population units. The sample is expressed as closely as the size of the sample allows, so that accurate estimates of population characters can be obtained from the sample. For qualitative or quantitative data collection, data on a finite set or subset of units are used in sampling or sample survey. A sample survey is an essential data collection criterion for fulfilling the requirements. Surveys vary in different situations and are handled in various fields such as government sector, industry, physical sciences, economical, etc. With some statistical analysis, they are used as the science and art of controlling and measuring reliability. Sampling are especially helpful in providing data about the target population when

(i) It is required that results present with maximum accuracy or with a fixed budget including a minimum number of units in a fixed degree of reliability.

(ii) The units for the characteristic have significant variance.

(iii) We cannot calculate an accurate count of the population because it would be too expensive and too destructive.

(iv) The investigation is broad in nature and the population is not fully identified.

(v) There are also restricted resources (time, money and other resources).

The purpose of sampling methods is to extract the maximum possible, reliable information about the population under study, save time, increase chances of accuracy. They only deal with large populations and helpful for inaccessible populations and other resources. Any statistical survey aims to obtain information about a population that is a group of units defined according to objective of the survey. The sample surveys recent evolutions have proved that the sampling methodologies provide realistic and reliable results for planned scientific surveys.

1.1.1 Census v/s Sampling

A census is that in which all the population units are studied. Furthermore, data collected through the census is not without errors. They may be experimental or response errors and errors due to nonresponse. It may take more effort, time, and cost to collect information for a large population in census surveys. In the sampling method, we choose a part of the population in a way that contributes a sample representative of the population. A sample must be such that all characteristics of the population, including the variability within population units are followed in the sample as close to the sample size. So, that reliable estimates of the population characters can be formed from the sample. On the other hand, if the population is not finite, resources are limited, then sampling is the only method that gives the necessary information about the population under study. Cochran (1977) defined the principal advantages of sampling over census as “reduced cost”, “greater speed”, “greater scope” and “greater accuracy”.

1.2 Sampling techniques

In sampling theory, various sampling designs are available for various situations and the nature of the population. A primary sampling method is simple random sampling (SRS), which is the easiest method compared to other sampling methods. In this sampling, the sample is drawn unit by unit with an equal probability of selection for each unit at each draw. Sampling is divided into two classes.

(i) Sampling with replacement: In with replacement sampling (WR sampling), all the population units are accessible for choice at each draw since the drawn unit is replaced before the next draw.

(ii) Sampling without replacement: In without replacement sampling (WOR), the unit is once chosen is not considered for encouraging choices.

WOR sample contains more information about the population as compared to a WR sample. Other types of sampling methods can be viewed as modifications of SRS. Some other essential sampling designs are used from time to time for different situations as

(i) Stratified random sampling

(ii) Cluster sampling

(iii) Systematic sampling

(iv) Two Stage sampling

(v) Sub sampling or Multistage sampling etc.

1.2.1 Stratified random sampling

In stratified random sampling, the population size N units with heterogeneity are partitioned into L non-overlapping and specific groups that make internally homogeneous as far as possible. These are called strata, and independent simple random samples of predetermined sizes are drawn from each stratum to estimate the overall population parameters. Stratified schemes are beneficial for survey techniques and are commonly used by government sectors, private sectors, industry, economics, health-related issues, gated buildings, applied statistics or other inaccessible areas. Owing to the economy and its efficacy, stratified random sampling is the most preferred sampling design used for obtaining information from heterogeneous populations.

There are several reasons for stratification. The principal ones are the following:

(i) Gain in the precision of the estimate.

(ii) Administrative convenience.

(iii) Sampling problems may vary in various parts of the population. In these strata, different sampling plans may be used.

(iv) If data of known precision are for specific subdivisions, these subdivisions may be treated as a "population" in their own right.

The practical implementation of stratified random sampling needs the solution to the following three fundamental problems.

(i) Determine the optimum number of strata.

(ii) Determine the optimum strata boundaries.

(iii) Determine if the optimum sample sizes from various strata.

Notations in stratified random sampling

In the stratified random sampling, with the study of the properties of estimates from a stratified sample and some notations are given for h^{th} stratum which is given as

N_h : Stratum size

n_h : Sample size

$f_h = \dfrac{n_h}{N_h}$: Sampling fraction

$W_h = \dfrac{N_h}{N}$: Stratum weight

y_{hi} : Value of the i^{th} unit

$\overline{Y}_h = \dfrac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$: Stratum mean

$\overline{y}_h = \dfrac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$: Sample mean

$S_h^2 = \dfrac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(y_{hi} - \overline{Y}_h\right)^2$: Stratum variance

$s_h^2 = \dfrac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(y_{hi} - \overline{y}_h\right)^2$: Sample variance

$\overline{Y} = \dfrac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi} = \sum_{h=1}^{L} W_h \overline{Y}_h$: Overall population mean

$\overline{y}_{st} = \sum_{h=1}^{L} W_h \overline{y}_h$: Stratified sample mean

$V(\overline{y}_{st}) = \sum_{h=1}^{L}\left(\dfrac{1}{n_h} - \dfrac{1}{N_h}\right) W_h^2 S_h^2$: Sampling variance

$v(\overline{y}_{st}) = \sum_{h=1}^{L}\left(\dfrac{1}{n_h} - \dfrac{1}{N_h}\right) W_h^2 s_h^2$: Unbiased estimate of $V(\overline{y}_{st})$.

This thesis assumes that the population under study is already stratified into L strata. The detailed observe of all these sampling designs are beyond the extent of this thesis. Since stratified random sampling is the maximum broadly utilized sampling layout, in this thesis, the emphasis is on the optimization problems appearing in univariate and multivariate stratified random sampling and their solutions through mathematical programming problem

(MPP) or other advanced techniques.

1.2.2 Allocation problems in stratified random sampling

Determining the sample sizes from numerous strata is known as allocation and several varieties of allocations are available. The generally famous of them are Equal, Proportional, and Optimum allocations. Among these allocations, the optimum allocation is the best.

Equal allocation

The subsample sizes are equivalent in this allocation, for the h^{th} stratum, i.e.,

$$n_h = n/L;\ h = 1, 2, \ldots, L.$$

This approach is used when there is insufficient knowledge about the strata standard deviations and the strata sizes are not given or approximately equivalent.

Proportional allocation

Due to its simplicity, this is the most commonly used allocation. Bowley (1926) had first suggested the proportional allocation. Even if the standard deviations S_h or their approximation s_h are not available, it can be easily used. The allocation of a given total sample size of n to various strata is proportional to the size of the strata, i.e.,

$$n_h \propto N_h \tag{1.2.1}$$

or

$$n_h = KN_h, \tag{1.2.2}$$

where K is the constant of proportionality. Now taking the sum of all relevant strata on both sides of the equation (1.2.2), we have

$$n = KN$$

and value of $K = n/N$. After putting value of K in equation (1.2.1), we have the proportional allocation for h^{th} stratum is given as

$$n_h = nW_h.$$

Optimum allocation

The variance of the calculation for the fixed cost is minimized, or the cost for fixed variance is minimized in optimum allocation. If the cost is linear, from the form $C = c_0 + \sum_{h=1}^{L} c_h n_h$, where c_0 and c_h are the overhead and measurement cost per-unit in the stratum h^{th}, respectively, then the optimum allocation is handled by

$$n_h = \frac{nW_hS_h/\sqrt{c_h}}{\sum_{h=1}^{L} nW_hS_h/\sqrt{c_h}};\ h = 1, 2, \ldots, L.$$

Tschuprow (1923) first gave the optimum allocation for fixed total sample size but the outcome remained unknown. Further, Neyman (1934) rediscovered the result for optimum allocation. Later on, Stuart (1954), was used Cauchy Schwarz inequality to work out optimal allocation under more general conditions. For further details, interested readers may view Cochran (1977) and Sukhatme *et al.* (1984).

1.2.3 Multivariate stratified random sampling design

In multivariate stratified sampling, several characteristics are defined on each population unit. The optimum allocation for one characteristic will not be optimum for other characteristics. Thus, the problem of allocation becomes more complex. The problem of sample allocation in multivariate stratified sampling has drawn the author's attention for a long time, starting apparently with Neyman (1934). In such circumstances, a compromise criterion is required to an optimum allocation for all the characteristics in some sense. Since these are based on a compromise criterion, such allocations are referred to as compromise allocations in multivariate stratified sampling literature.

1.2.4 Allocation problems in multivariate stratified random sampling design: compromise allocations

The problem of allocation is complex in multivariate stratified sampling due to the contradictory existence of the characteristics. An allocation that is optimal for all characteristics in some sense is necessary in such situations to define a compromising criterion. Yates (1960) suggested the idea for evaluation population means when fixed cost criterion be achieved by minimizing the weighted sum of sampling variances of the estimated sampling means.

Cochran (1963) proposed the average of individual optimum allocations for various characteristics. Chatterjee (1967) worked out a compromise allocation by reducing the amount of the proportional increases in variances due to the use of non optimum allocations and assumed the evaluation costs for the different characteristics in a given stratum. Kokan (1963) described the convex programming formulation to the problem of allocation for multivariate stratified sampling. An analytical solution to this convex programming problem was developed by Kokan and Khan (1967). They demonstrated how the sample allocation problem could be presented as a convex programming problem in other designs, such as two-stage sampling, double sampling and response errors. Soland (1967) also dealt with the case of multivariate stratified sampling, where prior knowledge is required on the unknown stratum mean of all variants. He discussed and formulated the stratification problem suggested by Dantzig (1953) as a nonlinear programming problem (NLPP) and then formulated other multivariate stratified sampling problems so that nonlinear programming can solve. Ericson (1970) acknowledged the problem of minimizing the posterior variance of the total population average subject to total budgetary constraint. Also addressed the situation when estimating more than one population characteristics defined on the basis of the strata are sufficiently identical in terms of the various characteristics. Dantzig (1977) gave an integer solution to this problem. Ahsan (1978) described the multivariate allocation problem where prior knowledge on the unknown within stratum means of p characters is available with known parameters in terms of a multivariate normal distribution. Ahsan and Khan (1982) tackled this problem by taking into account the subsequent variances of population means when multipurpose sampling. Sukhatme *et al.* (1984) attained the compromise allocation by minimizing the sum of variances under linear cost constraints for characteristics of p. Khan *et al.* (1997) considered the problem of resolving the optimum compromise allocation as an NLPP and Jahan *et al.* (2001) used the technique of dynamic programming (DP) to work out compromise allocations. Many authors who worked on this issue are Millr *et al.* (2007), Khan *et al.* (2008), Ansari *et al.* (2009), Khan *et al.* (2011), Varshney *et al.* (2012, 2015, 2017), Khowaja *et al.* (2013), Shafiullah. *et al.* (2015), Haseen *et al.* (2016), Raghav *et al.* (2017), Ansari *et al.* (2018) and others, used different compromise criterion to solve such allocation problems for different situations.

1.3 Integer allocation

Integer sample sizes are required for practical purposes. The easiest way to work out with integer sample sizes is to round off the non integer sample sizes to the relative integer values. Generally, the rounded-off values work well, as the variance capabilities are uniform at the optimum stage (see Cochran (1977)), and feasibility also maintained for large sample sizes. In multivariate stratified random sampling, rounded off integer solution may fail to give a nearly optimal result for small sample sizes. They might also violate the cost constraint if the various strata measurement costs are high (see Khan *et al.* (1997)). In such situations, in the formulation, the integer restrictions will also be imposed as additional constraints and make acquiring compromise allocations more complex.

1.4 Problem of nonresponse

If a census is conducted with a large population, several errors occur that can not be supervised adequately. While in an adequately designed sample survey, the estimation of the margin of error is possible to predict, and hence a decision can be made about the accuracy of results. Generally, two types of errors, which occur during sampling, have been recognized in actual practice in the estimators, first one is "sampling errors", and the other is "nonsampling errors". Sampling errors arise due to sampling techniques. The nonsampling errors may occur at any stage of the survey, beginning from the planning until completing the survey report, incomplete returns in mailed questionnaires, inaccurate measurement, recording observations, and respondents. Nonsampling errors are of four types, such as nonresponse errors, measurement errors, tabulation errors and computational errors.

In the case of nonresponse, the estimation of the parameter of interest under traditional sampling techniques results in significant bias, as some units do not respond, thus not covering the target population as the sampling strategy implies. Nonresponse was first discussed by Hansen and Hurwitz (1946) with a mailed questionnaire. The method proposed by Hansen and Hurwitz produces an unbiased estimator of population mean and problem deal with deterministic nonresponce model. In order to improve the response rate, El-badry (1956) generalized the technique of Hansen Hurwitz's by sending large amount of questionnaires to all groups that had not replied. Subsequently, other authors have worked on nonresponse theory as Srinath (1971), Singh and Singh (1979, 1985), and others.

1.5 Optimization

For finding out the best solution under some circumstances, optimization is the best tool for finding them. For designing, constructing and maintaining any framework, designers get to make different technology and administrative choices at a few stages. Optimization can be characterized as finding the conditions that give the most extreme or least value of a function. The first stage of optimization defines decision variables for required benefits, efforts and collected other relevant data. The second stage continues the process by analyzing the mathematical model and selecting the appropriate mathematical programming for finding the best solution and the third stage consists of finding an optimal solution with the computer's help.

The idea of optimization had been developed in the 18^{th} century. Many authors namely, Newton, Lagrange, and Cauchy have worked on optimization problems and found innovative differential calculus techniques for applications in physics and geometry. However, they provided solutions in some basic situations, which are practically acceptable. The problem of optimization solved in recent years appeared to get more complicated. The most remarkable example of the rapid growth of optimization techniques appeared with Bellman's implementation of DP in 1957 and Pontryagin's maximum theory in 1958. The methods solved the problems of optimum dynamic system operation. Many mathematicians have been interested in developing mathematical programming techniques to solve the problem of maximizing or minimizing multiple variables over the last 50 years. Optimization includes a wide variety of cases and implementations. Some common examples of optimization that occur in practice are an economic problem, commercial problem, aerodynamics and science of engineering, etc. Most of the issues in studying production processes and manufacturing facilities can be reduced to optimization.

1.6 Mathematical programming problems

Mathematical programming consists of the maximization or minimization of decision variables called the objective function and restricted by several constraints in equations and inequalities. These problems may exist in different fields, for example, management sciences, numerical analysis, engineering sciences, economics, agriculture, statistical analysis, sample surveys, reliability, etc. The formulation of such problems aims to find the best solution among all possible solutions under limited sources for competing activities. These activities may be expressed as a set of constraints defined by the nature of the problem under study.

The mathematical model of the problem may be represented as

$$\begin{array}{ll} \text{Minimize (or Maximize)} & Z = f(\underline{x}) \\ \text{subject to} & g_i(\underline{x})\ [\leq, =, \geq]\, b_i\ ;\ i = 1, 2, \ldots, m \\ \text{and} & \underline{x} \geq \underline{0}, \end{array}$$

where one of the signs $[\leq, =, \geq]$ holds for each i. $\underline{x}' = (x_1, x_2, \ldots, x_n)$ is the vector of decision variables and b_i; $i = 1, 2, \ldots, m$ are the right hand side values. $f(\underline{x})$ and $g_i(\underline{x})$, $(i = 1, 2, \ldots, m)$ are functions of $x_1, x_2, \ldots, x_n$ and restrictions in $\underline{x} \geq \underline{0}$ are called non-negativity restrictions. Since the maximization of $f(x)$ is equivalent to minimization of $[-f(x)]$, hence the MPP may be taken as either a maximization problem or minimization problem. The MPPs may be categorized into two classes:

- Linear programming problem (LPP) - when all involved functions are linear in an MPP.
- Nonlinear programming problem (NLPP) - when at least one of the involved functions is nonlinear.

An NLPP is more challenging to solve in comparison to LPP for apparent reasons. Like the simplex method for LPP, there is no single method to solve every NLPP. Appropriate techniques are evolved by exploiting some unique characteristics of a distinct NLPP. Depending upon the situation of the involved functions and restrictions on the decision variables, an MPP can be further be placed in one or more of the following categories arranged below:

(i) Integer programming problem

(ii) Multiobjective programming problem

(iii) Chance constrained programming problem

(iv) Stochastic programming problem, etc.

Integer programming problem

In integer optimization problems (to be maximized or minimized), we assume that the selection of variables as non fractional or discrete values and integer problems can be usually be restricted or unrestricted with some constraints. The objective and constraint functions can be linear and nonlinear. An integer problem is determined as linear if, through relaxing the integer restriction on the variables, the resulting functions are strictly linear.

Multiobjective programming problem

Multiobjective programming is a multicriteria decision-making field that has to do with MPPs involving several objective functions to be optimized at once. Multiobjective optimization has been applied in many scientific areas, including engineering, economics, and logistics, where optimal decisions must be made in the face of trade-offs.

Chance constrained programming problem

Chance constrained programming was first proposed by Charnes and Cooper (1963) as a method for solving optimization problems under uncertainty. Some constraints may be deterministic in a stochastic programming problem, and the remainder may include random elements. On the other hand, in a chance programming problem, the set of constraints does not necessarily exist, but it holds at some point with the given probabilities. As such, we are given a set of probability measures that indicate the extent of the random constraints being violated.

Stochastic programming problem

It refers to the problem of mathematical programming, which includes assessing relative risks and uncertainties in various alternatives of choice for management decisions. It deals with situations where stochastic (or random or probabilistic) variables represent some or all of the problem's parameters, rather than deterministic quantities. The values of the random variable can be varied according to its nature and form of a problem. For example, concrete strength has to be taken as a random variable in the design of concrete structures because the compressive strength of concrete varies considerably from sample to sample. Any machined component's actual dimension must be taken as a random variable in mechanical systems design because the dimension may be anywhere within a specified tolerance range. Likewise, in aircraft and rocket design, the actual loads acting on the vehicle depend on the prevailing atmospheric conditions at the launch date, which can not be predicted accurately in advance.

Therefore, the loads in these flight vehicle's configuration must be regarded as random variables. The stochastic optimization problem is also abbreviated as a stochastic linear or nonlinear problem, depending on the nature of the equations involved (in terms of random variable's) in the problem. The basic concept is used to solve the stochastic problem and turn the stochastic problem into an equivalent deterministic problem using standard techniques like linear, geometric, dynamic, nonlinear programming.

Lexicographic goal programming, Fuzzy programming technique, Particle swarm optimization (PSO), Genetic algorithm (GA), Goal programming problem, DP and many more are useful for computational techniques to solve different MPPs forms. Some optimization techniques are discussed below.

Lexicographic goal programming technique

In this technique, the objective functions are arranged according to their importance by the designer. The optimal solution is then achieved by the objective functions beginning with the most important one and continuing order according to the importance of objective functions. Target subscriptions show not only the target function number but also aim priorities.

The essential objective function is reduced after ordering, according to the original constraints. Suppose this problem has a particular solution that is the solution to the entire problem of multiobjective optimization. Otherwise, the second most critical objective feature would be reduced. Now a new constraint is introduced, in addition to the initial constraints. This new constraint ensures that the most relevant objective function conserves its optimum value. If this problem has a unique solution, then the initial problem is solved. Otherwise, the cycle continues, as mentioned above. Fishburn (1974) critically discusses the lexicographic orders, utilities, and decision Rule.

Fuzzy programming technique

Fuzzy programming methodology is used in this study to find an optimal solution for problems of the multivariate sample survey. The definition of fuzzy set theory was first introduced by Zadeh (1965), and Zimmermann (1978) proposed the fuzzy approach to solving multiobjective problems. He showed that efficient solutions are obtained from fuzzy linear programming. Narasimhan (1980) and Ignizio (1982) had explored and improved the use of goal programming in a fuzzy set theory to solve multiobjective problems.

In the real world, problems are generally vague and not well defined. To overcome such problems in real life, fuzzy mathematical programming was developed. The fuzzy concept is used to describes the goals and constraints in MPP and understand the degree of decision-makers expectations.

Particle swarm optimization

The PSO, defined by Kennedy and Eberhart (1995), is a stochastic optimization technique based on a population. PSO's development aimed to emulate cooperative foraging living organisms and social behaviors such as birds, fish, bees, etc. Several striking PSO characteristics make it an excellent evolutionary metaheuristic. In PSO, the population is called a swarm and any variable of the swarm is knows as PSO particles. The particles of the swarm head in the search space have a limited speed, so as to reach the desired goal.

Using their cognitive and collective ability the search particles are able to adjust their best position and the best position so far explored by a swarm, respectively. Kennedy and Eberhart (1995) developed an equation (1.6.1) to modify each particle's velocity and thus modified the individual particle's position according to this modified velocity in equation (1.6.2).

$$\begin{aligned} V_p(t) =& V_p(t-1) + C_1 * rand(.) * (X_p^{best} - X_p(t-1)) \\ & + C_2 * rand(.) * (X_s^{best} - X_p(t-1)) \end{aligned} \tag{1.6.1}$$

$$X_p(t) = X_p(t-1) + V_p(t), \tag{1.6.2}$$

where, $V_p(t) \in [V_{min}, V_{max}]$ represents individual particle velocity $(p = 1, 2, \ldots, N)$ at t^{th} iteration. C_1 and C_2, respectively, denote particle cognitive and collaborative ability, called coefficients of acceleration. Rand $(.)$ is a random value between 0 and 1. X_p represents the particle position p, best of X_p^{best} and best of X_s^{best} are used to denote the best location (up to t^{th} iterations) found so far for an individual particle and the entire swarm (global best), respectively.

Genetic algorithm

The GA is the most widely used to solve optimization problems. This algorithm was initially invented in 1895 from Darwinian evolutionary theory and was first implemented at Michigan University by Holland (1975). According to this theory, organisms that are more environmentally adaptable, survive. Information transmitted from each generation to the next generation is contained in chromosomes, thus transmitting inherited properties.

In this algorithm, the better population is combined according to the principle of survival of the best, and this solution will be repeated more often in the next generation depending on the suitability of each solution. This process will continue to reach an optimal solution.

1.6.1 Mathematical programming problems in sample survey

In many decision-making problems, sample surveys, environmental, social, economic, and technical areas have more than one multiple objective functions. It is important to mention that various goals are always unmeasurable and in conflict with each other in optimization problems. The problem of obtaining statistical information on population characteristics can be formulated as an optimization problem of minimizing the cost of the survey subject to the restriction that the loss of accuracy is within the prescribed limit or, in turn, minimizing the loss of accuracy subject to the limitation that the cost of the survey remains within the budget. The compromise allocation was obtained by Ghufran *et al.* (2012) when the cost function is assumed quadratic and random. Also, Varshney *et al.* (2012) worked out compromise allocation by using a quadratic cost function for large sample size in the estimation of more than one parameters in stratified sampling with a fixed budget, and Haseen *et al.* (2012) suggested a fuzzy approach for solving double sampling design in the presence of nonresponse. Raghav *et al.* (2014) has discussed the various multiobjective optimization techniques in multivariate stratified sample surveys in case of nonresponse. Shafiullah. *et al.* (2014) introduced nonresponse using quadratic cost function by fuzzy geometric programming technique. Muhammad *et al.* (2015) formulated the allocation problem as a multiobjective multivariate nonlinear programming problem and worked out compromise allocation for the fixed Gamma cost function. Haseen *et al.* (2016) discussed the problem of nonresponse in multiobjective stochastic multivariate stratified sampling. The fundamental paper by Charnes *et al.* (1955) presented the application of mathematical programming to statistics. Wagner (1959) suggested solving the problem through the dual approach. Other areas of application of mathematical programming in statistics developed simultaneously by several authors.

Generally, one best allocation for one item is not best for other allocations. We use some compromise criterion in a survey with numerous items until there is a strong correlation between items. If useful empirical data are available, we compute the individually optimum allocation and see to what extent there is disagreement. Chatterjee (1967) suggested an alternative compromise allocation to choose the n_h by minimizing the average of the proportional increases in variance, taken over the variables, which is given as

$$n_h = \frac{n\sqrt{\sum\limits_{j=1}^{p} {n'_{jh}}^2}}{\sum\limits_{h=1}^{L}\sqrt{\sum\limits_{j=1}^{p} {n'_{jh}}^2}}.$$

If j denotes a variable, the value of n_h is determined as above, where n'_{jh} is the optimum sample size in stratum h for variable j.

Individual optimum values differ only slightly. In some surveys, the optimum allocations vary so much that the above criterion does not work for individual variates. Some other principle is a requirement to obtain the allocation to be used. Yates (1960) suggested two useful principles. The one is applying to surveys with a specified purpose, in which the loss can be measured in terms of money or utility due to an error of a known value in an estimation. With p variates and quadratic loss functions, it may be appropriate to represent the total expected loss as the weighted sum of the variances of the estimated population means or total. For the means, we have

$$L = \sum_{j=1}^{p} a_j V(\bar{y}_{jst}) = \sum_{j=1}^{p} a_j \sum_{h=1}^{L} W_h^2 S_{jh}^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right),$$

where S_{jh}^2 is the variance of the j^{th} variate in the stratum h $(h = 1,2,\ldots,L)$ and $a_j \geq 0$ is weight assigned to j^{th} variable such that $\sum\limits_{j=1}^{p} a_j = 1$. After interchanging the order of summation, we get

$$L=\sum_{h=1}^{L}\frac{W_h^2}{n_h}\left(\sum_{j=1}^{p}a_jS_{jh}^2\right)-\frac{1}{N}\sum_{h=1}^{L}W_h\left(\sum_{j=1}^{p}a_jS_{jh}^2\right).$$

With a linear function for the costs of sampling, we have

$$C=c_0+\sum_{h=1}^{L}c_hn_h.$$

The value of n_h is obtained by minimizing the product of $(C-c_0)$ and the term in L (the term containing n_h), by using Cauchy-Schwartz inequality,

$$n_h\propto\frac{W_h}{\sqrt{c_h}}\sqrt{\sum_{j=1}^{p}a_jS_{jh}^2}.$$

The proportionality constant is found out by satisfying the given L or C constraint. For particular, suppose that the value of L is defined and that it is possible to ignore the term finite population correction (fpc). We have

$$n_h=n\frac{(W_hA_h/\sqrt{c_h})}{\sum_{h=1}^{L}(W_hA_h/\sqrt{c_h})},$$

where $A_h=\sqrt{\sum_{j=1}^{p}a_jS_{jh}^2}$.

The required total sample size is

$$n=\frac{1}{L}\left(\sum_{h=1}^{L}\frac{W_hA_h}{\sqrt{c_h}}\right)\left(\sum_{h=1}^{L}W_hA_h\sqrt{c_h}\right).$$

In the second principle, the required variance V_j is specified for every variate. For population means, we have

$$\sum_{h=1}^{L}\frac{W_h^2S_{jh}^2}{n_h}-\sum_{h=1}^{L}\frac{W_hS_{jh}^2}{N}\le V_j;\quad j=1,2,\ldots,p.$$

The inequality sign “$\le$” is used to achieve the most economical allocation that may provide variances smaller than the desired $V_j; j=1,2,\ldots,p$.

In this approach the cost C is to be minimized subject to tolerances $V_j; j = 1, 2, \ldots, p$ and $0 \le n_h \le N_h$. Hence the problem becomes NLPP. The solution algorithms were given by Hartley and Hocking (1963), Chatterjee (1966), Zukhovitsky and Avdeyeva (1966), Kokan and Khan (1967), Huddleston *et al.* (1970), Cochran (1977) and Sukhatme *et al.* (1984) worked out compromise allocations in stratified random sampling. For a notion of sampling optimization, Díaz-García and Garay Tapia (2007) discussed in details. Bakhshi *et al.* (2010) found the optimal sample numbers with a probabilistic cost constraint. The allocation problem was formulated by Khan *et al.* (2011) as a multiobjective stochastic convex programming problem and a compromise solution achieved by using the approximation technique of Chebychev. Varshney *et al.* (2011) workout an allocation by minimizing the trace of the variance-covariance matrix of the estimator of the population mean. The randomness of multiobjective multivariate stratified sample surveys was explored by Ali *et al.* (2013a) and the compromise allocation was derived using various multiobjective optimization techniques. Some other authors such as Khan and Ali (2014), Haseen *et al.* (2015), Varshney *et al.* (2017) and Haq *et al.* (2020) also discussed this concept.

1.7 Reliability

From World War II, reliability theory has been broadly developed over the years with a significant defense personnel contribution. According to the Aeronautical (1994), reliability is the probability that a system will perform satisfactorily for at least a given period under stated conditions. The reliability of an item at time t is denoted as

$$R(t) = 1 - F(t) = Pr(T \ge t) \text{ for } t \ge 0.$$

$R(t)$ is the probability that the item does not fail in the time interval $(0, t]$. It has the following characteristics:

(i) R(0) = 1, since the device is assumed operable at $t = 0$.

(ii) $R(\infty) = 0$, since no device can work.

(iii) R(T) is a non increasing function between 0 and 1.

A system's reliability may be optimized subject to the resource constraints to determine the optimum number of redundant elements for each level provided that the reliability of every element is known. There are many strategies available in the literature to solve problems in different situations. A basic algorithm for solving the redundant optimization problem was developed by Misra (1972). Lientz (1974) applied a probabilistic approach for allocating components by using a conditional enumeration to maximize the system reliability subject to cost constraint for the given component reliability value. The same technique described for reliability optimization problems were also used by McLeavey and McLeavey (1976). A detailed study of the literature on the reliability maximization of redundant systems reveals that many researchers have illustrated the system reliability optimization problems obtained using different techniques with an essential significant cost constraint under consideration.

Availability

Availability is a success criteria for devices that can be repaired. This technology combines with both the reliability and maintainability aspects of a system. Availability is the possibility of a system not being out of operation or not in need of repair when it is needed. The range of availability lies between 0 to 1. In general, when the device is not available, then availability as equally to reliability, because of the increased costs and inconvenient. There are two significant differences in reliability and availability, which are as follows:

(i) Reliability is an interval function, whereas availability is an point function at a given time.

(ii) The reliability function prevents a system's failure within the specified time period, but the availability function does not depend on such constraints.

1.7.1 Mathematical programming problems in reliability

Reliability is used to assess system structures and significant design measures in many industrial environments, such as telecommunication and manufacturing facilities. The design of such hardware systems may be formulated as a reliability optimization problem. Such problems aim to either maximize reliability, availability, and performance or minimize the cost. In a real-life system, data is uncontrollable and unavoidable in some situations and always not give a precise result. Therefore, to handle the inaccurate data, the MPPs may be used to formulate the optimization problems and the appropriate decision may be taken by using the proper mathematical programming technique.

The problems of the reliability optimization in various system designs has been reported by Tillman *et al.* (1980), Sasaki and Shingai (1983), Chern and Jan (1986), Misra (1986), Sung and Cho (2000), Kuo *et al.* (2001), etc. Prasad and Kuo (2000) focused on reliability design using three principles: system configuration, problem category, and optimization techniques. Out of these three criteria of research classification, system configuration, and optimization approaches are still evolving research areas in optimal reliability design of sophisticated systems. Some authors also utilized the fuzzy technique in reliability optimization problems, such as Huang (1997) formulated the problem with multiple objectives of series system reliability by using a fuzzy concept. Mahapatra and Roy (2006) worked on the reliability of series and complex systems under the cost constraint with generalized triangular fuzzy numbers, and Mahapatra and Roy (2009) discussed the reliability for a series and parallel system using the triangular intuitionistic fuzzy concept. Mahapatra and Roy (2011) analyzed the reliability and cost model for series system with a fuzzy parametric geometric approach.

Ben-Daya *et al.* (2000) and Osaki (2002) obtained different models of reliability and maintenance and explored their best solutions, and applied them to various systems. Reliability has expanded its scope across other sectors in recent years, thus functioning as an integral element of quality within its structure and development cycle. To preserve the reliability of complex systems at a higher level, the structural design of the system or system components demands greater reliability, or both are simulated. Many authors have recently focused on the theory of reliability, such as Moghaddam *et al.* (2008), Ali *et al.* (2013b), Gruber *et al.* (2013), Zhou *et al.* (2015), Alam (2016), Gupta *et al.* (2016), Kumar and Garg (2016), Loganathan and Gandh (2016), Ardakan *et al.* (2017), Fallahnezhad and Najafian (2017), Ghnimi *et al.* (2017), Kim and Kim (2017), Kakkar *et al.* (2019), Mousavi *et al.* (2019), Zhiyuan *et al.* (2019), and Li *et al.* (2020).

In Reliability, the MPP is to maximize the system reliability may be formulated as

$$\begin{array}{lll} \text{Maximize} & R = R_{sys} & \\ \text{subject to} & h_{ij}(\underline{x}) \le B; & i = 1,2,\ldots,N,\ j = 1,2,\ldots,M_i \\ & \sum\limits_{j=1}^{M_i} x_{ij} \ge 1; & \forall\, i = 1,2,\ldots,N \\ \text{and} & x_{ij} \ge 0; & \forall\, i = 1,2,\ldots,N,\ j = 1,2,\ldots,M_i, \end{array} \tag{1.7.1}$$

where N is the number of subsystems, M_i is the number of design alternative in subsystem i, and x_{ij} is the number of components of design alternative j in stage i. R and R_{sys} represent the objective function and overall reliability of the system, respectively. We consider that B is total available amount of budget and $h_{ij}(\underline{x})$ is a functions of $\underline{x}$. $\sum_{j=1}^{M_i} x_{ij}$ represents the multiple choice constraint and x_{ij} defines the decision variables. The objective function of the problem is to find all decision variable's values for which the objective function under the given constraints may be maximized.

The problem of the system reliability has been defined by Taghizadeh and Hafezi (2012), Yusuf (2014), Rao and Naikan (2014) and several other authors. A multiobjective redundancy allocation problem has been determine by Khalili-Damghani *et al.* (2014) adopting a decision support system and multiple constraints under fuzziness. Reliability optimization problems have been investigated by some other authors such as Gupta *et al.* (2016), Kumar and Garg (2016), Salah and Temraz (2019), Ivezic *et al.* (2019), and Mellal *et al.* (2020) etc.

The real-life situations, such as projects and case studies, may be formulated as MPPs, and a solution may be worked out by using the appropriate optimization technique.

1.8 Some statistical distributions

Several distributions have been proposed for the use of lifetime analysis. We have considered some distributions in the present work, namely exponential, Gamma, Weibull, and normal distributions. A brief description of the distributions used in this thesis is given below. We start with an expository description of exponential distribution, which enjoys the most widely used model in lifetime analysis and reliability.

1.8.1 Exponential distribution

The exponential distribution is the most commonly used continuous probability distribution in modeling reliability and life testing analysis. In several various environments, exponential distribution exists, as with radioactive or particle decays or the time among events in a Poisson process where events are constantly appearing.

The exponential distribution with parameter λ, the probability density function (pdf) is given as

$$f(x) \quad = \lambda \exp(-\lambda x), \quad \lambda \geq 0. \tag{1.8.1}$$

The distribution is supported for variable x on the interval $[0, \infty)$.

1.8.2 Gamma distribution

Gamma distribution is the most popular distribution for analyzing lifetime data. These distributions possess both scale and shape parameters and can be quite flexible in analyzing any positive real data. The pdf of Gamma distribution is given by

$$f(x;\lambda,k)=\frac{\lambda(\lambda x)^{(k-1)}\exp(-\lambda x)}{\Gamma(k)},\quad \lambda\geq 0,\ k\geq 0,$$

where λ and k are scale and shape parameters, respectively and variable x on the interval $[0,\infty)$.

A Gamma distribution arises naturally in the processes for which the waiting times between the Poisson distributed events are relevant and a general form of statistical distribution related to beta distribution. Gamma distribution is taken from the Poisson assumptions, the number of events happens in a time interval of t, at a rate of λ is given by

$$P(r)=\frac{(\lambda t)^{t}e^{-\lambda t}}{r!},\quad \lambda\geq 0.$$

Probability that there are at least k events happen at a time t, as defined below

$$F(t)=\sum_{r=k}^{\infty}P(r)=\int_{0}^{t}\frac{\lambda^{k}z^{(k-1)}e^{-\lambda z}}{(k-1)!}dz,$$

where the integral is placed by a summation. The time distribution for the first event's occurrence follows an exponential distribution, and a sum of exponential distribution follows Gamma distribution. In queuing theory, the Gamma distribution for an λ which assumes integer values is known as the Erlang distribution.

1.8.3 Weibull distribution

Weibull distribution is a positive real-valued continuous probability distribution that describes in probability theory and statistics area. It is named given by Swedish mathematician Waloddi Weibull and given detail in the article wide applicability (1951). Whereas, it was first developed by Fréchet (1928) and utilized by Rosin and Rammler (1933) to describe a particle size distribution.

The pdf of the Weibull distribution with parameters r, θ and γ is given by

$$f(x;r,\theta,\gamma) \quad = \frac{r(x-\gamma)^{(r-1)}}{\theta^r}\exp\left(-\left(\frac{x-\gamma}{\theta}\right)^r\right), \quad x>\gamma,\ r>0,\ \theta\geq 0, \tag{1.8.2}$$

where r and θ are shape and scale parameters, respectively and γ is a location parameter. The Weibull distribution is a suitable model for numerous problems, and it is an extensively used distribution in reliability and survival analysis due to its various shapes of the hazard rate function. For more details, the interested reader may refer Hallinan (1993), Horst (2008) etc.

1.8.4 Normal distribution

One of the most important continuous probability distributions in the entire statistics field is the normal distribution because it fits many natural phenomena, such as heights, blood pressure, measurement error, and IQ scores. The normal distribution was first appeared in 1733 as an approximation to the probability for sums of binomially distributed quantities to lie between two values when Abraham de Moivre communicated it to some of his contemporaries. A search by Daw and Pearson (1972) confirmed that several copies of this note had been bound up with library copies of De Moivre's miscellanea analytica, printed in 1733 or later. The normal distribution has been named after various scientists, including Laplace and Gauss. Kac (1975) recalled that it has also been named after Quetelet and Maxwell.

The normal distribution is often referred to as the Gaussian distribution in honor of Karl Friedrich Gauss, who derived the equation of a normal pdf. As with any probability distribution, the normal distribution parameters define its shape and probabilities entirely and have two parameters, the mean and standard deviation, and do not have just one form. However, the shape changes based on the parameter values.

The pdf of a normal random variable X is given by

$$f\left(x;\, \mu,\, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad \mu \in \mathbb{R},\, \sigma^2 > 0,$$

where μ and σ^2 are location and scale parameters, respectively and x lies between the real numbers. The constants μ, σ, and σ^2 are respectively, the mean, standard deviation, and variance of the normal distribution.

1.9 Problems discussed in subsequent chapters

The summary of the thesis presented in six chapters and followed in the sequel outlining the contents chapter by chapter.

In **Chapter 1**, provide a primary reading material to equip the reader for understanding the problems discussed in the following chapters. Although the topics considered have not been discussed in detail, it is attempted to provide sufficient references so that interested readers can look into the references to have complete knowledge of the topics. Finally, we provided the motivation and summary of the present work.

The **Chapter 2**, deals with the problem of nonresponse theory, which Hansen and Hurwitz (1946) initially introduce. The nonresponse problem is solved using the Gamma cost function and comparison by existing compromise allocations with the proposed method. The numerical illustration by lexicographic goal programming technique for the practical utility of the method. This problem is based on my published research work Varshney and Mradula (2019) and published in the *Journal of Statistical Computation and Simulation.*

In **Chapter 3**, we study a procedure for determining the optimum stratum boundaries and sample size with a nonlinear fixed cost function for each stratum. In health surveys, these auxiliary variables normally characterize positively skewed distributions from the Gaussian distribution families such as Weibull, Gamma, Log-normal, etc. Thus, this research investigates if the proposed parametric based mathematical programming approach for determining the optimum stratum boundaries yields a gain in efficiency over other methods. This research also tries to determine if the proposed method works for skewed distributions such as the exponential and Weibull distribution when both real and simulated data sets are used in the mathematical programming problem formulation of the stratification problem.

The **Chapter 4**, deals with stochastic programming under some probabilistic conditions and multivariate stratified sampling as a stochastic multiobjective nonlinear programming problem by using of a nonlinear cost function which is prudent by introducing the labour cost as part of the total cost of a survey. The stochastic nonlinear programming problem is solved using the chance constraint and modified E-model methods, and a solution is found out by fuzzy goal programming problem, and Lagrange's methods. Here also, a comparison study is made. This problem is based on my accepted research work in the *Journal of King Saud University - Science.*

In **Chapter 5**, we study the ratio type estimators where the population mean is unknown, but the population median of study variable is known, and use of auxiliary information for improving the precision of estimates of the population mean or total. The problem is formulated in two parts,
(i) with linear cost function and solved using the integer programming technique and Lagrange multiplier technique and numerical illustration for the method's practical utility. This part is based on my published research work Mradula *et al.* (2019) and published in the *Journal of Communications in Statistics - Simulation and Computation.*
(ii) With nonlinear cost function and solved it using integer programming problem with

MATLAB software.

A comparative study of different other estimators has been made.

In **Chapter 6**, the availability and cost function of a series-parallel system containing a multiobjective model under a fixed time environment. The evolution of the solution methods has been discussed with an illustration using particle swarm optimization. The proposed method then compared with different existing optimization techniques such as genetic algorithm and goal programming technique and results show the effectiveness of the proposed approach in terms of availability and a profound reduction in the cost maintenance with system time. We use simulation studies by using R software to evaluate the availability and determine the cost function by MATLAB.

CHAPTER 2

Compromise allocation in multivariate stratified sampling design in the presence of nonresponse with Gamma cost function

2.1 Introduction

To obtain maximum information from different characteristics defined on each unit of the population, the homogeneity within population units will not remain there. Stratified sampling is considered as the best suited sampling design in comparison to other available designs. The optimum allocation determines the size of the sample to be taken from each stratum of the population. It is determined by minimizing either the variance of the estimate for a fixed cost of the survey or the cost of the survey for fixed variance of the estimate. Tschuprow (1923) and Neyman (1934) derived the formula for optimum allocation to estimate population mean for univariate case. Cochran (1977) and Sukhatme *et al.* (1984) discussed optimum allocation in stratified random sampling. To obtain maximum information from different characteristics defined on each unit of the population, the stratified sample survey converts into its multivariate case and termed as multivariate stratified sample survey in sampling literature. In multivariate stratified sample survey, an optimum allocation of any characteristic may not be remained optimum for other characteristics. Kokan and Khan (1967) formulated the problem of allocation for multivariate case by minimizing the cost of enumerating all the characteristics, defined on each unit, of a survey for desired precisions on various characters in a convex programming problem.

Usually, in sample surveys, when more than one characteristic are defined, optimum allocation of one characteristic may not be optimum for others. In such situations, a compromise criterion is needed to get an allocation, which may be near-optimum to other characteristics in some sense, for all characteristics, and such allocation is termed as “compromise allocation” in sampling literature. Many authors Dalenius (1957), Yates (1960), Chatterjee (1968), Khan *et al.* (1997), Shafiullah. *et al.* (2014), Ali *et al.* (2015), Varshney *et al.* (2015), Haseen *et al.* (2016), Raghav *et al.* (2017), Varshney *et al.* (2017), Ansari *et al.* (2018) and others, worked out for by defining new estimators or used existing estimators in different situations for multivariate case.

In multivariate stratified sample surveys, when the required information is not obtained from selected units of the population, the situation becomes nonresponse. The extent of nonresponse depends on various factors such as the type of the target population, time of the survey, and type of the survey. To solve the nonresponse problem in stratified sampling, each stratum is further divided into two mutually exclusive and exhaustive groups, i.e., one of the respondents and other nonrespondents. Hansen and Hurwitz (1946) constructed the estimator for the population mean by considering these two groups. They derived the expression of variance and worked out the optimum sampling fraction for the nonrespondents. For fixed linear cost function, Varshney *et al.* (2011, 2012), formulated the problem of nonresponse and worked out compromise allocation for a multivariate stratified sample survey. In this chapter, the problem of obtaining compromise allocation is discussed for multivariate stratified sample survey and fixed nonlinear cost function in the presence of nonresponse. If labour time is considered in the given cost function, the cost function may not remain linear. The lexicographic goal programming technique is used to solve the problem of allocation to determine sample sizes of respondent and nonrespondent groups. The numerical examples are included to illustrate the practical utility of the method.

In this chapter, the compromise allocations are worked out for multivariate stratified populations in nonresponse under the Gamma cost function. The contents of this chapter is based on my published research work Varshney and Mradula (2019).

2.2 Formulation of the problem

In multivariate stratified sample survey it is assumed that the population is divided into L strata to maintain homogeneity within strata and p characteristics are defined on each population unit. For the h^{th} stratum, N_h, $W_h = N_h/N$, $\overline{Y}_h$ and S_h^2; where $h = 1,2,\ldots,L$ denote the stratum size, stratum weight, stratum mean and stratum variance, respectively. Furthermore, h^{th} stratum is divided into two mutually exclusive and exhaustive groups; one for respondents of size N_{h1} and other group for nonrespondents of size $N_{h2} = N - N_{h1}$. Obviously the true values of N_{h1} and N_{h2} and/or their estimates are not known prior to the sample observations.

For the given sample of size n such that $n = \sum_{h=1}^{L} n_h$, let n_h denotes the size of sample which is selected randomly from h^{th} stratum of the population. Also, out of n_h units, n_{h1} units belong to the group of respondents and the remaining $n_{h2} = n_h - n_{h1}$ units belong to the group of nonrespondents. To overcome the problem of nonresponse, a second attempt was made and subsamples of sizes

$$r_h = n_{h2}/k_h; \quad h = 1,2,\ldots,L, \tag{2.2.1}$$

are drawn from nonrespondents groups, where $k_h \geq 1$ and $1/k_h$ represents the sampling fraction among nonrespondents in the h^{th} stratum. The unbiased estimates of N_{h1} and N_{h2}

may be taken as $\hat{N}_{h1} = n_{h1}N_h/n_h$ and $\hat{N}_{h2} = n_{h2}N_h/n_h$, respectively.

Using Hansen-Hurwitz technique, an estimator of stratum mean $\overline{Y}_{jh}$ for j^{th} characteristic in the h^{th} stratum is given by

$$\overline{y}_{jh(w)} = \frac{n_{h1}\overline{y}_{jh1} + n_{h2}\overline{y}_{jh2(r_h)}}{n_h}, \tag{2.2.2}$$

where $\overline{y}_{jh1}$ and $\overline{y}_{jh2(r_h)}$; $j = 1,2,\ldots,p$ denote the sample means based on n_{h1} units of respondents and r_h units of subsample of nonrespondents group respectively for the h^{th} stratum.

It can be seen that $\overline{y}_{jh(w)}$ is an unbiased estimate of the stratum mean $\overline{Y}_{jh}$ of the h^{th} stratum for the j^{th} characteristic with a variance

$$V(\overline{y}_{jh(w)}) = \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S^2_{jh} + \frac{W^2_{h2}S^2_{jh2}}{r_h} - \frac{W_{h2}S^2_{jh2}}{n_h}, \tag{2.2.3}$$

where S^2_{jh} is the stratum variance of j^{th} characteristic in the h^{th} stratum; $j = 1,2,\ldots,p$ and $h = 1,2,\ldots,L$ given as

$$S^2_{jh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{jhi} - \overline{Y}_{jh})^2,$$

where y_{jhi} denotes the value of the i^{th} unit of the h^{th} stratum for j^{th} characteristic and $\overline{Y}_{jh} = \sum_{i=1}^{N_h} y_{jhi}/N_h$ is the stratum mean of y_{jhi}.
S^2_{jh2} is the stratum variance of j^{th} characteristic in the h^{th} stratum; $j = 1,2,\ldots,p$ and $h = 1,2,\ldots,L$ given as

$$S^2_{jh2} = \frac{1}{\hat{N}_{h2} - 1}\sum_{i=1}^{\hat{N}_{h2}}(y_{jhi} - \overline{Y}_{jh2})^2,$$

$\overline{Y}_{jh2} = \sum_{i=1}^{\hat{N}_{h2}} y_{jhi}/\hat{N}_{h2}$ is the stratum mean of y_{jhi} and $W_{h2} = N_{h2}/N_h$ is the stratum weight of nonrespondent in h^{th} stratum.

An unbiased estimate of overall population mean $\overline{Y}_j$ of the j^{th} characteristic is given by

$$\bar{y}_{j(w)} = \sum_{h=1}^{L} W_h \bar{y}_{jh(w)}.$$

The variance of $\bar{y}_{j(w)}$ will be given as

$$V(\bar{y}_{j(w)}) = \sum_{h=1}^{L} W_h^2 V(\bar{y}_{jh(w)}).$$

After ignoring fpc, it is given by

$$V(\bar{y}_{j(w)}) = \sum_{h=1}^{L} \frac{W_h^2 \left(S_{jh}^2 - W_{h2} S_{jh2}^2\right)}{n_h} + \sum_{h=1}^{L} \frac{W_h^2 W_{h2}^2 S_{jh2}^2}{r_h}$$

$$= V_j \,; \quad j = 1, 2, \ldots, p. \tag{2.2.4}$$

The inclusion of cost of measuring various sampling units in survey may be considered as the problem of sample allocation of multivariate stratified sample survey. Khan *et al.* (2008) and Varshney *et al.* (2011) discussed the method by minimizing the variability of estimator for fixed cost function given as

$$C = \sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + \sum_{h=1}^{L} c_{h2} r_h,$$

where C denotes total budget available for a survey, c_{h0} per-unit cost of making first attempt, c_{h1} is the per-unit cost for n_h units from respondent group and c_{h2} is the per-unit cost for r_h units from nonrespondent group.

It is important to determine a true functional form of the cost function so that the appropriate form should be considered. In most of the situations, the per-unit measurement cost, travel cost within strata, reward to respondent and nonrespondent, and labour cost are important factors for a survey, where reward given to a respondent and nonrespondent may reflect the preciousness of the respondent's and non-respondent's view point, availability, approachability, time, etc., and labour cost may be multiple of time units consumed to collect the data from the respondents and nonrespondents units. If the time period, is specified with rate λ for obtaining the information from the sample, follows an exponential distribution (Hogg and Craig (1978), Krysan *et al.* (1994) and Ross (2009)), then probability distribution function of time is approximated by the expression $f(x) = \lambda e^{-\lambda x}$, $\forall\, x > 0$ and $\lambda = 1/(\text{average time})$. For the h^{th} stratum, the sum of independently identically distributed (i.i.d.) random variables follows Gamma distribution with parameters (n_h, λ). Therefore the time for collecting the information from all strata follows Gamma function with the parameters $(\sum_{h=1}^{L} n_h, \lambda)$.

Muhammad *et al.* (2015) used the following cost function, with Gamma function for considering labour time, which is given as

$$C' = \sum_{h=1}^{L} c'_h n_h + \sum_{h=1}^{L} t_h n^{\delta} + \omega \int_0^{\infty} t\lambda e^{-\lambda t} \frac{(\lambda t)^{\sum_{h=1}^{L} n_h - 1}}{(\sum_{h=1}^{L} n_h - 1)!} dt,$$

where $C' = C - c_0$, c'_h unit cost along with reward paid to respondents in stratum h, t is time taken by the interviewer, δ represents the effect of travel within strata, and Gamma function represents the effect of labour cost. $c'_h = c_h + r_h$, where r_h is reward paid within the h stratum, equally to all n_h units and ω is the cost of unit time.

The aggregate expected cost for labour time may be taken as by using the following expression for different choices of λ which may be estimated by using the methods given in Hogg and Craig (1978) and Ross (2009).

$$\omega \sum_{h=1}^{L} E\left(T_h\right) = \omega \sum_{h=1}^{L} \left(\int_0^{\infty} t\lambda e^{-\lambda t} \frac{(\lambda t)^{\sum_{h=1}^{L} n_h - 1}}{(\sum_{h=1}^{L} n_h - 1)!} dt \right)$$

$$= \omega \sum_{h=1}^{L} \frac{n_h}{\lambda}.$$

In the view of above costs, the following cost function may be considered as

$$C_0 = \sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + \sum_{h=1}^{L} c_{h2} r_h + \omega_1 \sum_{h=1}^{L} \frac{n_h}{\lambda_1} + \omega_2 \sum_{h=1}^{L} \frac{r_h}{\lambda_2}, \tag{2.2.5}$$

where $C_0 = C - c_0$ denotes total budget available for a survey, ω_1 denotes the cost of unit time at first attempt, ω_2 denotes the cost of unit time at second attempt and $\frac{1}{\lambda_1}$, $\frac{1}{\lambda_2}$ are average times for labour individual during the collection of information from sample n_h and subsample r_h units respectively.

2.3 Solution using a lexicographic goal programming approach

For the lexicographic goal programming approach (Díaz-García and Cortez (2006)), the order for variances of characteristics is taken, according to their importance, where the most important characteristic is preferred first. Thus the problem becomes the estimation of p - variances in the order $V_1, V_2, \ldots, V_p$. At the first stage of the solution, the following integer nonlinear programming problem (INLPP) has been obtained.

$$\begin{aligned}
&\text{Minimize} && V_1 \\
&\text{subject to} && \sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + \sum_{h=1}^{L} c_{h2} r_h + \omega_1 \sum_{h=1}^{L} \frac{n_h}{\lambda_1} + \omega_2 \sum_{h=1}^{L} \frac{r_h}{\lambda_2} \le C_0 \\
& && 2 \le n_h \le N_h,\ 2 \le r_h \le \hat{n}_{h2} \\
&\text{and} && n_h,\ r_h \text{ are integers; } h = 1, 2, \ldots, L.
\end{aligned} \tag{2.3.1}$$

Let V_1^0 be the optimum value of V_1 and $x_1 \ge 0$ is the increase in V_1^0 for not using the optimum value of the variance for particular characteristic, such that $V_1 - V_1^0 \le x_1$.

At the second stage, the following INLPP will be given as

$$\begin{aligned}
&\text{Minimize} && V_2 + x_1 \\
&\text{subject to} && V_1 - x_1 \le V_1^0
\end{aligned} \tag{2.3.2}$$

$$\sum_{h=1}^{L}(c_{h0}+c_{h1}W_{h1})n_h+\sum_{h=1}^{L}c_{h2}r_h+\omega_1\sum_{h=1}^{L}\frac{n_h}{\lambda_1}+\omega_2\sum_{h=1}^{L}\frac{r_h}{\lambda_2}\le C_0$$

$$x_1\ge 0,\ 2\le n_h\le N_h,\ 2\le r_h\le \hat{n}_{h2}$$

and $\quad n_h,\ r_h$ are integers; $h=1,2,\dots,L.$

Continuing in this manner, at the p^{th} stage, the final solution will be obtained by solving the following INLPP:

$$\begin{aligned}
\text{Minimize}\quad & V_p+\sum_{j=1}^{p-1}x_j\\
\text{subject to}\quad & V_j-x_j\le V_j^0;\ j=1,2,\dots,p-1\\
& \sum_{h=1}^{L}(c_{h0}+c_{h1}W_{h1})n_h+\sum_{h=1}^{L}c_{h2}r_h+\omega_1\sum_{h=1}^{L}\frac{n_h}{\lambda_1}+\omega_2\sum_{h=1}^{L}\frac{r_h}{\lambda_2}\le C_0\\
& x_j\ge 0;\ j=1,2,\dots,p-1\\
& 2\le n_h\le N_h,\ 2\le r_h\le \hat{n}_{h2}\\
\text{and}\quad & n_h,\ r_h \text{ are integers; } h=1,2,\dots,L,
\end{aligned}\tag{2.3.3}$$

where $x_j\ge 0;\ j=1,2,\dots,p-1$ are goal variables whose values are to be determined. These variables will give the minimum possible increases in the variances of p characteristics. $V_j^0;\ j=1,2,\dots,p$ denote the variances under individual optimum allocation for fixed cost for j^{th} characteristic. The solutions of these p - INLPPs are obtained by using MATLAB.

2.4 Numerical illustration

The following three numerical examples are presented to illustrate the use of lexicographic goal programming approach. The values of S^2_{jh2} and S^2_{jh} are not available in general. Let us assume the value of S^2_{jh2}/S^2_{jh} is constant and given as 0.25 for $j=1,2,\dots,p$ and $h=1,2,\dots,L$. However, these ratios may vary from stratum to stratum and from characteristic to characteristic. So it may be handled accordingly. The cost for a unit time of labour is denoted by ω (say $100,150$, etc. per hour per individual may be considered).

For time, the estimates of Gamma function $\int_0^\infty \lambda e^{-\lambda t} \frac{(\lambda t)^{\sum_{h=1}^{L} n_h - 1}}{(\sum_{h=1}^{L} n_h - 1)!} dt$ may be replaced by expected time $\sum_{h=1}^{L} E(T_h) = \sum_{h=1}^{L} \frac{n_h}{\lambda}$ for different values of λ, to collect information from respondent and nonrespondent units (say 15 minutes, 20 minutes, etc., on average from an individual).

Example 1. Consider a population of size $N = 3850$ divided into four strata and two characteristics. The following data are from Khan *et al.* (2008). The values of N_h, S_{1h}^2, S_{2h}^2, W_{h1}, W_{h2}, c_{h0}, c_{h1} and c_{h2} are given in Table 2.1. The total cost C_0 of the survey is fixed as 5000 units.

Table 2.1: Data with four strata and two characteristics

h	N_h	S_{1h}^2	S_{2h}^2	W_{h1}	W_{h2}	c_{h0}	c_{h1}	c_{h2}
1	1214	4817.72	8121.15	0.70	0.30	1	2	3
2	822	6251.26	7613.52	0.80	0.20	1	3	4
3	1028	3066.16	1456.40	0.75	0.25	1	4	5
4	786	6207.25	6977.72	0.72	0.28	1	5	6

For the first stage, the INLPP given by equation (2.3.1) with the given values, is formulated as

$$\begin{aligned}
\text{Minimize} \quad V_1 = {} & \frac{443.0974013}{n_1} + \frac{270.715638}{n_2} + \frac{204.9419482}{n_3} \\
& + \frac{240.6056486}{n_4} + \frac{10.7780449}{r_1} + \frac{2.849638294}{r_2} \\
& + \frac{3.415699137}{r_3} + \frac{5.070828722}{r_4} \\
\text{subject to} \quad & 2.4n_1 + 3.4n_2 + 4n_3 + 4.6n_4 + 3r_1 + 4r_2 + 5r_3 + 6r_4 \\
& + 25(n_1 + n_2 + n_3 + n_4) + 25(r_1 + r_2 + r_3 + r_4) \leq 5000 \\
& 2 \leq n_1 \leq N_1, \quad 2 \leq r_1 \leq \hat{n}_{12} \\
& 2 \leq n_2 \leq N_2, \quad 2 \leq r_2 \leq \hat{n}_{22} \\
& 2 \leq n_3 \leq N_3, \quad 2 \leq r_3 \leq \hat{n}_{32} \\
& 2 \leq n_4 \leq N_4, \quad 2 \leq r_4 \leq \hat{n}_{42} \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ n_4,\ r_1,\ r_2,\ r_3,\ r_4 \text{ are integers.}
\end{aligned} \tag{2.4.1}$$

The solution of INLPP given in (2.4.1) is obtained by using MATLAB, for characteristic $j = 1$, given as

$$n_{1,1}^* = 50,\ n_{1,2}^* = 37,\ n_{1,3}^* = 32,\ n_{1,4}^* = 35,\ r_{1,1}^* = 8,\ r_{1,2}^* = 4,\ r_{1,3}^* = 4,\ r_{1,4}^* = 5,$$

with minimum value of the objective function that is variance V_1 as $V_1^* = 33.3852$.

For the second stage, the INLPP given in (2.3.2) may be formulated as

$$\begin{aligned}
\text{Minimize} \quad V_2 = {} & \frac{746.921876}{n_1} + \frac{329.7093584}{n_2} + \frac{97.34568757}{n_3} + \frac{270.4706345}{n_4} \\
& + \frac{18.16836996}{r_1} + \frac{3.470624826}{r_2} + \frac{1.622428126}{r_3} \\
& + \frac{5.700241329}{r_4} + x_1 \\
\text{subject to} \quad & \frac{443.0974013}{n_1} + \frac{270.715638}{n_2} + \frac{204.9419482}{n_3} + \frac{240.6056486}{n_4} \\
& + \frac{10.7780449}{r_1} + \frac{2.849638294}{r_2} + \frac{3.415699137}{r_3} \\
& + \frac{5.070828722}{r_4} - x_1 \le 33.3852, \\
& 2.4n_1 + 3.4n_2 + 4n_3 + 4.6_4 + 3r_1 + 4r_2 + 5r_3 + 6r_4 \\
& + 25(n_1 + n_2 + n_3 + n_4) + 25(r_1 + r_2 + r_3 + r_4) \le 5000 \\
& x_1 \ge 0,\ 2 \le n_1 \le N_1, \quad 2 \le r_1 \le \hat{n}_{12} \\
& 2 \le n_2 \le N_2, \quad 2 \le r_2 \le \hat{n}_{22} \\
& 2 \le n_3 \le N_3, \quad 2 \le r_3 \le \hat{n}_{32} \\
& 2 \le n_4 \le N_4, \quad 2 \le r_4 \le \hat{n}_{42} \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ n_4,\ r_1,\ r_2,\ r_3,\ r_4 \text{ are integers.}
\end{aligned} \tag{2.4.2}$$

The optimal solution provided by MATLAB for the INLPP, given in (2.4.2), is given as

$$n_{2,1}^* = 54,\ n_{2,2}^* = 38,\ n_{2,3}^* = 27,\ n_{2,4}^* = 35,\ r_{2,1}^* = 8,\ r_{2,2}^* = 4,\ r_{2,3}^* = 4,\ r_{2,4}^* = 5,$$

$$x_1^* = 0.3516,$$

with minimum value of the objective function that is variance V_1 and V_2 as $V_1^* = 33.7368$ and $V_2^* = 38.8776$, respectively. The optimal values of the sampling fractions among non-respondents are given as

$$k_h^* = W_{h2} n_h^* / r_h^*, \; h = 1,2,3,4.$$

$$1/k_1^* = 0.4938272, \; 1/k_2^* = 0.5263158, \; 1/k_3^* = 0.5925926, \; 1/k_4^* = 0.5102041.$$

For this example, the total sample size of the first phase sampling is $n = 154$ and for second phase of sampling is $r = 21$.

Example 2. The data, given in Table 2.2, for three strata and two characteristics, are taken from the population of size $N = 200$. The total available budget is fixed as 1000 units.

Table 2.2: Data for three strata and two characteristics

h	N_h	S_{1h}^2	S_{2h}^2	W_{h1}	W_{h2}	c_{h0}	c_{h1}	c_{h2}
1	60	4	9	0.70	0.30	1	2	3
2	90	16	4	0.80	0.20	1	3	4
3	50	196	289	0.75	0.25	1	4	5

For this example, at the first stage, the following INLPP is to be solved

$$\begin{aligned}
\text{Minimize} \quad V_1 &= \frac{0.333}{n_1} + \frac{3.0780}{n_2} + \frac{11.48438}{n_3} + \frac{0.0081}{r_1} \\
&\quad + \frac{0.0324}{r_2} + \frac{0.1914062}{r_3} \\
\text{subject to} \quad & 2.4n_1 + 3.4n_2 + 4n_3 + 3r_1 + 4r_2 + 5r_3 \\
& + 25(n_1 + n_2 + n_3) + 25(r_1 + r_2 + r_3) \leq 1000 \\
& 2 \leq n_1 \leq N_1, \quad 2 \leq r_1 \leq \hat{n}_{12} \\
& 2 \leq n_2 \leq N_2, \quad 2 \leq r_2 \leq \hat{n}_{22}
\end{aligned} \tag{2.4.3}$$

$$2 \le n_3 \le N_3, \quad 2 \le r_3 \le \hat{n}_{32}$$

and $\quad n_1,\ n_2,\ n_3,\ r_1,\ r_2,\ r_3$ are integers.

The solution of the above INLPP obtained by using MATLAB for $j = 1$ is given as

$$n^*_{1,1} = 3,\ n^*_{1,2} = 9,\ n^*_{1,3} = 16,\ r^*_{1,1} = 2,\ r^*_{1,2} = 2,\ r^*_{1,3} = 2,$$

with minimum value of the objective function that is variance V_1 as $V_1^* = 1.2867$.

For the second stage, the problem is formulated as

$$\begin{aligned}
\text{Minimize} \quad & V_2 = \frac{0.74925}{n_1} + \frac{0.7695}{n_2} + \frac{16.93359}{n_3} + \frac{0.333}{r_1} + \frac{0.0081}{r_2} \\
& + \frac{0.2822266}{r_3} + x_1 \\
\text{subject to} \quad & \frac{0.333}{n_1} + \frac{3.0780}{n_2} + \frac{11.48438}{n_3} + \frac{.0081}{r_1} + \frac{.0324}{r_2} \\
& + \frac{0.1914062}{r_3} - x_1 \le 0.8388, \\
& 2.4n_1 + 3.4n_2 + 4n_3 + 3r_1 + 4r_2 + 5r_3 \\
& + 25(n_1 + n_2 + n_3) + 25(r_1 + r_2 + r_3) \le 1000 \\
& x_1 \ge 0,\ 2 \le n_1 \le N_1, \quad 2 \le r_1 \le \hat{n}_{12} \\
& 2 \le n_2 \le N_2, \quad 2 \le r_2 \le \hat{n}_{22} \\
& 2 \le n_3 \le N_3, \quad 2 \le r_3 \le \hat{n}_{32} \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ r_1,\ r_2,\ r_3 \text{ are integers.}
\end{aligned} \tag{2.4.4}$$

The solution provided by MATLAB is

$$n^*_{2,1} = 3,\ n^*_{2,2} = 7,\ n^*_{2,3} = 18,\ r^*_{2,1} = 2,\ r^*_{2,2} = 2,\ r^*_{2,3} = 2,\ x^*_1 = 0.0180,$$

with minimum value of the objective function that is variance V_1 and V_2 as $V_1^* = 1.3047$ and $V_2^* = 1.4727$, respectively.

Example 3. In this example the data of Sukhatme *et al.* (1984) is taken for five strata and three characteristics with total available budget of 10000 units.

Table 2.3: Data for five strata and three characteristics

h	N_h	S^2_{1h}	S^2_{2h}	S^2_{3h}	W_{h1}	W_{h2}	c_{h0}	c_{h1}	c_{h2}
1	395	12	56	41.3	0.75	0.25	1	4	6
2	382	80	2132	23.1	0.65	0.35	1	6	7
3	439	1113	565	10.9	0.70	0.30	1	7	9
4	368	84	355	11.5	0.75	0.25	1	9	11
5	416	247	68	38.8	0.72	0.28	1	11	12

For the first stage, we have to solve the problem

$$\begin{aligned}
\text{Minimize} \quad V_1 &= \frac{0.4388203}{n_1} + \frac{2.663113}{n_2} + \frac{49.60277}{n_3} + \frac{2.66616}{n_4} + \frac{9.938173}{n_5} \\
&+ \frac{0.002437891}{r_1} + \frac{0.08937845}{r_2} + \frac{1.206554}{r_3} + \frac{0.044436}{r_4} \\
&+ \frac{0.2094497}{r_5} \\
\text{subject to} \quad & 4n_1 + 4.9n_2 + 5.9n_3 + 7.75n_4 + 8.92n_5 + 6r_1 + 7r_2 + 9r_3 + 11r_4 \\
& + 12r_5 + 25(n_1 + n_2 + n_3 + n_4 + n_5) \\
& + 25(r_1 + r_2 + r_3 + r_4 + r_5) \le 10000 \\
& 2 \le n_1 \le N_1, \quad 2 \le r_1 \le \hat{n}_{12} \\
& 2 \le n_2 \le N_2, \quad 2 \le r_2 \le \hat{n}_{22} \\
& 2 \le n_3 \le N_3, \quad 2 \le r_3 \le \hat{n}_{32} \\
& 2 \le n_4 \le N_4, \quad 2 \le r_4 \le \hat{n}_{42} \\
& 2 \le n_5 \le N_5, \quad 2 \le r_5 \le \hat{n}_{52} \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ n_4,\ n_5,\ r_1,\ r_2,\ r_3,\ r_4,\ r_5 \text{ are integers.}
\end{aligned} \tag{2.4.5}$$

The solution of the above INLPP may be obtained by using MATLAB and is given as

$$n^*_{1,1} = 13,\ n^*_{1,2} = 32,\ n^*_{1,3} = 138,\ n^*_{1,4} = 31,\ n^*_{1,5} = 59,\ r^*_{1,1} = 2,\ r^*_{1,2} = 6,$$

$$r^*_{1,3} = 20,\ r^*_{1,4} = 4,\ r^*_{1,5} = 8,$$

with minimum value of the objective function that is variance V_1 as $V^*_1 = 0.83892$.

For the second stage, the following problem may be formulated as given as

$$
\begin{aligned}
\text{Minimize}\quad V_2 = {} & \frac{2.047828}{n_1}+\frac{70.97196}{n_2}+\frac{25.1802}{n_3}+\frac{11.2677}{n_4}+\frac{2.736015}{n_5}\\
& +\frac{0.03413047}{r_1}+\frac{2.381936}{r_2}+\frac{0.6124914}{r_3}+\frac{0.187795}{r_4}\\
& +\frac{0.05766226}{r_5}+x_1\\
\text{subject to}\quad & \frac{0.4388203}{n_1}+\frac{2.663113}{n_2}+\frac{49.60277}{n_3}+\frac{2.66616}{n_4}+\frac{9.938173}{n_5}\\
& +\frac{0.002437891}{r_1}+\frac{0.08937845}{r_2}+\frac{1.206554}{r_3}+\frac{0.044436}{r_4},\\
& +\frac{0.2094497}{r_5}-x_1 \le 0.8388\\
& 4n_1+4.9n_2+5.9n_3+7.75n_4+8.92n_5+6r_1+7r_2+9r_3\\
& +11r_4+12r_5+25(n_1+n_2+n_3+n_4+n_5)\\
& +25(r_1+r_2+r_3+r_4+r_5)\le 10000\\
& x_1\ge 0,\ 2\le n_1\le N_1,\quad 2\le r_1\le \hat{n}_{12}\\
& 2\le n_2\le N_2,\quad 2\le r_2\le \hat{n}_{22}\\
& 2\le n_3\le N_3,\quad 2\le r_3\le \hat{n}_{32}\\
& 2\le n_4\le N_4,\quad 2\le r_4\le \hat{n}_{42}\\
& 2\le n_5\le N_5,\quad 2\le r_5\le \hat{n}_{52}\\
\text{and}\quad & n_1,\ n_2,\ n_3,\ n_4,\ n_5,\ r_1,\ r_2,\ r_3,\ r_4,\ r_5 \text{ are integers.}
\end{aligned}
\tag{2.4.6}
$$

The INLPP (2.4.6) gives the following solution by using MATLAB

$$n^*_{2,1}=17,\ n^*_{2,2}=93,\ n^*_{2,3}=92,\ n^*_{2,4}=38,\ n^*_{2,5}=36,\ r^*_{2,1}=2,\ r^*_{2,2}=16,$$

$$r^*_{2,3}=14,\ r^*_{2,4}=5,\ r^*_{2,5}=5,\ x^*_1=0.2418,$$

with minimum value of the objective function that is variance V_2 as $V^*_2=2.0274$, respectively.

For the third stage, there is need to solve the following INLPP

$$
\begin{aligned}
\text{Minimize} \quad V_3 = {} & \frac{1.510273}{n_1} + \frac{0.7689739}{n_2} + \frac{0.4857774}{n_3} + \frac{0.36501}{n_4} + \frac{1.561138}{n_5} \\
& + \frac{0.02517122}{r_1} + \frac{0.02580803}{r_2} + \frac{0.01181621}{r_3} + \frac{0.0060835}{r_4} \\
& + \frac{0.03290141}{r_5} + x_1 + x_2 \\
\text{subject to} \quad & \frac{0.4388203}{n_1} + \frac{2.663113}{n_2} + \frac{49.60277}{n_3} + \frac{2.66616}{n_4} + \frac{9.938173}{n_5} \\
& + \frac{0.002437891}{r_1} + \frac{0.08937845}{r_2} + \frac{1.206554}{r_3} + \frac{0.044436}{r_4} \\
& + \frac{0.2094497}{r_5} - x_1 \leq 0.8388 \\
& \frac{2.047828}{n_1} + \frac{70.97196}{n_2} + \frac{25.1802}{n_3} + \frac{11.2677}{n_4} + \frac{2.736015}{n_5} \\
& + \frac{0.03413047}{r_1} + \frac{2.381936}{r_2} + \frac{0.6124914}{r_3} + \frac{0.187795}{r_4} \\
& + \frac{0.05766226}{r_5} - x_2 \leq 2.0274 \\
& 4n_1 + 4.9n_2 + 5.9n_3 + 7.75n_4 + 8.92n_5 + 6r_1 + 7r_2 + 9r_3 \\
& + 11r_4 + 12r_5 + 25(n_1 + n_2 + n_3 + n_4 + n_5) \\
& + 25(r_1 + r_2 + r_3 + r_4 + r_5) \leq 10000 \\
& x_1 \geq 0,\ x_2 \geq 0,\ 2 \leq n_1 \leq N_1, \quad 2 \leq r_1 \leq \hat{n}_{12} \\
& 2 \leq n_2 \leq N_2, \quad 2 \leq r_2 \leq \hat{n}_{22} \\
& 2 \leq n_3 \leq N_3, \quad 2 \leq r_3 \leq \hat{n}_{32} \\
& 2 \leq n_4 \leq N_4, \quad 2 \leq r_4 \leq \hat{n}_{42} \\
& 2 \leq n_5 \leq N_5, \quad 2 \leq r_5 \leq \hat{n}_{52} \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ n_4,\ n_5,\ r_1,\ r_2,\ r_3,\ r_4,\ r_5 \text{ are integers.}
\end{aligned}
\tag{2.4.7}
$$

The INLPP (2.4.7) gives the following solution provided by MATLAB is

$$n^*_{3,1} = 23,\ n^*_{3,2} = 73,\ n^*_{3,3} = 100,\ n^*_{3,4} = 36,\ n^*_{3,5} = 46,\ r^*_{3,1} = 2,\ r^*_{3,2} = 13,$$

$$r^*_{3,3} = 15,\ r^*_{3,4} = 2,\ r^*_{3,5} = 7,\ x^*_1 = 0.1439,\ x^*_2 = 0.00141,$$

with minimum value of the objective function and variances are V_1, V_2 and V_3 as $V^*_1 = 0.9823649$, $V^*_2 = 2.028786$ and $V^*_3 = 0.1482332$, respectively. The corresponding trace value $V^*_1 + V^*_2 + V^*_3 = 3.159384$.

2.5 Some other methods to workout compromise allocation

In this section a study is carried out to compare the proposed allocation with compromise allocations suggested by others as follows:

2.5.1 Cochran's averaged allocation

Cochran (1977) proposed the character wise average of the individual optimum allocations as the compromise allocation. In the existence of nonresponse with Gamma cost function this allocation is reformulated as

$$n_h = \frac{1}{p}\sum_{j=1}^{p} \frac{C_0\sqrt{W_h^2\left(S^2_{jh} - W_{h2}S^2_{jh2}\right)/(c_{h0} + c_{h1}W_{h1} + \omega_1/\lambda_1)}}{\sum_{h=1}^{L}\sqrt{W_h^2\left(S^2_{jh} - W_{h2}S^2_{jh2}\right)(c_{h0} + c_{h1}W_{h1} + \omega_1/\lambda_1)} + \sum_{h=1}^{L} W_h W_{h2} S^2_{jh2}\sqrt{\omega_2/\lambda_2 + c_{h2}}} \tag{2.5.1}$$

$$r_h = \frac{1}{p}\sum_{j=1}^{p} \frac{C_0 W_h W_{h2} S_{jh2}/\sqrt{c_{h2}}}{\sum_{h=1}^{L}\sqrt{W_h^2\left(S^2_{jh} - W_{h2}S^2_{jh2}\right)(c_{h0} + c_{h1}W_{h1} + \omega_1/\lambda_1)} + \sum_{h=1}^{L} W_h W_{h2} S^2_{jh2}\sqrt{\omega_2/\lambda_2 + c_{h2}}}; \tag{2.5.2}$$

$h = 1,2,\ldots,L;\ j = 1,2,\ldots,p.$

2.5.2 Chatterjee's averaged allocation

Chatterjee (1968) proposed the compromise allocation by minimizing the sum of relative increases in the minimum variances in the estimates. For nonresponse, this allocation, with Gamma cost function, reformulated as

$$n_h = \frac{C_0\sqrt{\sum_{j=1}^{p} n^{0^2}_{jh}}}{\sum_{h=1}^{L}(c_{h0} + c_{h1}W_{h1} + \omega_1/\lambda_1)\sqrt{\sum_{j=1}^{p} n^{0^2}_{jh}} + \sum_{h=1}^{L} c_{h2} + \omega_2/\lambda_2\sqrt{\sum_{j=1}^{p} r^{0^2}_{jh}}} \tag{2.5.3}$$

$$r_h = \frac{C_0\sqrt{\sum_{j=1}^{p} r_{jh}^{0^2}}}{\sum_{h=1}^{L}\left(c_{h0}+c_{h1}W_{h1}+\omega_1/\lambda_1\right)\sqrt{\sum_{j=1}^{p} n_{jh}^{0^2}}+\sum_{h=1}^{L} c_{h2}+\omega_2/\lambda_2\sqrt{\sum_{j=1}^{p} r_{jh}^{0^2}}}; \quad (2.5.4)$$

$h = 1,2,\ldots,L;\ j = 1,2,\ldots,p,$

where n_{jh}^{0} and r_{jh}^{0} denote the individual optimum allocations for j^{th} characteristic, respectively.

2.5.3 Khan *et al.* (2008) compromise allocation

Khan *et al.* (2008) proposed the compromise allocation by minimizing the weighted sum of the variances with Gamma cost function is reformulated as

$$n_h = \frac{C_0\sqrt{W_h^2\left(A_h^2-W_{h2}B_h^2\right)/\left(c_{h0}+c_{h1}W_{h1}+\omega_1/\lambda_1\right)}}{\sum_{h=1}^{L}\sqrt{W_h^2\left(A_h^2-W_{h2}B_h^2\right)\left(c_{h0}+c_{h1}W_{h1}+\omega_1/\lambda_1\right)}+\sum_{h=1}^{L} W_h W_{h2} B_h\sqrt{\omega_2/\lambda_2+c_{h2}}} \quad (2.5.5)$$

$$r_h = \frac{C_0 W_h W_{h2} B_h/\sqrt{c_{h2}+\omega_2/\lambda_2}}{\sum_{h=1}^{L}\sqrt{W_h^2\left(A_h^2-W_{h2}B_h^2\right)\left(c_{h0}+c_{h1}W_{h1}+\omega_1/\lambda_1\right)}+\sum_{h=1}^{L} W_h W_{h2} B_h\sqrt{\omega_2/\lambda_2+c_{h2}}}, \quad (2.5.6)$$

where $A_h^2 = \sum_{j=1}^{p} a_j S_{jh}^2$ and $B_h^2 = \sum_{j=1}^{p} a_j S_{jh2}^2;\ \ h = 1,2,\ldots,L;\ j = 1,2,\ldots,p.$

2.5.4 Minimizing trace allocation

Sukhatme *et al.* (1984) obtained by minimizing the trace of variance-covariance matrix of estimator used. This compromise allocation is reformulated as

$$n_h = \frac{C_0\sqrt{W_h^2\left(A_h^2-W_{h2}B_h^2\right)/\left(c_{h0}+c_{h1}W_{h1}+\omega_1/\lambda_1\right)}}{\sum_{h=1}^{L}\sqrt{W_h^2\left(A_h^2-W_{h2}B_h^2\right)\left(c_{h0}+c_{h1}W_{h1}+\omega_1/\lambda_1\right)}+\sum_{h=1}^{L} W_h W_{h2} B_h\sqrt{\omega_2/\lambda_2+c_{h2}}} \quad (2.5.7)$$

$$r_h = \frac{C_0 W_h W_{h2} B_h/\sqrt{c_{h2}+\omega_2/\lambda_2}}{\sum_{h=1}^{L}\sqrt{W_h^2\left(A_h^2-W_{h2}B_h^2\right)\left(c_{h0}+c_{h1}W_{h1}+\omega_1/\lambda_1\right)}+\sum_{h=1}^{L} W_h W_{h2} B_h\sqrt{\omega_2/\lambda_2+c_{h2}}}, \quad (2.5.8)$$

where $A_h^2 = \sum_{j=1}^{p} S_{jh}^2$ and $B_h^2 = \sum_{j=1}^{p} S_{jh2}^2$; $h = 1, 2, \ldots, L$; $j = 1, 2, \ldots, p$.

Table 2.4: Allocations with trace values and cost incurred for example 1

S.No	Allocations	n_1	n_2	n_3	n_4	r_1	r_2	r_3	r_4	Trace	Cost
1	Minimizing Trace	54	37	26	34	8	4	3	5	73.9552	4875.8
2	Khan *et al.* Allocation	54	37	26	34	8	4	3	5	73.9552	4875.8
3	Cochran Allocation	55	38	27	35	7	4	3	4	73.6677	4931.2
4	Chatterjee Allocation	55	38	27	35	8	4	3	5	72.2673	4990.2
5	Proposed Allocation	54	38	27	35	8	4	4	5	72.2482	4992.8

2.6 Discussion

In this chapter, the problem is formulated as INLPP with fixed Gamma cost function. To obtain its solution, a lexicographic goal programming approach is used to provide the best solution. Tables 2.4, 2.5 and 2.6 show the comparison of the proposed allocation with other compromise allocations. Table 2.4 shows that the allocations obtained by minimizing trace and Khan *et al.* (2008) have the same trace value while the proposed allocation gives the minimum trace value compared to all other allocations. Table 2.5 shows that all other allocations are not advisable in comparison to the proposed allocation even these allocations have the minimum trace value but violate the cost constraint. Table 2.6 shows that the allocations obtained by minimizing trace and Khan *et al.* (2008) have the same trace value and cost. Cochran's allocation gives a higher trace value, while Chatterjee's allocation violates the cost constraint with a minimum trace value. The proposed allocation gives the solution by utilizing the maximum cost as possible in comparison to others.

Tables 2.7, 2.8 and 2.9 give the percentage increase in the variances when individual optimum allocations and proposed allocation are used for all characteristics. It can be seen that the percentage increase in the variances by proposed allocation is considerably less than the percentage increase when the individual optimum allocation for one characteristic is used for all characteristics.

Table 2.5: Allocations with trace values and cost incurred for example 2

S.No	Allocations	n_1	n_2	n_3	r_1	r_2	r_3	Trace	Cost
1	Minimizing Trace	4	7	20	2	2	2	2.5113	1062.4
2	Khan *et al.* Allocation	4	7	20	2	2	2	2.5113	1062.4
3	Cochran Allocation	4	7	20	2	2	3	2.4324	1092.4
4	Chatterjee Allocation	4	7	20	2	2	2	2.5113	1062.4
5	Proposed Allocation	3	7	18	2	2	2	2.7594	977.0

Table 2.6: Allocations with trace values and cost incurred for example 3

S.No	Allocations	n_1	n_2	n_3	n_4	n_5	r_1	r_2	r_3	r_4	r_5	Trace	Cost
1	Minimizing Trace	20	90	89	38	37	4	16	13	5	5	3.0533	9963.6
2	Khan *et al.* Allocation	20	90	89	38	37	4	16	13	5	5	3.0533	9963.6
3	Cochran Allocation	37	68	77	37	58	3	12	12	5	7	3.2834	9988.6
4	Chatterjee Allocation	21	91	90	38	37	3	16	13	5	5	3.0305	10022.4
5	Proposed Allocation	23	73	100	36	46	2	13	15	2	7	3.1594	9998.0

Table 2.7: Percentage increase in the variances when individual optimum for one characteristic is used for all characteristics (example 1)

Characteristics	Percentage increase in the variances		
	1	2	Proposed
1	0	2.0186	1.0532
2	1.0094	0	0.9126

Table 2.8: Percentage increase in the variances when individual optimum for one characteristic is used for all characteristics (example 2)

Characteristics	Percentage increase in the variances		
	1	2	Proposed
1	0	6.4044	1.3988
2	1.3959	0	1.2374

Table 2.9: Percentage increase in the variances when individual optimum for one characteristic is used for all characteristics (example 3)

Characteristics	Percentage increase in the variances			
	1	2	3	Proposed
1	0	94.7409	41.6681	15.508
2	28.0618	0	22.4040	7.3134
3	15.4577	14.1215	0	5.6085

2.7 Conclusion

On the basis of the comparative study carried out in Section 2.5, it may be concluded that the proposed allocation is suitable for working out a usable compromise allocation for multivariate stratified surveys with the Gamma cost function.

CHAPTER 3

On determination of optimum strata boundaries in sample surveys

3.1 Introduction

Many researchers adopted stratified random sampling to obtain the required information in every field of study , that is, health related studies, market shares related studies in business, etc. Many studies arrange individuals into convivial strata of socio-economic elements. Let the population under study is stratified into L strata and the number of units in each stratum is N_h $(h = 1, 2, \ldots, L)$ such that the total of population units be $N = \sum_{h=1}^{L} N_h$. The main objective to construct the strata boundaries is to control homogeneity internally. That is, the stratum variances S_h^2 may be minimized as possible. If the study variable's distribution is known, then the optimum stratum boundaries (OSB) may be obtained by dividing the range of the distribution at various possible points between the range. The problem of OSB was discussed by Dalenius (1950). He considered the study variable as a stratification variable and constructed minimal equations to obtain OSB. But, these equations may not be explained clearly due of their implicit nature. The problem of determining OSB is discussed by many authors, some of them are Dalenius and Gurney (1951), Mahalanobis (1952), Hansen *et al.* (1953), Dalenius and Hodges (1959), Sethi (1963), Unnithan (1978), Lavallée and Hidiroglou (1988), Hedlin (2000), etc. Unnithan (1978) suggested an iterative procedure that needs a suitable initial solution for the skewed population. Lavallée and Hidiroglou (1988) discussed an algorithm to construct stratum boundaries for a power allocated (applying an exponential value q, where $0 < q < 1$, to the stratum population value under Neyman allocation to allow for a sufficient spread of the sample allocation) stratified sample. Later on, Hidiroglou and Srinath (1993) presented a more general form of the algorithm. After Lavallée and Hidiroglou's algorithm, a modified algorithm that incorporated the different relationships between the stratification and study variables were proposed by Sweet and Sigman (1998) and Rivest (2002). There are various other algorithms discussed in the literature, for example, Niemiro (1999) suggested the use of a random search approach, and Nelder and Mead (1965) used the simplex technique for stratification (Lednick and Wieczorkowski (2003)). Later on, Kozak (2004) discussed a modified arbitrary search technique. Gunning and Horgan (2004) proposed another method for stratification, primarily based on a geometric progression. This strategy was studied with three other techniques, Dalenius and Hodges (1959), Ekman (1959), and Lavallée and Hidiroglou (1988), which expressed that the geometric progression technique is efficient than Horgan (2006). The usefulness of Gunning and Horgan's geometric progression technique was compared, and it has shown that the geometric progression approach is much less efficient than Lavallée and Hidiroglou's method. Several stratification techniques have been suggested by Lavallée (1988), Khan *et al.* (2002), Khan *et al.* (2005, 2015), Kozak *et al.* (2007), etc., when the distributions of the study variables are known. They constructed the problems of determining OSB as the optimization problems, and solved by using DP. The DP technique was first proposed by Bühler and Deutler (1975), to solve the OSB problems. Khan *et al.* (2009) worked on the OSB problem in health surveys by using auxiliary variables, and many other authors also solved OSB problems such as Reddy *et al.* (2016), Danish *et al.* (2017), Danish (2018), Hidiroglou and Kozak (2018), Danish *et al.* (2019a,b), Reddy and Khan (2019) and Zhanga *et al.* (2020). Moreover, some authors solved the multiobjective NLPP with fixed linear and nonlinear cost functions. For example Varshney *et al.* (2012, 2015, 2017) and Ansari *et al.* (2018), etc. Recently, Haq *et al.* (2020) also discussed the solution of multiobjective NLPP.

In this chapter, a procedure is discussed to solve the OSB problem and sample size with a nonlinear fixed cost function for each stratum is determined. The application of the proposed method is given with empirical investigations using real and simulated data sets. Generally, in health surveys with the help of auxiliary variables, we identify positively skewed distributions which are considered into the family of Gaussian distribution such as Weibull, Gamma, Log-normal, etc. Thus in this chapter, the problem is formulated as an NLPP for determining the OSB, which shows a gain in efficiency over other methods as defined in the literature. The proposed method is discussed for the exponential and Weibull distribution using real and simulated data sets for the NLPP. This NLPP is solved by minimizing the coefficients of variation (CV) of unknown population parameters. Since the formulated NLPP may be defined as a multistage decision problem, therefore it is solved by using suitable mathematical programming technique. After getting OSBs, they are used to compute each stratum's sample size with a fixed nonlinear cost function. A numerical example with a real data set of skewed population, where the auxiliary variables follow exponential and Weibull distributions, is presented to illustrate the proposed procedure. Also, using a simulated data set, a comparative study is included with some other methods.

3.2 Formulation of the OSB problem as an NLPP

Let the population be stratified into L strata and the estimation of the population mean of the study variable is needed. Let x_0 and x_L be the lowest and upper values of the study variable in the population. Then the problem of optimum stratification can be described as to find the intermediate stratum boundaries $x_1 \leq x_2 \leq \cdots \leq x_{L-1}$ such that the variance of the stratified sample mean given as

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h,$$

and variance is given as

$$V(\bar{x}_{st}) = \sum_{h=1}^{L} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 \sigma_h^2, \tag{3.2.1}$$

where σ_h^2 is the stratum variance for the h^{th} stratum, $h = 1, 2, \ldots, L$. If fpc is ignored, then equation (3.2.1) restated as

$$V(\bar{x}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h}. \tag{3.2.2}$$

In this chapter, the square of the coefficient of variation instead of variance is suggested to use because it is unit free and also has positive value. Thus, the square of the coefficient of variation may be defined as

$$(CV)^2 = \frac{1}{\overline{X}^2} \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h}, \tag{3.2.3}$$

where

$$CV = \frac{SD(\bar{x}_{st})}{\overline{X}}$$

$$(CV)^2 = \frac{V(\bar{x}_{st})}{\overline{X}^2}.$$

In a stratified sample survey for the total cost of the survey, the basic cost function may be considered as (Cochran (1977))

$$C = c_0 + \sum_{h=1}^{L} c_h n_h,$$

or

$$C - c_0 = C_0 = \sum_{h=1}^{L} c_h n_h,$$

where C_0 denotes the budget needed for measuring the units to conduct a survey, c_h is per-unit measurement cost to measure all p characteristics for the h^{th} stratum, n_h is the sample size of units selected from the h^{th} stratum and c_0 is the overhead cost amount. In practical conditions the travel cost within the stratum may be suggested, then the cost function becomes nonlinear. Beardwood *et al.* (1959) showed that the travel cost of approaching n_h units within stratum may be taken as $t_h\sqrt{n_h}$. By adding this cost component, the total cost C of the survey is given by

$$C = c_0 + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h\sqrt{n_h},$$

or

$$C_0 = C - c_0 = \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h\sqrt{n_h}, \tag{3.2.4}$$

where $t_h\sqrt{n_h}$ is the travel cost for h^{th} stratum. The cost capacity in equation (3.2.4) is now quadratic in $\sqrt{n_h}$.

A method of constructing OSB for strata is needed for study variable for getting substantial gains in the precision of the estimates. Let $f(y)$ be the probability density function of the auxiliary variable y. The problem of determining the OSB is then equivalent to finding the $L-1$ intermediate points $x_1 \le x_2 \le \cdots \le x_{L-1}$ in the interval $[x_0, x_L]$ such that equation (3.2.2) is minimum. Let $x_L - x_0 = d$.

The values of W_h and σ_h in equation (3.2.2) are obtained as

$$W_h = \int_{x_{h-1}}^{x_h} f(y)dy \tag{3.2.5}$$

$$\sigma^2_{h\lambda} = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} y^2 f(y)dy - \mu^2_{h\lambda}, \tag{3.2.6}$$

where $$\mu_{h\lambda} = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} y f(y)dy \tag{3.2.7}$$

and $(x_{h-1},\, x_h)$ are the boundaries of h^{th} stratum.

When the distribution of $f(y)$ function is known, then using equations (3.2.5) to (3.2.7) could be expressed as a the function of x_h and x_{h-1} only.

Let $f_h\left(x_h, x_{h-1}\right) = \frac{1}{\overline{X}^2}\left(\frac{W_h^2\sigma_h^2}{n_h}\right)$. Then, the problem of determination of OSB can be expressed as the optimization problem which is given as

$$\begin{aligned}
\text{Minimize} \quad & \sum_{h=1}^{L} f_h\left(x_h, x_{h-1}\right) \\
\text{subject to} \quad & C = c_0 + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{l} t_h\sqrt{n_h} \\
& a = x_0 \le x_1 \le x_2 \le \cdots \le x_{L-1} \le x_L = b \\
\text{and} \quad & l_h \ge 0;\ h = 1,2,\ldots,L,
\end{aligned} \tag{3.2.8}$$

where a is the minimum value of all lower bounds of the study variable and b is the maximum value of all upper bounds of the study variable and $l_h = x_h - x_{h-1}$; $h = 1,2,\ldots,L$ and $l_h \ge 0$ denotes the range or the width of the h^{th} stratum.

Range of the distribution is expressed as a function of stratum width

$$\sum_{h=1}^{L} l_h = \sum_{h=1}^{L} x_h - x_{h-1} = b - a = x_L - x_0 = d. \tag{3.2.9}$$

The h^{th} stratification point x_h; $h = 1,2,\ldots,L$ can be defined as

$$x_h = x_{h-1} + l_h.$$

Substituting equation (3.2.9) as a constraint, the algorithm can be defined as the problem of determining optimum strata widths (OSW) $l_1, l_2, \ldots, l_L$ and written as

$$\begin{aligned}
\text{Minimize} \quad & \sum_{h=1}^{L} f_h\left(x_h, x_{h-1}\right) \\
\text{subject to} \quad & C = c_0 + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{l} t_h \sqrt{n_h} \\
& \sum_{h=1}^{L} l_h = d \\
\text{and} \quad & l_h \geq 0;\ h = 1, 2, \ldots, L.
\end{aligned} \tag{3.2.10}$$

Since x_0 is initial value. Then initial term, $f_1\left(l_1, x_0\right) = f_1\left(x_0 + l_1, x_0\right)$ in the objective function of equation (3.2.10) is a function of l_1 only. If l_1 is known, the second term $f_2\left(l_2, x_1\right)$ will become a function of l_2 alone and so on. The objective function and constraints of the MPP (3.2.10) are separable, therefore, any suitable mathematical programming technique may be used to solve the MPP (3.2.10). The MPP may be expressed as a function of l_h which may be given as

$$\begin{aligned}
\text{Minimize} \quad & \sum_{h=1}^{L} f_h\left(l_h\right) \\
\text{subject to} \quad & C = c_0 + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{l} t_h \sqrt{n_h} \\
& \sum_{h=1}^{L} l_h = d \\
\text{and} \quad & l_h \geq 0;\ h = 1, 2, \ldots, L.
\end{aligned} \tag{3.2.11}$$

3.3 The solution procedure

The MPP (3.2.11) is a multistage decision problem where the objective function and constraints may be expressed as separable functions of l_h, so the DP technique may be applicable to solve the formulated problem. Using this technique, a multivariable problem may be decomposed into stages, and each stage will be converted into a single variable subproblem. The DP technique is implemented by solving a recursive equation based on Bellman's principle of optimality.

The k^{th} subproblem of (3.2.11) which involves first $k(\leq L)$ strata may be formulated as

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{h=1}^{k} f_h\left(l_h\right) \\
\text{subject to} \quad & C = c_0 + \sum_{h=1}^{k} c_h n_h + \sum_{h=1}^{k} t_h \sqrt{n_h} \\
& \sum_{h=1}^{k} l_h = d_k \\
\text{and} \quad & l_h \geq 0;\ h = 1,2,\ldots,k,
\end{aligned}
\tag{3.3.1}
$$

where $d_k \leq d$ is the total width for k strata. Also, if we put $k = L$ then $d_k = d$ and the transformation functions are used as

$$
\begin{aligned}
& d_k = l_1 + l_2 + \ldots + l_k \\
& d_{k-1} = l_1 + l_2 + \ldots + l_{k-1} = d_k - l_k \\
& d_{k-2} = l_1 + l_2 + \ldots + l_{k-2} = l_{k-1} - l_{k-1} \\
& \vdots \\
& d_2 = l_1 + l_2 = d_3 - l_3 \\
& d_1 = l_1 = d_2 - l_2.
\end{aligned}
\tag{3.3.2}
$$

If $f\left(k, d_k\right)$ denotes the minimum value of the objective function of (3.3.1), then

$$
f\left(k, d_k\right) = \min\left[\sum_{h=1}^{k} f_h\left(l_h\right) \middle| \sum_{h=1}^{k} l_h = d_k \text{ and } l_h \geq 0;\ h = 1,2,\ldots,k\right]. \tag{3.3.3}
$$

By using the following recurrence relation, the minimum value of the objective function of (3.2.11) may be obtained

$$
f\left(k, d_k\right) = \min_{0 \leq l_k \leq d_k} \left[f_k\left(l_k\right) + f\left(k-1, d_k - l_k\right)\right],\ 2 \leq k \leq L \tag{3.3.4}
$$

and consequently, the required solution to the problem.

3.4 Numerical example

In this section, the numerical illustrations are placed to express the proposed technique's application to the real and simulated data sets for different populations. The OSBs are obtained with a nonlinear cost function.

3.4.1 Real data set

For this numerical illustration, the data from a hospital are considered, as taken by Khan *et al.* (2009). The population of size $N = 393$ and number of beds is 1000. For the whole population with auxiliary variable y, a common model may be easily available for estimation purposes. Strata formation may be initiated by ordering the units from low to high values based on an auxiliary variable. A model is required with

$$E_M(x_h) = y_h^T \beta,\ V_M(x_h) = \sigma^2 y_h^\gamma, \tag{3.4.1}$$

in most of the populations, for $0 \le \gamma \le 2$. Valliant *et al.* (2000) defined a model which is given as

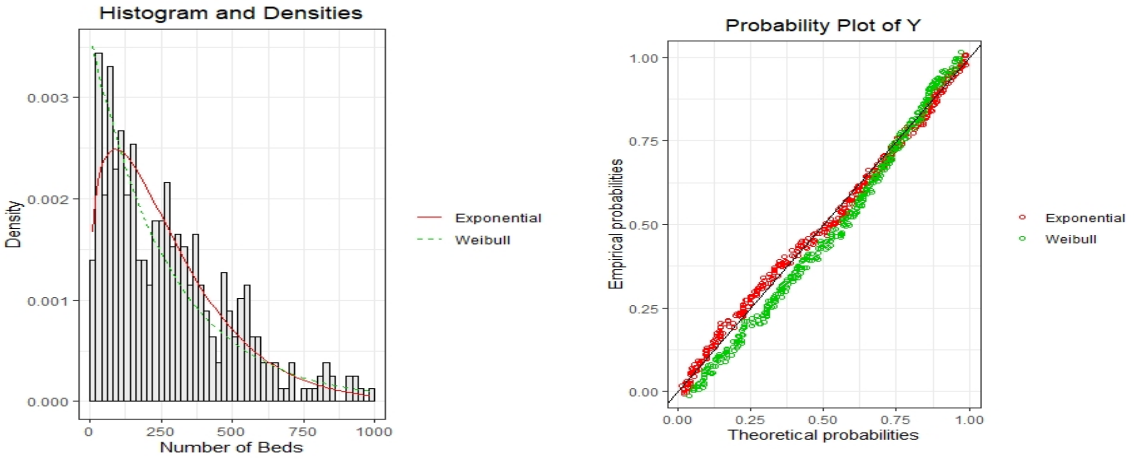
$$E_M(x_h) = \beta_0 + \beta_{\gamma/2} y_h^{\gamma/2} + \beta_\gamma y_h^\gamma. \tag{3.4.2}$$

In equation (3.4.2), the fitted model is quadratic. Here the number of patients discharged is denoted by x and the number of beds y is considered to be best for the model $\gamma = 2$. The R^2 value for this model is 0.855 and estimated parameters are highly significant. A probability plot of the number of beds (y) is obtained to determine the distribution of y.

To identify the distribution of $f(y)$, a relative histogram of y is plotted and in Figure 3.1 (a), it has shown that the distribution of y follows the right-skewed distribution. This distribution has some similar properties as that of exponential and Weibull distributions. It may be seen that the probability plot of y, in Figure 3.1 (b), shows that the points are clustered around a straight line, which enables us that it may be considered as exponential and

Weibull distributions. The MLE method is used to estimate the parameters; the minimum and maximum values for exponential and Weibull distributions are given in Table 3.1. The total amount available for conducting the survey is assumed to be $C = 800$ units and cost values as given in Table 3.3.

Figure 3.1: Histogram and probability plot for exponential and Weibull distributions using a real data set



(a) **(b)**

Table 3.1: Parametric values for real data

Parameters	Exponential	Weibull
Shape		1.18198
Scale	0.0036	280.349
Location		9.35936
Minimum	10	10
Maximum	986	986

3.4.2 Simulated data sets

For exponential and Weibull distributions with auxiliary and study variables, the simulated data sets were randomly generated using R software. The size of each population is taken as 1000 units. In this data set, a weak relationship appears between y and x for predictive power in its regression models given in equations (3.4.1) and (3.4.2) for $\gamma = 2$. The R^2 value for exponential distribution is 0.00342829 and for Weibull distribution is 0.0002398003. The parametric values for exponential and Weibull distributions are estimated using R software and given in Table 3.2. The total amount available for conducting the survey is assumed to be $C = 1500$ units and cost values as given in Table 3.3.

Table 3.2: Parametric values for simulated data

Parameters	Exponential	Weibull
Shape		0.7510
Scale	0.069724	0.1713
Location		6.5001
Minimum	0.01286	6.65001
Maximum	132.19440	96.4011

Table 3.3: Cost values for real and simulated data sets

h	1	2	3	4	5
c_h	2	3	4	5	6
t_h	1	1.5	2	2.5	3

3.4.3 Determination of OSB under exponential distribution

Let the stratification variable y follows the exponential distribution with parameter $\lambda \geq 0$ and it is given as

$$f(y) \quad = \lambda \exp(-\lambda y), \quad 0 \leq y < \infty \tag{3.4.3}$$

In most of the cases, the populations' sizes are often finite, so it is assumed that the most considerable value of y in the population is D, that is, $0 \leq y \leq D$. here $y_0 = 0$ and $y_L = D$. Using equations (3.2.5) to (3.2.7), it may be written as

$$W_h = \exp(-\lambda x_{h-1})\left(1 - \exp(-\lambda l_h)\right) \tag{3.4.4}$$

$$\mu_h = \frac{\left(x_{h-1} + \frac{1}{\lambda}\right)\left(1 - \exp(-\lambda l_h)\right) - l_h \exp\left(-\lambda l_h\right)}{1 - \exp\left(-\lambda l_h\right)} \tag{3.4.5}$$

and $$\sigma_h^2 = \frac{\frac{1}{\lambda^2}\left(1-\exp\left(-\lambda l_h\right)\right)^2 - l_h^2 \exp\left(-\lambda l_h\right)}{\left(1-\exp\left(-\lambda l_h\right)\right)^2}, \tag{3.4.6}$$

by using equations (3.4.4) to (3.4.6), the problem of OSB may be expressed as

$$\begin{aligned}
&\text{Minimize} && \frac{1}{\overline{X}^2}\sum_{h=1}^{L} \frac{\exp\left(-\lambda x_{h-1}\right)^2 \frac{\left(1-\exp(-\lambda l_h)\right)^2}{\lambda^2} - l_h^2 \exp\left(-\lambda l_h\right)}{n_h} \\
&\text{subject to} && C = c_0 + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{l} t_h \sqrt{n_h} \\
& && \sum_{h=1}^{L} l_h = d \\
&\text{and} && l_h \geq 0;\ h = 1,2,\ldots,L.
\end{aligned} \tag{3.4.7}$$

3.4.4 Determination of OSB under Weibull distribution

The Weibull distribution with three parameters is continuous probability distribution. It may be frequently used as a model in life testing. Due to its moderate skewness nature, it may be used in fitting of distributions of different nature. By substituting different values for shape parameter, the distribution may be converted into the exponential distribution, the Rayleigh distribution and the normal distribution. If the stratification variable y follows the Weibull distribution on the interval $[x_0, x_L]$ with parameters θ, γ, and r. Then the probability density function is given as

$$\begin{aligned}
f\left(y; r, \theta, \gamma\right) &= \frac{r\left(y-\gamma\right)^{(r-1)}}{\theta^r} \exp\left(-\left(\frac{y-\gamma}{\theta}\right)^r\right); \quad y \geq \gamma,\ r > 0,\ \theta > 0, \\
& \qquad -\infty < \gamma < \infty,
\end{aligned} \tag{3.4.8}$$

where r and θ are shape and scale parameters, respectively and γ is a location parameter. By using the equations (3.2.5) to (3.2.7), the expression of $W_h^2\sigma_h^2$ for Weibull distribution may be expressed as

$$
\begin{aligned}
&\theta^2\Gamma\left(\frac{2}{r}+1\right)\left[\exp\left(-\frac{(x_{h-1}-\gamma)}{\theta}\right)^r-\exp\left(-\frac{(x_{h-1}+l_h-\gamma)}{\theta}\right)^r\right]\times\\
&\left[\text{\c{Q}}\left(\frac{2}{r}+1,\left(\frac{(x_{h-1}-\gamma)}{\theta}\right)^r\right)-\text{\c{Q}}\left(\frac{2}{r}+1,\left(\frac{(x_{h-1}+l_h-\gamma)}{\theta}\right)^r\right)\right]\\
&-\left\{\theta\Gamma\left(\frac{1}{r}+1\right)\left[\text{\c{Q}}\left(\frac{1}{r}+1,\left(\frac{(x_{h-1}-\gamma)}{\theta}\right)^r\right)-\text{\c{Q}}\left(\frac{1}{r}+1,\left(\frac{(x_{h-1}+l_h-\gamma)}{\theta}\right)^r\right)\right]\right\}^2.
\end{aligned}
\tag{3.4.9}
$$

With the help of equation (3.4.9), the OSB problem may be expressed for Weibull distribution given as

$$
\begin{aligned}
\text{Minimize}\quad &\frac{\theta^2}{\overline{X}^2}\Gamma\left(\frac{2}{r}+1\right)\sum_{h=1}^{L}\left[\exp\left(-\frac{(x_{h-1}-\gamma)}{\theta}\right)^r-\exp\left(-\frac{(x_{h-1}+l_h-\gamma)}{\theta}\right)^r\right]\times\\
&\left[\text{\c{Q}}\left(\frac{2}{r}+1,\left(\frac{(x_{h-1}-\gamma)}{\theta}\right)^r\right)-\text{\c{Q}}\left(\frac{2}{r}+1,\left(\frac{(x_{h-1}+l_h-\gamma)}{\theta}\right)^r\right)\right]\\
&-\left\{\theta\Gamma\left(\frac{1}{r}+1\right)\left[\text{\c{Q}}\left(\frac{1}{r}+1,\left(\frac{(x_{h-1}-\gamma)}{\theta}\right)^r\right)-\text{\c{Q}}\left(\frac{1}{r}+1,\left(\frac{(x_{h-1}+l_h-\gamma)}{\theta}\right)^r\right)\right]\right\}^2\\
\text{subject to}\quad &C=c_0+\sum_{h=1}^{L}c_hn_h+\sum_{h=1}^{L}t_h\sqrt{n_h}\\
&\sum_{h=1}^{L}l_h=d\\
\text{and}\quad &l_h\geq 0;\ h=1,2,\ldots,L.
\end{aligned}
\tag{3.4.10}
$$

Table 3.4: Exponential distribution with real data

Strata	Proposed Method			Cumulative Method			Geometric Method			LH Kozak Method		
	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV
2	377.719	267 95	0.00043291	310	199 140	0.00043538	99.297	16 263	0.0010732	14.5	360 33	0.00057049
3	216.95 488.92	149 86 63	0.00022775	210 460	126 150 27	0.00032987	46.198 213.428	6 45 166	0.00059169	205.2 465.47	131 83 74	0.00023137
4	185.093 382.081 736.488	141 73 42 26	0.00014855	160 310 510	113 45 38 57	0.00016941	31.511 99.297 312.90	4 10 52 108	0.00042105	140.13 302.8 514.27	83 59 46 54	0.00014648
5	115.687 208.877 174.045 176.35 301.041	73 93 25 19 67	0.00011856	110 260 410 560	68 65 30 16 43	0.00012116	25.048 62.741 157.15 393.642	3 4 17 55 72	0.00032521	114.11 244.24 374.37 556.56	59 50 27 26 47	0.000121401

Table 3.5: Exponential distribution with simulated data

Strata	Proposed Method			Cumulative Method			Geometric Method			LH Kozak Method		
	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV
2	23.2897	454 181	0.00048208	65.43828	694 26	0.00070949	1.303848	2 487	0.00139678	.03	3 486	0.00120285
3	13.9933 32.8972	176 164 150	0.00026155	43.62552 87.25104	663 41 3	0.00059483	.2796179 6.079797	2 27 342	0.00125505	11.76 29.39	203 159 140	0.00026845
4	9.6428 21.7705 42.1427	177 115 102 64	0.00017333	32.71914 65.43828 98.15742	636 51 5 2	0.0004931	.1294893 1.303847 13.12863	3 2 92 212	0.000829234	6.62 17.64 35.26	94 114 96 103	0.00019243
5	8.4737 19.1287 36.2957 84.1627	152 102 95 69 86	0.00016307	26.17531 52.35063 78.52594 104.7013	577 72 12 3 2	0.0003586	.0815897 .5176423 3.284159 20.8362	5 2 8 155 100	0.00068524	5.3 12.35 22.92 40.55	66 54 67 55 69	0.00016535

Table 3.6: Weibull distribution with real data

Strata	Proposed Method			Cumulative Method			Geometric Method			LH Kozak Method		
	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV
2	465.512	291 79	.0021	310	187 148	0.002290	99.297	22 259	0.002806	14.5	2.0035 273.25	0.0044
3	299.728 637.57	185 96 37	.0019	210 460	109 95 76	0.001934	46.198 213.428	5 55 159	0.002544	205.2 465.47	93.9993 102.974 77.5834	0.0028
4	235.158 478.879 731.023	121 89 46 23	.0018	160 310 510	65 59 55 54	0.001865	31.511 99.297 312.901	4 15 67 96	0.002445	140.13 302.8 514.27	52.1677 63.3189 56.9886 54.9677	0.0022
5	194.761 386.268 583.957 784.408	84 76 46 26 18	.0017	110 260 410 560	33 55 43 28 39	0.001821	25.048 62.741 157.154 393.642	2 6 23 53 66	0.002385	114.11 244.24 374.37 556.56	35.0316 47.0952 38.8307 36.2760 39.8578	0.0021

Table 3.7: Weibull distribution with simulated data

Strata	Proposed Method			Cumulative Method			Geometric Method			LH Kozak Method		
	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV	OSB	n_h	CV
2	29.67701	514 143	0.00236	38.44711	703 20	0.002692	25.05165	585.4197 96.7097	0.0024	6.5	124 403	0.0032
3	21.59541 54.25821	390 222 2	0.00190	25.63140 51.26281	616 73 2	0.002125	15.97831 39.27732	428.7157 178.4842 14.6174	0.0020	12.49 22.48	321 129 104	0.0023
4	18.67781 36.88421 78.12591	313 219 37 3	0.00180	19.22355 38.44711 57.67066	503 124 15 2.0087	0.002005	12.76081 25.05165 49.18068	337.6088 153.6615 73.9587 2	0.0018	9.5 15.49 25.98	231 107 90 64	0.0023
5	13.20081 19.73321 32.22961 55.16961	165 132 85 70 2	0.00130	15.37884 30.75768 46.13653 61.51537	420 144 35 3 2	0.00196	11.1503 19.12728 32.81101 56.28416	330.1481 121.5897 59.0490 31.6042 2	0.0014	8.9 12.49 18.49 28.08	246 37 31 71 56	0.0017

3.4.5 Relative efficiency

The relative efficiency (R.E.) of the proposed method with other methods, as discussed, may be computed as

$$\text{R.E.} = \frac{CV_{\text{Other}}}{CV_{\text{Proposed}}} \times 100.$$

Table 3.8: R.E. of proposed method w.r.t other methods for exponential distribution with real and simulated data sets

Strata	R.E. for real data set			R.E. for simulated data set		
	Cumulative	Geometric	LH Kozak	Cumulative	Geometric	LH Kozak
2	100.571	247.904	131.780	147.173	289.740	249.513
3	144.839	259.798	101.589	227.425	479.851	102.638
4	114.042	283.439	98.607	284.489	478.413	111.019
5	102.193	274.299	102.396	219.906	420.212	101.398

Table 3.9: R.E. of proposed method w.r.t other methods for Weibull distribution with real and simulated data sets

Strata	R.E. for real data set			R.E. for simulated data set		
	Cumulative	Geometric	LH Kozak	Cumulative	Geometric	LH Kozak
2	109.047	133.619	209.524	114.068	101.695	135.593
3	101.789	133.895	147.368	111.842	105.263	121.053
4	103.611	135.833	122.222	111.389	100.000	127.778
5	107.118	140.294	123.529	150.769	107.692	130.769

3.5 Conclusion

The proposed method determines the OSB for the population with a nonlinear cost function, which provides optimum sample size for each stratum. Tables 3.4 and 3.6 represent the OSB, sample sizes, and CV values for the real data set, and Tables 3.5 and 3.7 represent the OSB, sample sizes, and CV values for the simulated data set for exponential and Weibull distributions. Figures 3.2 and 3.3 show that if the number of strata increases, then the CVs decrease. The CVs of the estimate, calculated by the proposed method for the distributions, as discussed, are lower than the CVs obtained by other methods. Therefore the precision of the estimates of parameters of the study variables is maximum for the proposed method. Hence, the determination of the OSB may be formulated as an MPP that minimizes the CV of the estimated population parameter. The solution procedure is implemented through MATLAB and R software.

In this chapter the integer restrictions on allocations are also suggested for obtaining the sample size and OSB for each stratum by using the real and simulated data sets and to explore the application of the proposed method for the nonlinear cost function. This chapter also presents the results obtained by the cumulative method, geometric method, and Lavallée and Hindiroglou's method to carry out a comparative study and this study helps to promote the application of the proposed technique. The R.E. of the proposed method over other methods are presented in Tables 3.8 and 3.9 where one can see the substantial gains by the proposed method over all other three methods in both real and simulated data sets for both distributions.

Figure 3.2: CV for exponential distribution for real and simulated data sets

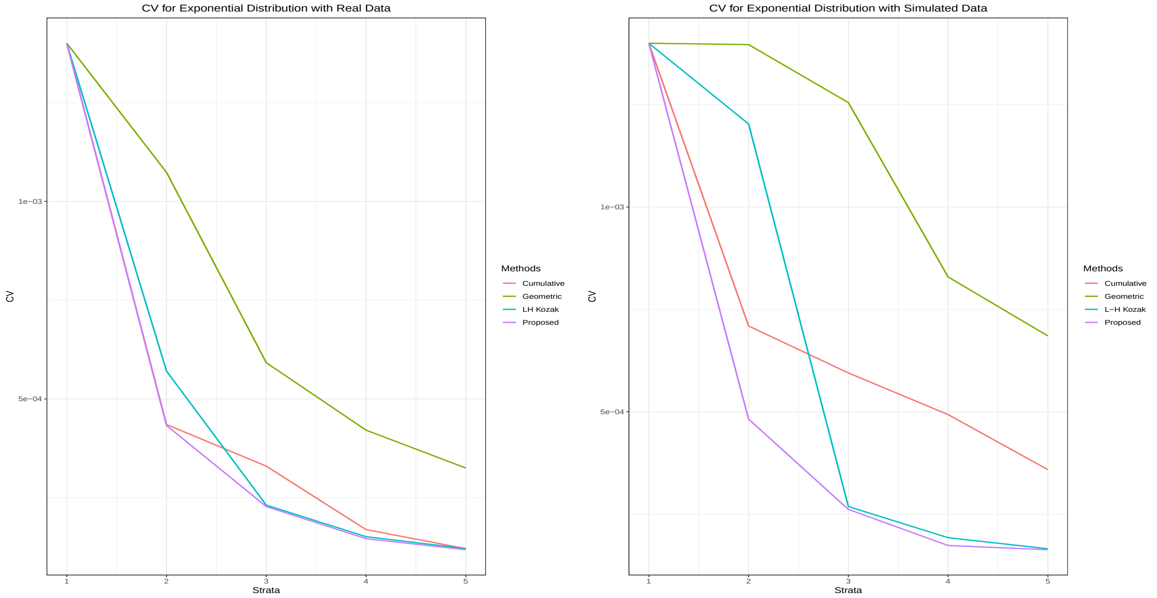
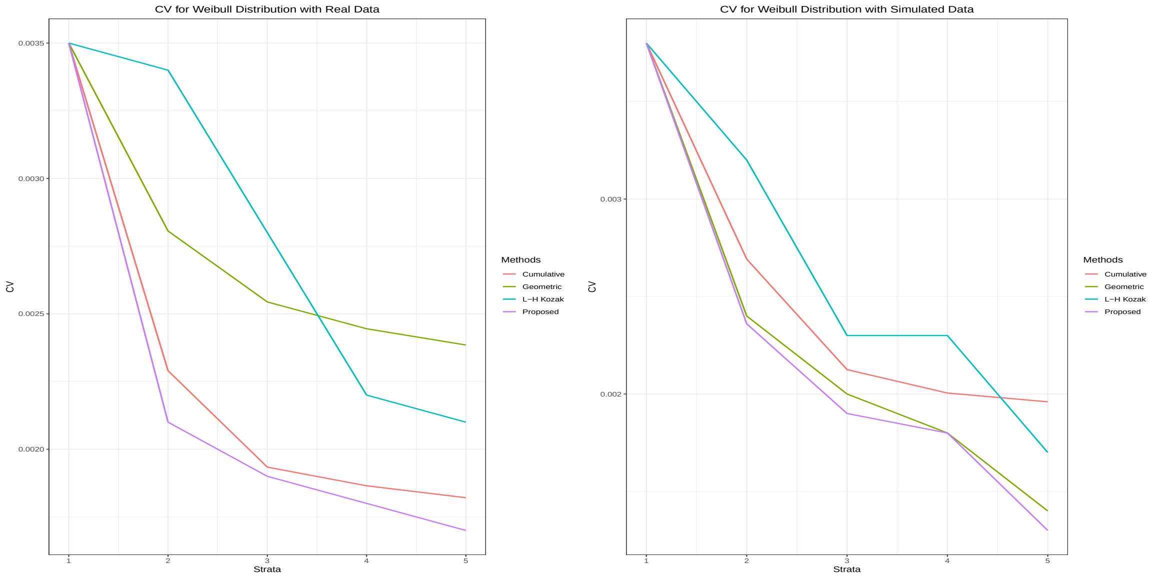


Figure 3.3: CV for Weibull distribution for real and simulated data sets



CHAPTER 4

On multivariate-multiobjective stratified sampling design under probabilistic environment: a fuzzy programming technique

4.1 Introduction

In many real-life practices, the populations may vary in their accessibility. Some parts of the population may be in remote locations, gated buildings or other inaccessible areas. Then the choice of sampling design affects the results of the survey. For such situations, stratified sampling seems to be the best choice of the sampling technique. For obtaining detailed information about the characteristics of the population, a multivariate stratified sample survey is carried out by splitting the population into L strata. It is assumed that all the p characteristics are defined in each unit of the population. The estimation of unknown population means of p characteristics is required and may be carried out using the NLPP. Cochran (1977) has been shown that the individual optimum allocation of one character may not remain optimum for others. A compromise criterion may be required to obtain the best allocation, which helps in getting precise information about population parameters. Therefore the allocation based on some compromised criterion is called compromise allocation in multivariate stratified sampling design.

Most of the authors (Neyman (1934), Kokan and Khan (1967), Chatterjee (1968), Ahsan (1975-1976), Khan *et al.* (1997), Semiz (2004), Kozak (2006), Varshney *et al.* (2012, 2015), Fatima *et al.* (2014), Muhammad *et al.* (2015), Muhammad and Husain (2017)) discussed the problems of allocation and worked out compromise allocation in multivariate stratified sample surveys. A compromise allocation is obtained either by suggesting different compromise criteria or using the suggested criteria under different conditions, i.e. in the availability of auxiliary information, presence of nonresponse, etc. In many sampling designs, the stratum variances are not known in advance but may be estimable. From the deterministic point of view, such problems may be formulated as NLPPs. However, if the nature of the estimated variances is also considered, it will be an additional restriction to the problem, and therefore the compromise allocation may not be obtained easily. For such situations, the Stochastic Nonlinear Programming Problem (SNLPP) may help to work out the required compromise allocation to obtain sufficient information about population parameters (See (Charnes and Cooper (1963), Prékopa (1978), Díaz-García and Garay Tapia (2007), Kozak and Wang (2010), Haseen *et al.* (2016))). Díaz-García and Ramos-Quiroga (2014) discussed and provided results by solving SNLPPs with a fixed linear cost function. The concept of fuzzy set theory has been discussed by Zadeh (1965) and Bellman and Zadeh (1970). They utilized the fuzzy approach for dynamic issues. The idea of the fuzzy set was given by Zimmermann (1978) to convert the multiobjective linear programming problem into a single objective linear programming problem. Some authors discussed fuzzy programming methods (See (Gupta *et al.* (2013), Ali and Hasan (2014), Varshney *et al.* (2017), Haq *et al.* (2020)). Fuzzy programming is one of many available optimization models that deals with an optimization problem under uncertainty. The technique is flexible and thus helps decision-makers have a better understanding of their problems. This model can be applied when situations are not clearly defined and thus have uncertainty. In other words, if an exact value is not critical to the problem. Fuzzy programming has been studied and applied recently by several authors in different areas (Elsisi (2019a,b, 2020), Fakhrzad and Goodarzian (2019), Fathollahi-Fard *et al.* (2020), Lu *et al.* (2020), Elsisi and Soliman (2021), Goodarzian and Hosseini-Nasab (2021)).

Under the probabilistic environment, the use of the nonlinear cost function is proposed by considering the labour cost as part of the survey's total cost. In this chapter, the compromise criterion is suggested for determining the compromise allocation for a multiobjective-multivariate stochastic nonlinear programming problem for the fixed cost in the probabilistic situation. The solution procedure is also given to solve the formulated problem by using an appropriate nonlinear programming technique.

In Section 4.2, the notations and formulation of the problem are given. The formulated problems' solutions are suggested using two deterministic approaches: modified E-model and chance constraints in Section 4.3. In the modified E-model formulation, the solution strategy is recommended by utilizing a fuzzy goal programming technique. For the chance constraints model, the solutions are obtained using Lagrange multiplier and integer nonlinear programming techniques. Section 4.4, the numerical illustration is discussed by considering the Iris data set, and the data set is obtained by simulation carried out by the software R. The formulated problem is solved through the modified E-model approach and chance constraints technique. The solution procedures are suggested by utilizing fuzzy goal programming problem, Lagrange multiplier method and integer nonlinear programming problem. MATLAB software is used to solve the formulated NLPPs. A comparative study is included by considering some other allocations, as discussed in Section 4.4.1.1, with the proposed allocation. Finally, the conclusion has been made for using the proposed technique in Section 4.5.

4.2 Framework of the problem

Assuming a population of N units that is partitioned into strata of sizes $N_1, N_2, \ldots, N_L$ units such that $\sum_{h=1}^{L} N_h = N$.

For h^{th} stratum, the following notations are introduced as follows

$$\begin{aligned}
N_h &: \text{ Stratum size} \\
W_h = \frac{N_h}{N} &: \text{ Stratum weight} \\
n_h &: \text{ Sample size} \\
y_{hi} &: \text{ Observational value of } i^{th} \text{ stratum unit/stratum sample} \\
\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi} &: \text{ Stratum mean} \\
\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi} &: \text{ Sample mean} \\
S_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(y_{hi} - \overline{Y}_h\right)^2 &: \text{ Stratum mean square} \\
s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(y_{hi} - \overline{y}_h\right)^2 &: \text{ Sample mean square.}
\end{aligned}$$

Furthermore,

$\overline{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi} = \frac{1}{N}\sum_{h=1}^{L} N_h \overline{Y}_h = \sum_{h=1}^{L} W_h \overline{Y}_h$: Describes the overall population mean.

If the estimated value of $\overline{Y}$ is needed, then the stratified sample mean

$$\overline{y}_{st} = \sum_{h=1}^{L} W_h \overline{y}_h,$$

gives an unbiased estimate for $\overline{Y}$ with the sampling variance

$$V(\overline{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h S_h^2}{N}.$$

In a multivariate stratified population where p characteristics are given on each population element, then the p population means $\overline{Y}_j; j = 1,2,\ldots,p$ are to be estimated. Since the individual optimum allocation may not be optimum for other characteristics. Let y_{jhi} denotes the value obtained from i^{th} element in h^{th} stratum having j^{th} characteristic and $\overline{Y}_{jh} = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{jhi}$ be the stratum mean of y_{jhi}. Then the sample means for all characteristics in h^{th} stratum are calculated by

$$\overline{y}_{jh} = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{jhi}; \quad j = 1,2,\ldots,p.$$

For j^{th} characteristic, an unbiased estimate of the overall population mean $\overline{Y}_j$ is given by $\overline{y}_{jst}$ and is expressed by

$$\overline{y}_{jst} = \textstyle\sum_{h=1}^{L} W_h \overline{y}_{jh},$$

with its sampling variance

$$V(\overline{y}_{jst}) = \textstyle\sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h S_{jh}^2}{N},$$

where S_{jh}^2 is the stratum variance of j^{th} characteristics in h^{th} stratum for $j = 1,2,\ldots,p$, $h = 1,2,\ldots,L$ and can be calculated by

$$S_{jh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(y_{jhi} - \overline{Y}_{jh}\right)^2.$$

For a multivariate stratified sample survey, the linear cost function may be considered for the overall budget of the survey (Cochran (1977)) and may be expressed as

$$C = c_0 + \sum_{h=1}^{L} c_h n_h,$$

or

$$C - c_0 = C_0 = \sum_{h=1}^{L} c_h n_h,$$

where C_0 denotes the cost to measure all sampling units in all strata, c_h is the per-unit measurement cost of measuring p characteristics on the selected unit in h^{th} stratum, n_h is the h^{th} stratum sample size and c_0 is the overhead cost to conduct the survey. If the travel cost within the stratum come into consideration, then the cost function may not have remained linear. Beardwood *et al.* (1959) suggested the nonlinear cost function for this case. They showed that the distance between n arbitrarily dispersed points is proportional to $\sqrt{n}$. The nonlinear cost function, which includes travel costs, may be expressed as

$$C = c_0 + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h},$$

or

$$C_0 = C - c_0 = \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h}, \tag{4.2.1}$$

where $t_h\sqrt{n_h}$ is the travel cost incurred for h^{th} stratum. The cost function defined by (4.2.1) is quadratic in $\sqrt{n_h}$.

Practically, some other cost factors may be considered, like costs on the reward to respondents, labour costs, etc. Whenever the interviewers want to collect detailed information from the selected respondents, there will be a requirement for more human resources that may be available for a specified time. For these prerequisites, the labour costs may be used for conducting the survey, and therefore the cost function may be expressed as

$$C_0 = C - c_0 = \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} + \omega \sum_{h=1}^{L} E\left(T_h\right), \tag{4.2.2}$$

where ω indicates the cost of labour for a unit time and $\sum_{h=1}^{L} E\left(T_h\right)$ indicates the accumulated labour time to obtain information from all strata. Since the labour time is available with respect to time for sampling units within a stratum and follows an exponential distribution with rate λ and the pdf for time may be given as $f(x) = \lambda e^{-\lambda x}$; $x > 0$ and value of $\lambda = 1/(\text{average time})$. To approach n_h units in h^{th} stratum, the labour time has Gamma distribution with parameters (n_h, λ). Subsequently, the distribution of labour time to measure all sampling units within the stratum follows Gamma distribution with $(\sum_{h=1}^{L} n_h, \lambda)$ (Ross (2009)). Hence the expected labour time for all strata may be computed as for various values of λ (Muhammad and Husain (2017)).

$$\begin{aligned}
\sum_{h=1}^{L} E\left(T_h\right) &= \sum_{h=1}^{L}\left(\int_0^{\infty} t\lambda e^{-\lambda t}\frac{(\lambda t)^{n_h-1}}{(n_h-1)!}dt\right)\\
&= \sum_{h=1}^{L}\left(\frac{1}{(n_h-1)!}\int_0^{\infty} t\lambda e^{-\lambda t}(\lambda t)^{n_h-1}dt\right)\\
&= \sum_{h=1}^{L}\frac{n_h}{\lambda}.
\end{aligned}$$

The problem may be formulated in two ways by using a deterministic approach, and the solution is obtained either by minimizing the variance $V(\bar{y}_{jst})$ for a fixed cost or by minimizing the cost of the survey with variances of specified limits. Therefore two optimization problems may be described as

$$\begin{aligned}
&\text{Minimize} && V(\bar{y}_{jst});\ j=1,2,\ldots,p \ \text{ simultanously}\\
&\text{subject to} && \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h\sqrt{n_h} + \omega\sum_{h=1}^{L}\frac{n_h}{\lambda} \le C_0\\
& && 2 \le n_h \le N_h\\
&\text{and} && n_h \text{ are integers; } h=1,2,\ldots,L.
\end{aligned} \tag{4.2.3}$$

and

$$\begin{aligned}
&\text{Minimize} && \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h\sqrt{n_h} + \omega\sum_{h=1}^{L}\frac{n_h}{\lambda}\\
&\text{subject to} && V(\bar{y}_{jst}) \le V_0^j, \quad j=1,2,\ldots,p\\
& && 2 \le n_h \le N_h\\
&\text{and} && n_h \text{ are integers; } h=1,2,\ldots,L,
\end{aligned} \tag{4.2.4}$$

respectively.

If the true values of S_{jh}^2 are unknown, then they may be computed through a starter test or the values of past events if they are given (see Kozak (2006)).

4.3 Determination of identical probabilistic sampling variances and cost

If S_{jh}^2 are considered as random variables, then the problems defined in (4.2.3) and (4.2.4) become SNLPPs. These problems may be converted into their equivalent deterministic problems. Several techniques, like modified E-model, E-model, V-model, chance constraints, etc., are available to solve the deterministic problems (see Charnes and Cooper (1963)). In this chapter, modified E-model and chance constraints methods are used to convert the problems into the deterministic form.

4.3.1 Determination of probabilistic sampling variance through modified E-model

Consider the following formulated SNLPP for j^{th} characteristic which is given as

$$\begin{array}{ll} \text{Minimize} & V(\bar{y}_{jst}) \\ \text{subject to} & \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} + \omega \sum_{h=1}^{L} \frac{n_h}{\lambda} \le C_0 \\ & 2 \le n_h \le N_h \\ \text{and} & n_h \text{ are integers; } h = 1,2,\ldots,L,\ j = 1,2,\ldots,p, \end{array} \tag{4.3.1}$$

where

$$V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h S_{jh}^2}{N},$$

and S_{jh}^2 are random variables.

By considering the limiting distribution of S_{jh}^2 (see Melaku (1986), Díaz-García and Garay Tapia (2007)), let us define a random variable ζ_h which has an asymptotic normal distribution with mean $E(\zeta_h)$ and variance $V(\zeta_h)$.

For j^{th} characteristic, ζ_{jh} is defined for multivariate case and is given as

$$\zeta_{jh} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(y_{jhi} - \bar{Y}_{jh}\right)^2,$$

where y_{jhi} denotes the value of the i^{th} unit in h^{th} stratum for j^{th} characteristic and $\overline{Y}_{jh} = N_h^{-1}\sum_{i=1}^{N_h} y_{jhi}$ is stratum mean for h^{th} stratum. The random variable ζ_{jh} has an asymptotic normal distribution with mean $E(\zeta_{jh})$ and variance $V(\zeta_{jh})$. These are given as

$$E(\zeta_{jh}) = \frac{n_h}{n_h - 1} S_{jh}^2$$

and

$$V(\zeta_{jh}) = \frac{n_h}{(n_h - 1)^2}\left[C_{yjh}^4 - (S_{jh}^2)^2\right],$$

respectively, where C_{yjh}^4 is the fourth mean moment and can be calculated by the following expression

$$C_{yjh}^4 = \frac{1}{N_h}\sum_{i=1}^{N_h}\left(y_{jhi} - \overline{Y}_{jh}\right)^4 \text{ ; } h = 1,2,\ldots,L,\ j = 1,2,\ldots,p.$$

Let us define s_{jh}^2 which may be given as

$$s_{jh}^2 = \zeta_h - \frac{n_h}{n_h - 1}\left(\overline{y}_{jh} - \overline{Y}_{jh}\right)^2,$$

where

$$\frac{n_h}{n_h - 1} \to 1$$

and

$$\left(\overline{y}_{jh} - \overline{Y}_{jh}\right)^2 \to 0 \text{ in probability form.}$$

Then, the sample variance s_{jh}^2 has normal asymptotical distribution

$$s_{jh}^2 \xrightarrow{a} N(E(\zeta_{jh}), V(\zeta_{jh})).$$

It has also seen that the objective function in (4.3.1) is a linear function of s_{jh}^2. Therefore the objective function also follows normal distribution with mean and variance, which are given as

$$\begin{aligned} E(\widehat{V}(\overline{y}_{jst})) &= E\left(\sum_{h=1}^{L} \frac{W_h^2 s_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_{jh}^2}{N}\right) \\ &= \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h - 1} - \sum_{h=1}^{L} \frac{W_h}{N}\left(\frac{n_h}{n_h - 1}\right) S_{jh}^2 \end{aligned} \tag{4.3.2}$$

and

$$\begin{aligned} &V(\widehat{V}(\overline{y}_{jst})) \\ &\quad = V\left(\sum_{h=1}^{L} \frac{W_h^2 s_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_{jh}^2}{N}\right) \\ &\quad = \sum_{h=1}^{L} \frac{W_h^4}{n_h (n_h - 1)^2}\left(C_{yjh}^4 - (S_{jh}^2)^2\right) - \sum_{h=1}^{L} \frac{W_h^2}{N^2}\left[\frac{n_h}{(n_h - 1)^2}\left(C_{yjh}^4 - (S_{jh}^2)^2\right)\right], \end{aligned} \tag{4.3.3}$$

respectively.

Therefore, by using the modified E-model technique, the objective function may be redefined as

$$f_j(n_h) = k_1 E\left(\widehat{V}(\overline{y}_{jst})\right) + k_2\sqrt{V(\widehat{V}(\overline{y}_{jst}))}\ ,$$

where k_1 and k_2 are non-negative constants such that $k_1 + k_2 = 1$. The values of k_1 and k_2 will show the existence of the expectation and variance of $\widehat{V}(\overline{y}_{jst})$. Hence, the equivalent deterministic NLPP to the SNLPP for j^{th} characteristic given in (4.3.1), may be formulated as

$$\begin{aligned} &\text{Minimize} && f_j(n_h) \\ &\text{subject to} && \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} + \omega \sum_{h=1}^{L} \frac{n_h}{\lambda} \le C_0 \\ & && 2 \le n_h \le N_h \\ &\text{and} && n_h \text{ are integers; } h = 1,2,\ldots,L,\ j = 1,2,\ldots,p, \end{aligned} \tag{4.3.4}$$

where

$$\begin{aligned} f_j(n_h) =& k_1\left[\sum_{h=1}^{L}\frac{W_h^2S_{jh}^2}{n_h-1}-\sum_{h=1}^{L}\frac{W_hS_{jh}^2}{N}\left(\frac{n_h}{n_h-1}\right)\right] \\ &+k_2\left[\sum_{h=1}^{L}\frac{W_h^4}{n_h\left(n_h-1\right)^2}\left(C_{yjh}^4-(S_{jh}^2)^2\right)-\sum_{h=1}^{L}\frac{W_h^2}{N^2}\left(\frac{n_h}{(n_h-1)^2}\left(C_{yjh}^4-(S_{jh}^2)^2\right)\right)\right]^{1/2}. \end{aligned}$$

Since the objective function includes the values of the population variance S_{jh}^2, but these values are unknown in general, in that case, the sample variance s_{jh}^2 may be used. Therefore, the equivalent deterministic NLPP defined in (4.3.4) may be given as

$$\begin{aligned} \text{Minimize} \quad & f_j(n_h) = k_1\left[\sum_{h=1}^{L}\frac{W_h^2s_{jh}^2}{n_h-1}-\sum_{h=1}^{L}\frac{W_hs_{jh}^2}{N}\left(\frac{n_h}{n_h-1}\right)\right] \\ & +k_2\left[\sum_{h=1}^{L}\frac{W_h^4}{n_h\left(n_h-1\right)^2}\left(C_{yjh}^4-(s_{jh}^2)^2\right)-\sum_{h=1}^{L}\frac{W_h^2}{N^2}\left(\frac{n_h}{(n_h-1)^2}\left(C_{yjh}^4-(s_{jh}^2)^2\right)\right)\right]^{1/2} \\ \text{subject to} \quad & \sum_{h=1}^{L}c_hn_h+\sum_{h=1}^{L}t_h\sqrt{n_h}+\omega\sum_{h=1}^{L}\frac{n_h}{\lambda}\le C_0 \\ & 2\le n_h\le N_h \\ \text{and} \quad & n_h \text{ are integers; } h=1,2,\ldots,L,\ j=1,2,\ldots,p. \end{aligned} \tag{4.3.5}$$

The NLPP given in (4.3.5) may be extended as multiobjective - INLPP (MINLPP) for multivariate stratified sampling designs as given as

$$\begin{aligned} \text{Minimize} \quad & [f_1(n_h),\ f_2(n_h),\ldots,f_p(n_h)] \\ \text{subject to} \quad & \sum_{h=1}^{L}c_hn_h+\sum_{h=1}^{L}t_h\sqrt{n_h}+\omega\sum_{h=1}^{L}\frac{n_h}{\lambda}\le C_0 \\ & 2\le n_h\le N_h \\ \text{and} \quad & n_h \text{ are integers; } h=1,2,\ldots,L. \end{aligned} \tag{4.3.6}$$

4.3.1.1 Solution procedure by using the fuzzy goal programming technique

To solve the MINLPP given in (4.3.6), the fuzzy goal programming technique may be applied for multivariate sampling design. Since no technique is developed to solve multiobjective formulation of INLPP, in that case, the problem may be converted into a single objective problem by using suitable criterion. For such a case, the fuzzy goal programming technique may be used and applied by using following steps.

Stage 1: To get the solution of the MINLPP, a problem of single objective function is to be required by ignoring the remaining objective functions of other characteristics to work out the optimum solution for each characteristic as an ideal solution.

Stage 2: The step-1, is repeated for all characteristics, and p - optimum solution are obtained to give the optimum values of objective functions $(f_1,\ f_2, \ldots, f_p)$.

Stage 3: To compute the payoff matrix, the ideal solutions will give the upper and lower values for each objective function by defining U_j and L_j for j^{th} objective function; $j = 1, 2, \ldots, p$.

These values are computed as

$$U_j = Max\left\{f_1(n^*_{1h}),\ f_2(n^*_{2h}), \ldots, f_p(n^*_{ph})\right\}$$

and

$$L_j = Min\left\{f_1(n^*_{1h}),\ f_2(n^*_{2h}), \ldots, f_p(n^*_{ph})\right\},$$

where $f_j(n^*_{jh})$ is the optimum value of the objective function for the j^{th} characteristic with optimum allocation n^*_{jh}.

Stage 4: The membership function may be defined as

$$\mu_j(n_{jh}) = \begin{cases} 0 & \text{if } f_j(n_{jh}) \geq U_j \\ \frac{U_j(n_{jh}) - f_j(n_{jh})}{U_j(n_{jh}) - L_j(n_{jh})} & \text{if } L_j \leq f_j(n_{jh}) \leq U_j \\ 1 & \text{if } f_j(n_{jh}) \leq L_j \end{cases},$$

where $\mu_j(n_{jh})$ is a strictly monotonic decreasing function with respect to the solution $n_{jh},\ h = 1, 2, \ldots, L$.

Consider the variable ξ_j which is defined as

$$\xi_j = \frac{U_j(n_{jh}) - f_j(n_{jh})}{U_j(n_{jh}) - L_j(n_{jh})}.$$

Stage 5: By max-min method, we have $Max\,[Min(\xi_1, \xi_2, \ldots, \xi_p)]$, then

$$\begin{aligned} \text{Maximize} \quad & \xi \\ \text{subject to} \quad & \xi_1 \geq \xi \\ & \xi_2 \geq \xi \\ & \vdots \\ & \xi_p \geq \xi, \end{aligned}$$

where $\xi = \underset{j}{Min}\left\{\mu_j(n_{(jh)});\ j = 1, 2, \ldots, p\right\}$.

Finally, the mathematical programming formulation for the problem (4.3.6) is to be solved by using fuzzy goal programming as follows

$$\begin{aligned} \text{Maximize} \quad & \xi \\ \text{subject to} \quad & f_1 - \xi(U_1 - L_1) \geq L_1 \\ & f_2 - \xi(U_2 - L_2) \geq L_2 \\ & \vdots \\ & f_p - \xi(U_p - L_p) \geq L_p \\ & \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} + \omega \sum_{h=1}^{L} \frac{n_h}{\lambda} \leq C_0 \\ \text{and} \quad & \xi \in [0, 1],\ n_h \text{ are integers};\ h = 1, 2, \ldots, L. \end{aligned} \tag{4.3.7}$$

4.3.2 Determination of probabilistic sampling cost through chance constraints

In this section, the SNLPP is considered for minimizing the total cost of survey for a given bound to the estimated variance of the mean. This bound may be specified with tolerance limits for estimated variances of the estimates. This SNLPP may be formulated as

$$
\begin{aligned}
&\text{Minimize} && \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} + \omega \sum_{h=1}^{L} \frac{n_h}{\lambda} \\
&\text{subject to} && P\left[\widehat{V}(\overline{y}_{jst}) \leq V_0^j\right] \geq p_0\ ;\ \ j = 1,2,\ldots,p \\
& && 2 \leq n_h \leq N_h \\
&\text{and} && n_h \text{ are integers; } h = 1,2,\ldots,L,
\end{aligned}
\tag{4.3.8}
$$

where

$$
\widehat{V}(\overline{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 s_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_{jh}^2}{N}.
$$

Also $V_0^j \geq 0$ and p_0 is a predetermined probability such that $0 \leq p_0 \leq 1$.

Since s_{jh}^2 follows an asymptotic normal distribution with mean $E(\zeta_{jh})$ and variance $V(\zeta_{jh})$. Then the estimated $\widehat{V}(\overline{y}_{jst})$ in (4.3.8) also follows asymptotic normal distribution with the mean and variance defined in (4.3.2) and (4.3.3), respectively. After standardizing the function of the $\widehat{V}(\overline{y}_{jst})$ in (4.3.8), it may be re-expressed as

$$
P\left[\frac{\widehat{V}(\overline{y}_{jst}) - E\{\widehat{V}(\overline{y}_{jst})\}}{\sqrt{V\{\widehat{V}(\overline{y}_{jst})\}}} \leq \frac{V_0^j - E\{\widehat{V}(\overline{y}_{jst})\}}{\sqrt{V\{\widehat{V}(\overline{y}_{jst})\}}}\right] \geq p_0,
$$

where

$$
p_0 = \phi\left[\frac{V_0^j - E\{\widehat{V}(\overline{y}_{jst})\}}{\sqrt{V\{\widehat{V}(\overline{y}_{jst})\}}}\right],
$$

and $\phi(\cdot)$ represents the function of standard normal distribution. If e denotes the value of random variable that follows standard normal distribution such that $\phi(e) = p_0$, with these conditions, the inequality may be expressed as

$$
\phi\left[\frac{V_0^j - E\{\widehat{V}(\overline{y}_{jst})\}}{\sqrt{V\{\widehat{V}(\overline{y}_{jst})\}}}\right] \geq \phi(e).
$$

Therefore,

$$E\left\{\widehat{V}(\overline{y}_{jst})\right\}+e\sqrt{V\left\{\widehat{V}(\overline{y}_{jst})\right\}}-V_0^j\leq 0.$$

The equivalent deterministic NLPP for SNLPP in (4.3.8) may be given as

$$\begin{array}{ll}
\text{Minimize} & \sum_{h=1}^{L}c_hn_h+\sum_{h=1}^{L}t_h\sqrt{n_h}+\omega\sum_{h=1}^{L}\frac{n_h}{\lambda} \\
\text{subject to} & E\left\{\widehat{V}(\overline{y}_{jst})\right\}+e\sqrt{V\left\{\widehat{V}(\overline{y}_{jst})\right\}}-V_0^j\leq 0\ ;\ \ j=1,2,\ldots,p \\
 & 2\leq n_h\leq N_h \\
\text{and} & n_h \text{ are integers; } h=1,2,\ldots,L,
\end{array} \tag{4.3.9}$$

where

$$\begin{aligned}
E\left\{\widehat{V}(\overline{y}_{jst})\right\}+e\sqrt{V\left\{\widehat{V}(\overline{y}_{jst})\right\}}&=\left[\sum_{h=1}^{L}\frac{W_h^2S_{jh}^2}{n_h-1}-\sum_{h=1}^{L}\frac{W_hS_{jh}^2}{N}\left(\frac{n_h}{n_h-1}\right)\right] \\
+e\left[\sum_{h=1}^{L}\frac{W_h^4}{n_h\,(n_h-1)^2}(C_{yjh}^4-(S_{jh}^2)^2)\right.&\left.-\sum_{h=1}^{L}\frac{W_h^2}{N^2}\left(\frac{n_h}{(n_h-1)^2}(C_{yjh}^4-(S_{jh}^2)^2)\right)\right]^{1/2}.
\end{aligned} \tag{4.3.10}$$

The expression in (4.3.10) population variances S_{jh}^2 and these values remain not known in advance. Then s_{jh}^2 may be substituted in place of S_{jh}^2. Hence the equivalent deterministic NLPP for (4.3.8) may be given as

$$\begin{array}{ll}
\text{Minimize} & \sum_{h=1}^{L}c_hn_h+\sum_{h=1}^{L}t_h\sqrt{n_h}+\omega\sum_{h=1}^{L}\frac{n_h}{\lambda} \\
\text{subject to} & E\left\{\widehat{V}(\overline{y}_{jst})\right\}+e\sqrt{V\left\{\widehat{V}(\overline{y}_{jst})\right\}}-V_0^j\leq 0\ ;\ \ j=1,2,\ldots,p \\
 & 2\leq n_h\leq N_h \\
\text{and} & n_h \text{ integers; } h=1,2,\ldots,L,
\end{array} \tag{4.3.11}$$

where

$$\begin{aligned}
E\left\{\widehat{V}(\overline{y}_{jst})\right\}+e\sqrt{V\left\{\widehat{V}(\overline{y}_{jst})\right\}}&=\left[\sum_{h=1}^{L}\frac{W_h^2s_{jh}^2}{n_h-1}-\sum_{h=1}^{L}\frac{W_hs_{jh}^2}{N}\left(\frac{n_h}{n_h-1}\right)\right] \\
+e\left[\sum_{h=1}^{L}\frac{W_h^4}{n_h\,(n_h-1)^2}(C_{yjh}^4-(s_{jh}^2)^2)\right.&\left.-\sum_{h=1}^{L}\frac{W_h^2}{N^2}\left(\frac{n_h}{(n_h-1)^2}(C_{yjh}^4-(s_{jh}^2)^2)\right)\right]^{1/2}.
\end{aligned} \tag{4.3.12}$$

4.4 Application

A population of size $N = 9500$, with three strata and two characteristics are taken and obtained by carrying out a simulation of 150 observations of Iris data. Iris data set is available in the domain of Software R. These observations are divided into three strata where two characteristics (that is, length and width of a leaf of a particular species of flower) are measured on each unit of the population. The population units for three strata of sizes $3000, 3000$ and 3500, respectively are generated by the simulation of Iris data using the software R and the values of s^2_{jh} and C^4_{yjh} are computed and reported in Table 4.1. The values of c_h, t_h, ω and λ are assumed for numerical illustration accordingly and given in Table 4.1. The total cost for conducting the survey is taken as 1000 units.

Table 4.1: Three strata information with two characteristics

h	N_h	s^2_{1h}	s^2_{2h}	C^4_{y1h}	C^4_{y2h}	c_h	t_h	ω	λ
1	3000	0.01523817	0.02037975	0.00068061	0.001217395	2	1	100	20
2	3000	0.06898021	0.00957083	0.01434982	0.000268493	3	2	100	20
3	3500	0.16608490	0.01080517	0.08426042	0.000349415	3	3	100	20

4.4.1 Solution for modified E-model by fuzzy goal programming technique

Without loss of generality, $k_1 = k_2 = 0.5$ is taken. For the given numeric values, given in Table 4.1, the formulations of the NLPP for both characteristics are given as

For characteristic $j=1$

$$\begin{aligned}
\text{Minimize } f_1 &= 0.5\left[\frac{0.00151959}{(n_1-1)}+\frac{0.00687891}{(n_2-1)}+\frac{0.02254338}{(n_3-1)}-\frac{0.00481210n_1}{9500(n_1-1)}-\frac{0.02178320n_2}{9500(n_2-1)}-\frac{0.06118920n_3}{9500(n_3-1)}\right]\\
&\quad+0.5\left[\frac{0.000004459}{(n_1-1)^2n_1}+\frac{0.00009538}{(n_2-1)^2n_2}+\frac{0.00104419}{(n_3-1)^2n_3}-\frac{0.00004472n_1}{9500^2(n_1-1)^2}\right.\\
&\quad\left.-\frac{0.00095649n_2}{9500^2(n_2-1)^2}-\frac{0.00769289n_3}{9500^2(n_3-1)^2}\right]^{\frac{1}{2}}\\
\text{subject to}\quad & 2n_1+3n_2+3n_3+\sqrt{n_1}+2\sqrt{n_2}+3\sqrt{n_3}+5\,(n_1+n_2+n_3)\leq 1000\\
& 2\leq n_1\leq N_1\\
& 2\leq n_2\leq N_2\\
& 2\leq n_3\leq N_3\\
\text{and}\quad & n_1,\ n_2,\ n_3 \text{ are integers.}
\end{aligned} \tag{4.4.1}$$

For characteristic $j = 2$

$$\begin{aligned}
\text{Minimize } f_2 \quad &= 0.5\left[\frac{0.00203233}{(n_1-1)}+\frac{0.00095443}{(n_2-1)}+\frac{0.00146663}{(n_3-1)}-\frac{0.00643571n_1}{9500(n_1-1)}-\frac{0.00302237n_2}{9500(n_2-1)}-\frac{0.00398085n_3}{9500(n_3-1)}\right]\\
&\quad +0.5\Big[\frac{(0.00000797)}{(n_1-1)^2n_1}+\frac{0.00000176}{(n_2-1)^2n_2}+\frac{0.0000042865}{(n_3-1)^2n_3}-\frac{0.00007998n_1}{9500^2(n_1-1)^2}\\
&\quad -\frac{0.00001764n_2}{9500^2(n_2-1)^2}-\frac{0.00003158n_3}{9500^2(n_3-1)^2}\Big]^{\frac{1}{2}}\\
\text{subject to} \quad & 2n_1+3n_2+3n_3+\sqrt{n_1}+2\sqrt{n_2}+3\sqrt{n_3}+5\,(n_1+n_2+n_3)\le 1000 \qquad (4.4.2)\\
& 2\le n_1\le N_1\\
& 2\le n_2\le N_2\\
& 2\le n_3\le N_3\\
\text{and} \quad & n_1,\ n_2,\ n_3 \text{ are integers.}
\end{aligned}$$

The ideal solution for each problem is obtained by solving independently by using for both characteristics are worked out as given below

$$n_{11} = 18,\ n_{12} = 37,\ n_{13} = 67,\ \text{with } f_1 = 0.00034712,$$

$$n_{21} = 54,\ n_{22} = 33,\ n_{23} = 40,\ \text{with } f_2 = 0.00005878.$$

After getting ideal solutions, the pay off matrix may be computed and given in Table 4.2.

Table 4.2: Payoff matrix using the ideal solutions

	f_1	f_2
n^*_{1h}	0.00034712	0.00010331
n^*_{2h}	0.00047683	0.00005878

The upper and lower bounds of each objective function may be given as

$$f_1^l = 0.00034712,\ f_1^u = 0.00047683,$$

$$f_2^l = 0.00005878,\ f_2^u = 0.00010331.$$

Therefore, the values of L_j and U_j are obtained as

$$L_j = \underset{j}{Min}\ f_j(n^*_{jh}) = 0.000058780 \text{ and } U_j = \underset{j}{Max}\ f_j(n^*_{jh}) = 0.00047683.$$

Let $\mu_1(n_{1h})$ and $\mu_2(n_{2h})$ be the fuzzy membership function of the objective functions $f_j(n_{jh})$; $j = 1,2$ and they are used for developing a membership function for both characteristics as

$$\mu_1(n_{1h}) = \begin{cases} 0 & \text{if } f_1(n_{1h}) \geq 0.00047683 \\ \frac{0.00047683 - f_1(n_{1h})}{0.00013} & \text{if } 0.00034712 \leq f_1(n_{1h}) \leq 0.00047683 \\ 1 & \text{if } f_1(n_{1h}) \leq 0.00034712 \end{cases}$$

$$\mu_2(n_{2h}) = \begin{cases} 0 & \text{if } f_2(n_{2h}) \geq 0.00010331 \\ \frac{0.00010331 - f_2(n_{2h})}{0.00004453} & \text{if } 0.000058780 \leq f_2(n_{2h}) \leq 0.00010331 \\ 1 & \text{if } f_2(n_{2h}) \leq 0.000058780 \end{cases} .$$

By using the max-min addition operator, the objective function is revised as

$$\text{Maximize } \left\{7.274210973 - \left(\frac{f_1(n_{1h})}{0.00013} + \frac{f_2(n_{2h})}{0.00004453}\right)\right\}.$$

So as to maximize the above problem with subject to constraints as formulated as

$$\begin{aligned} \text{Maximize} \quad & \xi \\ \text{subject to} \quad & f_1 - .00012971\xi \geq 0.00034712 \\ & f_2 - .00004453\xi \geq 0.00005878 \\ & 2n_1 + 3n_2 + 3n_3 + \sqrt{n_1} + 2\sqrt{n_2} + 3\sqrt{n_3} + \\ & \qquad\qquad 5\left(n_1 + n_2 + n_3\right) \leq 1000 \\ & 2 \leq n_1 \leq N_1 \\ & 2 \leq n_2 \leq N_2 \\ & 2 \leq n_3 \leq N_3 \\ \text{and} \quad & \xi \in [0,1],\ n_1,\ n_2,\ n_3 \text{ are integers.} \end{aligned}$$

Using MATLAB the optimal solution to the above problem is obtained as

$\xi = 0.1890$, $n_1 = 33$, $n_2 = 35$, $n_3 = 56$, with variances $V(\bar{y}_{jst});\ j = 1,2$ under the proposed allocation given as

$$V(\bar{y}_{1st}) = 0.000638635,\ V(\bar{y}_{2st}) = 0.000114102.$$

Therefore, the trace will be

$$V(\bar{y}_{1st}) + V(\bar{y}_{2st}) = 0.00075274.$$

Where,

$$\begin{aligned} f_1 = \quad & 0.5\left[\frac{0.00151959}{(n_1-1)}+\frac{0.00687891}{(n_2-1)}+\frac{0.02254338}{(n_3-1)}-\frac{0.00481210n_1}{9500(n_1-1)}-\frac{0.02178320n_2}{9500(n_2-1)}-\frac{0.06118920n_3}{9500(n_3-1)}\right] \\ & +0.5\Big[\frac{0.000004459}{(n_1-1)^2n_1}+\frac{0.00009538}{(n_2-1)^2n_2}+\frac{0.00104419}{(n_3-1)^2n_3}-\frac{0.00004472n_1}{9500^2(n_1-1)^2} \\ & -\frac{0.00095649n_2}{9500^2(n_2-1)^2}-\frac{0.00769289n_3}{9500^2(n_3-1)^2}\Big]^{\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} f_2 = \quad & 0.5\left[\frac{0.00203233}{(n_1-1)}+\frac{0.00095443}{(n_2-1)}+\frac{0.00146663}{(n_3-1)}-\frac{0.00643571n_1}{9500(n_1-1)}-\frac{0.00302237n_2}{9500(n_2-1)}-\frac{0.00398085n_3}{9500(n_3-1)}\right] \\ & +0.5\Big[\frac{(0.00000797)}{(n_1-1)^2n_1}+\frac{0.00000176}{(n_2-1)^2n_2}+\frac{0.0000042865}{(n_3-1)^2n_3}-\frac{0.00007998n_1}{9500^2(n_1-1)^2} \\ & -\frac{0.00001764n_2}{9500^2(n_2-1)^2}-\frac{0.00003158n_3}{9500^2(n_3-1)^2}\Big]^{\frac{1}{2}}. \end{aligned}$$

4.4.1.1 Comparison with other allocations

In this section, a comparative study is carried out where the proposed method is compared with some other well-defined methods of allocation. Some of these methods are as follows:

Proportional allocation

For the fixed cost of the survey, the proportional allocation may be obtained by substituting $n_h = nW_h$ in the cost function and subsequently, stratum-wise allocations, which are rounded off to nearest integers, may be obtained as

$$n_1 = 40,\ n_2 = 40,\ n_3 = 46,$$

also the trace value, under proportional allocation, is computed as 0.0008066.

Cochran's average allocation

Cochran (1977) suggested the compromise criterion by taking the average of the individual optimum allocations n^*_{jh}; $h = 1,2,\ldots,L,\ j = 1,2,\ldots,p$. These allocations are obtained by solving the individual NLPP for each characteristic; that is, for j^{th} characteristic, the required NLPP may be formulated as

$$\begin{aligned}
&\text{Minimize} && V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_{jh}} \\
&\text{subject to} && \sum_{h=1}^{L} c_h n_{jh} + \sum_{h=1}^{L} t_h \sqrt{n_{jh}} + \omega \sum_{h=1}^{L} \frac{n_{jh}}{\lambda} \leq C_0 \\
&\text{and} && n_{jh} \text{ are integers, } h = 1,2,\ldots,L, \; j = 1,2,\ldots,p,
\end{aligned} \tag{4.4.3}$$

where $n_{jh}^* = (n_{j1}^*, n_{j2}^*, \ldots, n_{jL}^*)$ denotes the individual optimum allocation to the j^{th} characteristic. Therefore Cochran's compromise allocation is computed as

$$n_h = \frac{1}{p} \sum_{j=1}^{p} n_{jh}^*; \; h = 1,2,\ldots,L.$$

For the given numerical values, as given in Table 4.1, the compromise allocation suggested by Cochran (1977) is worked out as $n_1 = 36$, $n_2 = 35$ $n_3 = 53$. The variances for both characteristics are calculated as $V^*(\bar{y}_{1st}) = 0.000664098$ and $V^*(\bar{y}_{2st}) = 0.000111389$ respectively. Therefore the trace value is 0.00077549.

Sukhatme's compromise allocation

This compromise allocation is obtained by optimizing the trace of the variance-covariance matrix of the estimator. The solution to the following NLPP will give the desired compromise allocation (Sukhatme *et al.* (1984)) and the formulation of the NLPP is given as

$$\begin{aligned}
&\text{Minimize} && \sum_{j=1}^{p} V(\bar{y}_{jst}) = \sum_{j=1}^{p} \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} \\
&\text{subject to} && \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} + \omega \sum_{h=1}^{L} \frac{n_h}{\lambda} \leq C_0 \\
&\text{and} && n_h \text{ are integers, } h = 1,2,\ldots,L, \; j = 1,2,\ldots,p.
\end{aligned}$$

On the substitution of numerical values from Table 4.1, the solution is obtained as $n_1 = 25$, $n_2 = 36$ and $n_3 = 61$ and therefore the trace value is calculated as 0.000753271. In Table 4.3, it appears that the proposed allocation gives the least trace value compared to that values obtained by other allocations.

Table 4.3: Allocations with trace value and relative efficiency

S.No.	Allocations	Allocations				Cost	Trace	R.E. w.r.t Proportional Allocation
		n_1	n_2	n_3	n			
1	Proportional	40	40	46	126	1007.3	0.0008066	1.00000
2	Cochran	36	35	53	124	995.67	0.0007755	1.04010
3	Sukhatme	25	36	61	122	991.43	0.0007532	1.07089
4	Proposed	33	35	56	124	999.03	0.0007527	1.07155

Table 4.4: Percentage increase for all characteristics within the variances when the individual optimum for one characteristic is used

	Percentage increase in the variances		
Characteristics	1	2	Proposed
1	0	0.31866816	0.014920487
2	0.55578140	0	0.229092349

Furthermore, the R.E. of proposed allocation with respect to proportional allocation is maximum among the others. Table 4.4 shows the percentage increase in both characteristics' variances when the individual optimum allocation of one characteristic is used for both characteristics, and the proposed allocation is utilized. The proposed allocation provides lesser values of percentage increase in the variances with respect to individual allocations. Table 4.5 shows the percentage increase in both characteristics' variances when other allocations are used instead of individual allocations. For the proposed allocation, these values are minimum in comparison to others. Therefore it may be claimed that the suggested allocation may be regarded as the best allocation. Based on the above discussion, the proposed allocation works well in comparison to other allocations.

Table 4.5: Percentage increase in the variances of different characteristics under individual optimum allocation when different compromise allocations are used

Percentage increase when optimization is done with respect to the criteria				
Characteristics	Proportional	Cochran	Sukhatme	Proposed
1	15.36391	9.441428	2.40957	0.014920487
2	3.225985	7.91908	27.73566	0.229092349

4.4.2 Solution by chance constraints

When the cost of carrying out a sample survey is high and a specified limit on the variances are given, then this method may be used. With the specified values of $V_0^1 = 0.00080$ and $V_0^2 = 0.00040$, the value of e is 2.3263 such that $\phi(e) = p_0 = 0.99$. The equivalent deterministic problem of SNLPP may be given as

$$
\begin{array}{ll}
\text{Minimize} & \displaystyle\sum_{h=1}^{3} c_h n_h + \sum_{h=1}^{3} t_h\sqrt{n_h} + \omega \sum_{h=1}^{3} \frac{n_h}{\lambda} \\
\text{subject to} & \displaystyle\sum_{h=1}^{3} \frac{W_h^2 s_{1h}^2}{(n_h-1)} - \sum_{h=1}^{3} \frac{W_h s_{1h}^2}{9500}\left(\frac{n_h}{n_h-1}\right) \\
& \displaystyle +2.3263\left[\sum_{h=1}^{3} \frac{W_h^4}{n_h(n_h-1)^2}\left(C_{y1h}^4-(s_{1h}^2)^2\right)\right. \\
& \displaystyle \left.-\sum_{h=1}^{3} \frac{W_h^2}{(9500)^2}\left(\frac{n_h}{(n_h-1)^2}\left(C_{y1h}^4-(s_{1h}^2)^2\right)\right)\right]^{1/2} \leq 0.00080 \\
& \displaystyle\sum_{h=1}^{3} \frac{W_h^2 s_{2h}^2}{(n_h-1)} - \sum_{h=1}^{3} \frac{W_h s_{2h}^2}{9500}\left(\frac{n_h}{n_h-1}\right) \\
& \displaystyle +2.3263\left[\sum_{h=1}^{3} \frac{W_h^4}{n_h(n_h-1)^2}\left(C_{y2h}^4-(s_{2h}^2)^2\right)\right. \\
& \displaystyle \left.-\sum_{h=1}^{3} \frac{W_h^2}{(9500)^2}\left(\frac{n_h}{(n_h-1)^2}\left(C_{y2h}^4-(s_{2h}^2)^2\right)\right)\right]^{1/2} \leq 0.00040 \\
& 2 \leq n_1 \leq N_1 \\
& 2 \leq n_2 \leq N_2 \\
& 2 \leq n_3 \leq N_3 \\
\text{and} & n_1,\ n_2,\ n_3 \text{ are integers.}
\end{array}
\tag{4.4.4}
$$

The NLPP (4.4.4) solutions are obtained using MATLAB by the Lagrange multiplier technique and INLPP technique. The solutions are reported in Table 4.6.

Table 4.6: Allocations with the incurred cost

S.No.	Allocations	Allocations				Cost
		n_1	n_2	n_3	n	
1	Lagrange multiplier (non integer)	12.97	52.74	61.33	127	1045.10
2	Lagrange multiplier (rounded)	13	53	61	127	1044.59
3	Lagrange multiplier (integer)	13	53	62	128	1052.78
4	Stochastic (non integer)	18.66	36.95	66.30	122	997.630
5	Stochastic (rounded)	19	37	66	122	997.890
6	Stochastic (integer)	19	38	65	122	997.870

Table 4.6 shows the use of the Lagrange multiplier technique and INLPP technique to minimize the survey's total cost. If the continuous solution to the NLPP (4.4.4) is considered, then the nonlinear programming technique is preferable to that of the Lagrange multiplier technique. If the continuous solution is adjusted off to the closest whole number, then the nonlinear programming technique provides the survey's minimum cost. Furthermore, if integer restriction is a must, then the use of a nonlinear programming technique is advisable.

4.5 Conclusion

In general, stratum variances' true values may not be known in advance but may be estimated. In this way, the problem is defined as a multivariate-multiobjective SNLPP in this chapter. The formulated SNLPPs may be converted into their deterministic form using a modified E- model and chance constraints techniques. The formulated problems' solutions may be computed either by minimizing the sampling variances for a fixed cost or minimizing the cost for the fixed precision value of the variances of estimates. For numerical illustration, the data are generated by conducting simulation using R, and the formulated NLPPs may be solved by using MATLAB. The proposed compromise allocation provides the best outcomes for the given numerical application than that obtained by other compromise criteria, as discussed in this chapter. Furthermore, for large scale investigations, it is vital to select the appropriate method for attaining the study's objectives.

CHAPTER 5

Efficient estimation of population parameter in stratified random sampling design

5.1 Introduction

Tschuprow (1923) and Neyman (1934) discussed the procedure to workout an allocation by minimizing the variance of the sample mean for a fixed linear cost function in stratified random sampling to estimate the population parameters under study. Neyman (1934) used a Lagrange multiplier technique to find the optimum sample size for univariate case. In sampling theory, the use of auxiliary information for improving the precision of estimates of the population parameters (that is mean or total) may be used. Cochran (1940) suggested the use of ratio method of estimation if the correlation between y (study variable) and the x (auxiliary variable) is positive (high). In the situation of negative correlation between study and auxiliary variable, Robson (1957) and then Murthy (1964) suggested to use product method of estimation. Many authors have used information on auxiliary variables such as population mean, variance, kurtosis, skewness, etc., for estimating population mean and variance of the study variable. Das and Tripathi (1978), Srivastava and Jhajj (1980), Isaki

(1983), Prasad and Singh (1990), Singh and Kataria (1990), Ahmed *et al.* (2000), Kadilar and Cingi (2006a,b), Gupta and Shabbir (2008) studied the population variance of study variable by using population mean, variance, kurtosis and coefficient of variation of auxiliary variable in simple and stratified random sampling. Upadhyaya *et al.* (2011) worked on ratio and product exponential type estimators under large sample approximations. Olufadi and Kadilar (2014) estimated the population variance of interested variate in simple and two-phase sampling using the variance of auxiliary variables and got interesting results. In many practical situations, it has been observed that the population mean of the study variable is unknown, however, the population median of the study variable may be used to estimate population parameter. Median is easily available parameter without having exact information on every unit of population. Therefore the use of the study variable's median may be a better choice to improve efficiency without increasing the total cost of the survey. For the study of the modified ratio type estimators of population mean of the study variable, one may refer to Kadilar and Cingi (2003), Tailor and Sharma (2009), Yan and Tian (2010), Subramani and Kumarapandiyan (2012, 2013), Subramani (2013), Yadav *et al.* (2014), Abid *et al.* (2016) and Yadav *et al.* (2016). Many authors have used regression and product type estimators of population parameters for predictive estimation of population parameters.

In this chapter, the ratio type estimators are proposed where the population mean is unknown but the study variable's population median is known. The problem is formulated in a single objective NLPP with linear and nonlinear cost function and solved by using the integer programming technique and Lagrange multiplier technique. The numerical illustration with real data set and simulated data set also presented to its utility of the proposed work.

This chapter is based on my research paper *"Efficient estimation of population mean under stratified random sampling with linear cost function"* and published in Communications in

Statistics - Simulation and Computation.

5.2 Efficient estimation with linear cost function

In this section the problem is formulated as an NLPP with linear cost function.

5.2.1 Introduction

Let the stratified random sampling where a population of size N is divided into L strata, such that $\sum_{h=1}^{L} N_h = N$, let for the h^{th} stratum N_h, $\overline{Y}_h$, M_h, C_{yh}^2, C_{ymh}, C_{mh}^2 and ρ_h denote the stratum size, stratum mean, stratum median, stratum correlation coefficient with study variable, stratum correlation coefficient between study variable and median, stratum correlation coefficient with median and correlation coefficient between X and Y respectively, where $h = 1, 2, \ldots, L$.

$\beta_{1h} = \frac{N_h \sum_{i=1}^{N_h} \left(X_i - \overline{X}_h\right)}{\left(N_h - 1\right)\left(N_h - 2\right) S_{xh}^3}$: Stratum coefficient of skewness of auxiliary variable

$\beta_{2h} = \frac{N_h \sum_{i=1}^{N_h} \left(X_i - \overline{X}_h\right)}{\left(N_h - 1\right)\left(N_h - 2\right) S_{xh}^3} - \frac{3\left(N_h - 1\right)^2}{\left(N_h - 2\right)\left(N_h - 3\right)}$: Stratum coefficient of kurtosis of auxiliary variable

$QD_h = \frac{Q_{3h} - Q_{1h}}{2}$: Stratum quartile deviation

$G_h = \frac{4}{N_h - 1} \sum_{i=1}^{N_h} \left(\frac{2i - N_h - 1}{2N_h}\right) X_i$: Stratum gini's mean difference

$B(\cdot)$: Bias of the estimator

$MSE(\cdot)$: Mean square error of the estimator

$PRE(t_{pr_j(st)}, t_*) = \frac{MSE(t_*)}{MSE(t_{pr_j(st)})} * 100$: Percentage relative efficiency (PRE) of the estimator,

where t_* is defined as comparative estimator in Section 5.2.3 and PRE of obtained with respect to $t_{pr_j(st)}$, for $j = 1, 2$.

5.2.2 Proposed estimator $(t_{pr1(st)})$

Inspired by Srivastava (1967) and Subramani (2016), the ratio type estimator of population mean using the information on population median of the study variable in stratified random sampling is proposed and defined as

$$t_{pr1(st)} = \sum_{h=1}^{L} \bar{y}_h \left[1 - b_{ymh} log \left(\frac{m_h}{M_h} \right) \right]. \tag{5.2.1}$$

In order to study the properties of the proposed estimator, the following approximations have been used

$$\bar{y}_h = \sum_{h=1}^{L} \bar{Y}_h (1+e_0), \ m_h = M_h (1+e_1), \ b_{ymh} = \frac{s_{ymh}}{s^2_{mh}}, \ s_{ymh} = S_{ymh} (1+e_3),$$

$$s^2_{mh} = S^2_{mh} (1+e_2), \ b_{ymh} = \frac{S_{ymh} (1+e_3)}{S^2_{mh} (1+e_2)},$$

such that,

$$E(e_0) = 0, \ E(e_1) = \frac{\overline{M}_h - M_h}{M_h} = \frac{Bias(m_h)}{M_h}, \ E\left(e_0^2\right) = \lambda_h C^2_{yh}, \ E\left(e_1^2\right) = \lambda_h C^2_{mh},$$

$$E(e_0 e_1) = \lambda C_{ymh}, \ E(e_1 e_2) = \lambda_h \frac{\mu_{21h}}{\sigma_{ymh} M_h}, \ E(e_1 e_3) = \lambda_h \frac{\mu_{12h}}{\sigma^2_{mh} M_h},$$

$$B_h = \frac{S_{ymh}}{S^2_{mh}} = \frac{M_h \bar{Y}_h C_{ymh}}{M_h^2 C^2_{mh}} = \frac{\bar{Y}_h C_{ymh}}{M_h C^2_{mh}},$$

where

$$\mu_{rsh} = \frac{1}{N_h} \sum_{i=1}^{N_h} (m_{ih} - M_h)^r \left(y_{ih} - \bar{Y}_h\right)^s, \ \overline{M}_h = \frac{1}{{}^{N_h}C_{n_h}} \sum_{i=1}^{{}^{N}C_n} m_{ih},$$

m_{ih} is sample median of the i^{th} sample for the h^{th} stratum, respectively. $h = 1, 2, \ldots, L, \ i = 1, 2, \ldots, {}^{N_h}C_{n_h}$.

Applying above approximations, the proposed estimator may be explicited as

$$
\begin{aligned}
t_{pr1(st)} &= \sum_{h=1}^{L} \bar{y}_h \left[1 - b_{ymh} log\left(\frac{m_h}{M_h}\right)\right] \\
&= \sum_{h=1}^{L} \bar{Y}_h (1+e_0)\left[1 - \frac{S_{ymh}(1+e_3)}{S_{mh}^2(1+e_2)} log(1+e_1)\right] \\
&= \sum_{h=1}^{L} \bar{Y}_h (1+e_0)\left[1 - B_h(1+e_3)(1+e_2)^{-1}\left(e_1 - e_1^2/2 + \ldots\right)\right] \\
&= \sum_{h=1}^{L} \bar{Y}_h (1+e_0)\left[1 - B_h(1+e_3)\left(e_1 - e_1e_2 - e_1^2/2\right)\right] \\
&= \sum_{h=1}^{L} \bar{Y}_h \left[(1+e_0) - B_h(1+e_0)\left(e_1 - e_1e_2 - e_1^2/2 + e_1e_3\right)\right] \\
&= \sum_{h=1}^{L} \bar{Y}_h \left[(1+e_0) - B_h\left(e_1 + e_0e_1 - e_1e_2 + e_1e_3 - e_1^2/2\right)\right].
\end{aligned}
$$

Now,

$$
\left(t_{pr1(st)} - \bar{Y}_h\right) = \sum_{h=1}^{L} \bar{Y}_h \left[e_0 - B_h\left(e_1 + e_0e_1 - e_1e_2 + e_1e_3 - e_1^2/2\right)\right], \qquad (5.2.2)
$$

taking expectations on both sides of equation (5.2.2), the bias of suggested estimator is obtained as

$$
B\left(t_{pr1(st)}\right) = E\left[t_{st} - \bar{Y}_h\right]
$$

$$
B\left(t_{pr1(st)}\right) = \bar{Y}_h \left[B_h\left(\frac{\lambda_h\mu_{21h}}{\sigma_{ymh}M_h} + \frac{1}{2}\lambda_h C_{mh}^2 - \frac{\lambda_h\mu_{12h}}{\sigma_{mh}^2 M_h} - \lambda_h C_{ymh} - \frac{Bias(m_h)}{M_h}\right)\right].
$$

From equation (5.2.2), up to approximation of order one, it may be written as

$$
\left(t_{pr1(st)} - \bar{Y}_h\right)^2 = \sum_{h=1}^{L} \bar{Y}_h^2 \left[e_0 - B_h e_1\right]^2, \qquad (5.2.3)
$$

taking expectations on both sides of equation (5.2.3), the MSE of suggested estimator is obtained as

$$
\begin{aligned}
E\left(t_{pr1(st)}-\overline{Y}_h\right)^2 &= \sum_{h=1}^{L}\overline{Y}_h^2\left[E\left(e_0^2\right)+B_h^2E\left(e_1^2\right)-2B_hE\left(e_0e_1\right)\right] \\
&= \sum_{h=1}^{L}\overline{Y}_h^2\left[\lambda_hC_{yh}^2+B_h^2\lambda_hC_{mh}^2-2\lambda_hB_hC_{ymh}\right]. \\
MSE\left(t_{pr1(st)}\right) &= \sum_{h=1}^{L}\overline{Y}_h^2\lambda_h\left[C_{yh}^2+B_h^2C_{mh}^2-2B_hC_{ymh}\right],
\end{aligned}
$$

where $B_h=\sum_{h=1}^{L}\frac{S_{ymh}}{S_{mh}^2}$, $\lambda_h=\frac{N_h-n_h}{N_hn_h}$.

$$
MSE\left(t_{pr1(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)\left[C_{yh}^2+B_h^2C_m^2-2B_hC_{ymh}\right]. \tag{5.2.4}
$$

Ignoring fpc factor in equation (5.2.4), we have,

$$
MSE\left(t_{pr1(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h^2\frac{1}{n_h}\left[C_{yh}^2+B_h^2C_{mh}^2-2B_hC_{ymh}\right]. \tag{5.2.5}
$$

Cost function

Due to budget constraint, the cost of a survey is the major factor to workout sample allocations to various strata. Khan *et al.* (2008) introduced linear cost function and considered as fixed total cost C_0 of the survey in terms of a linear function of n_h; $h=1,2,\ldots,L$.

$$
C_0=C-c_0=\sum_{h=1}^{L}c_hn_h, \tag{5.2.6}
$$

where c_h is the per-unit cost of measurement of all the characteristics in the h^{th} stratum; $h=1,2,\ldots,L$ and c_0 is the overhead cost for the survey.

In this chapter, the problem is formulated by minimizing MSE of the estimator for fixed linear cost function given in (5.2.6). Then the optimization problem can be defined as

$$\begin{aligned} &\text{Minimize} \quad MSE\left(t_{pr1(st)}\right) \\ &\text{subject to} \quad \sum_{h=1}^{L} c_h n_h \leq C_0 \\ &\qquad\qquad\quad 2 \leq n_h \leq N_h \\ &\text{and} \qquad n_h \text{ are integers; } h = 1,2,\ldots,L. \end{aligned} \tag{5.2.7}$$

Using the cost function, the MSE may now be given as

$$MSE\left(t_{pr1(st)}\right) = \sum_{h=1}^{L} \overline{Y}_h^2 \frac{1}{n_h}\left[C_{yh}^2 + B_h^2 C_{mh}^2 - 2B_h C_{ymh}\right].$$

5.2.3 Some other estimators in sampling literature

In this section, some other estimators are discussed for the sake of comparison with the proposed estimator.

Unbiased estimator $\left(t_{0(st)}\right)$

The most widely used estimator of population mean of the study variable, is given as

$$t_0 = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

where $\overline{y}$ is sample mean and it is an unbiased estimator of population mean and up to first order of approximation it is given by

$$V\left(t_0\right) = \frac{1-f}{n} S_y^2 = \frac{1-f}{n} \overline{Y}^2 C_y^2,$$

where

$$C_y = \frac{S_y}{\overline{Y}},\ S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2,\ f = \frac{n}{N}.$$

To define unbiased estimator for stratified random sampling, the population is divided into

L strata and it is defined as

$$t_{0(st)} = \overline{y}_h = \sum_{i=1}^{n_h} \frac{1}{n_h} y_{ih},$$

where sample mean is an unbiased estimator of population mean and up to first order of approximation it is given by

$$V\left(t_{0(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} S_{yh}^2 = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \overline{Y}_h^2 C_{yh}^2,$$

where

$$C_{yh} = \frac{S_{yh}}{\overline{Y}_h},\ S_{yh}^2 = \sum_{i=1}^{N_h} \sum_{h=1}^{L} \frac{1}{N_h - 1} \left(Y_{ih} - \overline{Y}_h\right)^2,\ f_h = \frac{n_h}{N_h}.$$

Subramani estimator $\left(t_{s(st)}\right)$

Subramani (2016) proposed the estimator by using population median of the study variable and defined the ratio estimator for the population mean of the study variable which is given as

$$t_s = \overline{y}\left(\frac{M}{m}\right),$$

where M and m are the population and sample medians of study variable respectively. t_s is a biased estimator and its bias and the MSE, up to first order of approximation, are defined as

$$B\left(t_s\right) = \frac{1-f}{n} \overline{Y} \left[C_m^2 - C_{ym} - 2RC_{ym}\right]$$

and

$$MSE\left(t_s\right) = \frac{1-f}{n} \overline{Y}^2 \left[C_y^2 + R^2 C_m^2 - 2RC_{ym}\right],$$

where

$$R=\frac{\overline{Y}}{M},\ S_m^2=\frac{1}{N-1}\sum_{i=1}^{N}(m_i-M)^2,\ S_{ym}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\overline{y}_i-\overline{Y}\right)(m_i-M),$$

$$C_m=\frac{S_m}{M},\ C_{ym}=\frac{S_{ym}}{\overline{Y}M}.$$

By transforming the Subramani estimator into stratified random sampling for h^{th} stratum. It is defined as

$$t_{s(st)}=\overline{y}_h\left(\frac{M_h}{m_h}\right),$$

where M_h is the population median, m_h is sample median of study variable for h^{th} stratum respectively. Hence, $t_{s(st)}$ is a biased estimator and its bias and the MSE, up to first order of approximation, are given as

$$B\left(t_{s(st)}\right)=\sum_{h=1}^{L}\frac{1-f_h}{n_h}\overline{Y}_h\left[C_{mh}^2-C_{ymh}-2R_hC_{ymh}\right]$$

and

$$MSE\left(t_{s(st)}\right)=\sum_{h=1}^{L}\frac{1-f_h}{n_h}\overline{Y}_h^2\left[C_{yh}^2+R_h^2C_{mh}^2-2R_hC_{ymh}\right],$$

where

$$R_h=\frac{\overline{Y}_h}{M_h},\ S_{ymh}=\sum_{i=1}^{N_h}\sum_{h=1}^{L}\frac{1}{N_h-1}\left(\overline{y}_{ih}-\overline{Y}_h\right)(m_{ih}-M_h),\ S_{mh}^2=\sum_{i=1}^{N_h}\sum_{h=1}^{L}(m_{ih}-M_h)^2,$$

$$C_{mh}=\frac{S_{mh}}{M_h},\ C_{ymh}=\frac{S_{ymh}}{\overline{Y}_hM_h}\ ;\ h=1,2,\ldots,L.$$

Kadilar and Cingi estimators

Kadilar and Cingi (2004) proposed the ratio type estimators for simple random sampling as shown in Table 5.1. We restructured these existing estimators into stratified random sampling from simple random sampling, where population are divided into different strata

$h=1,2,\ldots,L$ as given in Table 5.2.

Table 5.1: Kadilar and Cingi (2004) estimators for simple random sampling

Estimators	Bias	MSE	Constant
$t_1=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}}\overline{X}$	$B\left(t_1\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_1^2$	$MSE\left(t_1\right)=\frac{1}{n}\left[R_1^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_1=\frac{\overline{Y}}{\overline{X}}$
$t_2=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}+C_x}\overline{X}+C_x$	$B\left(t_2\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_2^2$	$MSE\left(t_2\right)=\frac{1}{n}\left[R_2^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_2=\frac{\overline{Y}}{\overline{X}+C_x}$
$t_3=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}+\beta_2}\overline{X}+\beta_2$	$B\left(t_3\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_3^2$	$MSE\left(t_3\right)=\frac{1}{n}\left[R_3^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_3=\frac{\overline{Y}}{\overline{X}+\beta_2}$
$t_4=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}\beta_2+C_x}\overline{X}\beta_2+C_x$	$B\left(t_4\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_4^2$	$MSE\left(t_4\right)=\frac{1}{n}\left[R_4^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_4=\frac{\overline{Y}\beta_2}{\overline{X}\beta_2+C_x}$
$t_5=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}C_x+\beta_2}\overline{X}C_x+\beta_2$	$B\left(t_5\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_5^2$	$MSE\left(t_5\right)=\frac{1}{n}\left[R_5^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_5=\frac{\overline{Y}C_x}{\overline{X}C_x+\beta_2}$
$t_6=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}C_x+\rho}\overline{X}C_x+\rho$	$B\left(t_6\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_6^2$	$MSE\left(t_6\right)=\frac{1}{n}\left[R_6^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_6=\frac{\overline{Y}C_x}{\overline{X}C_x+\rho}$
$t_7=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}\rho+C_x}\overline{X}\rho+C_x$	$B\left(t_7\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_7^2$	$MSE\left(t_7\right)=\frac{1}{n}\left[R_7^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_7=\frac{\overline{Y}\rho}{\overline{X}\rho+C_x}$
$t_8=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}\beta_2+\rho}\overline{X}\beta_2+\rho$	$B\left(t_8\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_8^2$	$MSE\left(t_8\right)=\frac{1}{n}\left[R_8^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_8=\frac{\overline{Y}\beta_2}{\overline{X}\beta_2+\rho}$
$t_9=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}\rho+\beta_2}\overline{X}\rho+\beta_2$	$B\left(t_9\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_9^2$	$MSE\left(t_9\right)=\frac{1}{n}\left[R_9^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_9=\frac{\overline{Y}\rho}{\overline{X}\rho+\beta_2}$
$t_{10}=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}+\beta_1}\overline{X}+\beta_1$	$B\left(t_{10}\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_{10}^2$	$MSE\left(t_{10}\right)=\frac{1}{n}\left[R_{10}^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_{10}=\frac{\overline{Y}}{\overline{X}+\beta_1}$
$t_{11}=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}\beta_1+\beta_2}\overline{X}\beta_1+\beta_2$	$B\left(t_{11}\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_{11}^2$	$MSE\left(t_{11}\right)=\frac{1}{n}\left[R_{11}^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_{11}=\frac{\overline{Y}\beta_1}{\overline{X}\beta_1+\beta_2}$
$t_{12}=\frac{\bar{y}+b\left(\overline{X}-\bar{x}\right)}{\bar{x}+G}\overline{X}+G$	$B\left(t_{12}\right)=\frac{1-f}{n}\frac{S_x^2}{\overline{Y}}R_{12}^2$	$MSE\left(t_{12}\right)=\frac{1}{n}\left[R_{12}^2S_x^2+S_y^2\left(1-\rho^2\right)\right]$	$R_{12}=\frac{\overline{Y}}{\overline{X}+G}$

Table 5.2: Reconstructed estimators by using Kadilar and Cingi (2004) into stratified random sampling

Estimators	Bias	MSE	Constant
$t_{1(st)} = \frac{\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h}\bar{X}_h$	$B\left(t_{1(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{1h}^2$	$MSE\left(t_{1(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{1h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{1h} = \frac{\bar{Y}_h}{\bar{X}_h}$
$t_{2(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h + C_{xh}}\bar{X}_h + C_{xh}$	$B\left(t_{2(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{2h}^2$	$MSE\left(t_{2(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{2h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{2h} = \frac{\bar{Y}_h}{\bar{X}_h + C_{xh}}$
$t_{3(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h + \beta_{2h}}\bar{X}_h + \beta_{2h}$	$B\left(t_{3(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{3h}^2$	$MSE\left(t_{3(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{3h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{3h} = \frac{\bar{Y}_h}{\bar{X}_h + \beta_{2h}}$
$t_{4(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h\beta_{2h} + C_{xh}}\bar{X}_h\beta_{2h} + C_{xh}$	$B\left(t_{4(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{4h}^2$	$MSE\left(t_{4(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{4h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{4h} = \frac{\bar{Y}_h\beta_{2h}}{\bar{X}_h\beta_{2h} + C_{xh}}$
$t_{5(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h C_{xh} + \beta_{2h}}\bar{X}_h C_{xh} + \beta_{2h}$	$B\left(t_{5(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{5h}^2$	$MSE\left(t_{5(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{5h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{5h} = \frac{\bar{Y}_h C_{xh}}{\bar{X}_h C_{xh} + \beta_{2h}}$
$t_{6(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h C_{xh} + \rho_h}\bar{X}_h C_{xh} + \rho_h$	$B\left(t_{6(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{6h}^2$	$MSE\left(t_{6(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{6h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{6h} = \frac{\bar{Y}_h C_{xh}}{\bar{X}_h C_{xh} + \rho_h}$
$t_{7(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h\rho_h + C_{xh}}\bar{X}_h\rho_h + C_{xh}$	$B\left(t_{7(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{7h}^2$	$MSE\left(t_{7(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{7h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{7h} = \frac{\bar{Y}_h\rho_h}{\bar{X}_h\rho_h + C_{xh}}$
$t_{8(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h\beta_{2h} + \rho_h}\bar{X}_h\beta_{2h} + \rho_h$	$B\left(t_{8(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{8h}^2$	$MSE\left(t_{8(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{8h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{8h} = \frac{\bar{Y}_h\beta_{2h}}{\bar{X}_h\beta_{2h} + \rho_h}$
$t_{9(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h\rho_h + \beta_{2h}}\bar{X}_h\rho_h + \beta_{2h}$	$B\left(t_{9(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{9h}^2$	$MSE\left(t_{9(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{9h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{9h} = \frac{\bar{Y}_h\rho_h}{\bar{X}_h\rho_h + \beta_{2h}}$
$t_{10(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h + \beta_{1h}}\bar{X}_h + \beta_{1h}$	$B\left(t_{10(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{10h}^2$	$MSE\left(t_{10(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{10h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{10h} = \frac{\bar{Y}_h}{\bar{X}_h + \beta_{1h}}$
$t_{11(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h\beta_{1h} + \beta_{2h}}\bar{X}_h\beta_{1h} + \beta_{2h}$	$B\left(t_{11(st)}\right) = \sum_{h=1}^{L} \frac{1-f_h}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{11h}^2$	$MSE\left(t_{11(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{11h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{11h} = \frac{\bar{Y}_h\beta_{1h}}{\bar{X}_h\beta_{1h} + \beta_{2h}}$
$t_{12(st)} = \frac{\bar{y}_h + b\left(\bar{X}_h - \bar{x}_h\right)}{\bar{x}_h + G_h}\bar{X}_h + G_h$	$B\left(t_{12(st)}\right) = \sum_{h=1}^{L} \frac{1-f}{n_h} \frac{S_{xh}^2}{\bar{Y}_h} R_{12h}^2$	$MSE\left(t_{12(st)}\right) = \sum_{h=1}^{L} \frac{1}{n_h}\left[R_{12h}^2 S_{xh}^2 + S_{yh}^2\left(1-\rho_h^2\right)\right]$	$R_{12h} = \frac{\bar{Y}_h}{\bar{X}_h + G_h}$

Cochran estimator$(t_{cr(st)})$

Cochran (1940) proposed the ratio type estimator of population mean, which is given as

$$t_{cr} = \frac{\overline{y}}{\overline{x}}\overline{X},$$

bias and the MSE, are defined as

$$B\left(t_{cr}\right) = \lambda\overline{Y}\left(C_x^2 - C_{yx}\right)$$

and

$$MSE\left(t_{cr}\right) = \lambda\overline{Y}^2\left(C_y^2 + C_x^2 - 2\rho C_y C_x\right).$$

The Cochran (1940) estimator may also be used for stratified population which consists of L strata. For h^{th} stratum, it is defined as

$$t_{cr(st)} = \frac{\overline{y}_h}{\overline{x}_h}\overline{X}_h,\ h = 1,2,\ldots,L.$$

For L strata, $t_{cr(st)}$ is a biased estimator and its bias and the MSE, up to first order of approximation, are given as

$$B\left(t_{cr(st)}\right) = \sum_{h=1}^{L}\lambda_h\overline{Y}_h\left(C_{xh}^2 - C_{yxh}\right)$$

and

$$MSE\left(t_{cr(st)}\right) = \sum_{h=1}^{L}\lambda_h\overline{Y}_h^2\left(C_{yh}^2 + C_{xh}^2 - 2\rho_h C_{yh}C_{xh}\right),$$

where $C_{yh}^2 = \frac{S_{yh}^2}{\overline{Y}_h^2}$ is the coefficient of variation of variable y_h, $C_{xh}^2 = \frac{S_{Xh}^2}{\overline{X}_h^2}$ is the coefficient of variation of variable x_h, and $\rho_h = \frac{C_{yxh}}{C_{yh}C_{xh}}$ is the coefficient of correlation between y_h and x_h for h^{th} stratum.

Bahl and Tuteja $(t_{pe(st)})$

Bahl and Tuteja (1991) defined the product exponential type estimator for population mean $\overline{Y}$, which is given as

$$t_{pe} = \overline{y}\exp\left(\frac{\overline{x}-\overline{X}}{\overline{x}+\overline{X}}\right),$$

bias and the MSE of the estimator are given as

$$B\left(t_{pe}\right) = \lambda\overline{Y}\left(\frac{C_x^2}{8}\right)(4K-1)$$

and

$$MSE\left(t_{pe}\right) = \lambda\overline{Y}^2\left(C_y^2 + \frac{1}{4}C_x^2\left(1+4K\right)\right),$$

where $K = \rho\frac{C_x}{C_y}$.

The Bahl and Tuteja (1991) estimator may be used for stratified population which consists of L strata. For h^{th} stratum, it is defined as

$$t_{pe(st)} = \overline{y}_h exp\left(\frac{\overline{x}_h-\overline{X}_h}{\overline{x}_h+\overline{X}_h}\right),$$

For L strata, $t_{pe(st)}$ is a biased estimator and its bias and the MSE, up to first order of approximation, are given by

$$B\left(t_{pe(st)}\right) = \sum_{h=1}^{L}\lambda_h\overline{Y}_h\left(\frac{C_{xh}^2}{8}\right)(4K_h-1)$$

and

$$MSE\left(t_{pe(st)}\right) = \sum_{h=1}^{L}\lambda_h\overline{Y}_h^2\left(C_{yh}^2 + \frac{1}{4}C_{xh}^2\left(1+4K_h\right)\right),$$

where $K_h = \rho_h\frac{C_{xh}}{C_{yh}};\ h = 1,2,\ldots,L.$

Bahl and Tuteja $(t_{p(st)})$

Bahl and Tuteja (1991) proposed the ratio type estimator of population mean, which is defined as

$$t_p = \bar{y}\frac{\bar{x}}{\bar{X}}.$$

The bias and the MSE of this estimator may be given as

$$B(t_p) = \lambda \bar{Y} K C_x^2$$

and

$$MSE(t_p) = \lambda \bar{Y}^2 \left(C_y^2 + C_x^2 (1+2K)\right),$$

The Bahl and Tuteja (1991) may be used for stratified population which consists of L strata. For h^{th} stratum, it is defined as

$$t_{p(st)} = \bar{y}_h \frac{\bar{x}_h}{\bar{X}_h},$$

For L strata, $t_{p(st)}$ is a biased estimator and its bias and the MSE, up to first order of approximation, are given by

$$B\left(t_{p(st)}\right) = \sum_{h=1}^{L} \lambda_h \bar{Y}_h K_h C_{xh}^2$$

and

$$MSE\left(t_{p(st)}\right) = \sum_{h=1}^{L} \lambda_h \bar{Y}_h^2 \left(C_{yh}^2 + C_{xh}^2 (1+2K_h)\right),$$

where $K_h = \rho_h \frac{C_{xh}}{C_{yh}};\ h = 1, 2, \ldots, L.$

5.2.4 Formulation of the problem

To find out the efficient estimator of population parameter, among all estimators discussed in this chapter, a comparative study may be required. This study may be carried out by developing the problem of INLPP. For such problems, the MSE of the estimators, as discussed in this chapter, may be minimized for fixed linear cost function. The INLPP may be formulated as

$$\begin{aligned} \text{Minimize} \quad & MSE\,(t_*) \\ \text{subject to} \quad & \sum_{h=1}^{L} c_h n_h \leq C_0 \\ & 2 \leq n_h \leq N_h \\ \text{and} \quad & n_h \text{ are integers; } h = 1,2,\ldots,L. \end{aligned} \tag{5.2.8}$$

where t_* is one of the possible estimators, as defined in Section 5.2.3, whose MSE is to be compared with others.

5.2.5 Numerical illustrations

In this section, the proposed estimator is applied to both real data set and simulated data set. The simulated data set is generated by using the bivariate normal distribution. The real data set is taken from India's census is 2011, Lucknow, Uttar Pradesh which may be accessed by the link `http://censusindia.gov.in/2011census/dchb/0926_PART_B_DCHB_LUCKNOW.pdf` and relevant data with respect to characteristics correspond to the area of certain villages and number of households which may be considered as study variable and auxiliary variable respectively. The population is divided into four strata without overlapping. In this numerical illustration, the problem is solved by using integer programming and Lagrange multiplier techniques. These computations are obtained by MATLAB and R software. The data are summarized in Tables 5.3 and 5.4.

Real data set

Table 5.3: Obtained results from real data for study variable (y) and auxiliary variable (x)

Population $(N = 645,\ h = 4)$										
h	N_h	$\sum_{h=1}^{L} N_{1h}$	$\sum_{h=1}^{L} N_{2h}$	$\bar{Y}_h$	$\bar{X}_h$	M_{yh}	M_{xh}	C_{yh}^2	C_{xh}^2	c_h
1	237	27548.01	35286	116.2363	148.886	116.810	145.0	0.3148490	0.367079	2
2	164	50446.83	66728	307.6026	406.878	292.295	389.5	0.1839708	0.133430	3
3	90	10401.43	13819	547.4437	727.316	548.770	723.0	1.6424440	1.357187	4
4	154	21198.79	37954	757.0996	1355.50	727.165	1319	4.7946940	10.31789	5

Table 5.4: Obtained results from real data for study variable (y) and auxiliary variable (x) (Continue...)

Population $(N = 645,\ h = 4)$										
h	$C_{m_y h}$	$C_{m_x h}$	C_{xyh}	C_{ymh}	C_{xmh}	β_{1xh}	β_{2xh}	G_h	QD_h	ρ_h
1	0.1455367	0.1729461	2853.599	0.2006495	0.2258169	0.6228215	2.573214	65.0334	44.1125	0.9962281
2	0.3040596	0.3861879	9002.096	0.1423751	0.1154959	1.757425	5.233497	122.087	77.85875	0.9955207
3	3.8489530	4.248597	1630.967	2.4950080	2.369188	0.2950488	1.639132	117.2344	85.21125	0.9783495
4	8.7804220	21.65683	44928.72	6.2031650	13.71308	0.6505419	2.13844	451.3315	291.4388	0.9614615

Solution by integer nonlinear programming technique

The formulated problem for proposed estimator, as given in (5.2.8), for the given data may be written as

$$
\begin{aligned}
&\text{Minimize} && \frac{516.4248}{n_1} + \frac{11116.53}{n_2} + \frac{7526.421}{n_3} + \frac{240588.1}{n_4} \\
&\text{subject to} && 2n_1 + 3n_2 + 4n_3 + 5n_4 \leq 400 \\
& && 2 \leq n_1 \leq N_1 \\
& && 2 \leq n_2 \leq N_2 \\
& && 2 \leq n_3 \leq N_3 \\
& && 2 \leq n_4 \leq N_4 \\
&\text{and} && n_1,\ n_2,\ n_3,\ n_4 \text{ are integers.}
\end{aligned}
\tag{5.2.9}
$$

The solution of the above problem is obtained by using MATLAB which is given as

$n_1 = 4,\ n_2 = 16,\ n_3 = 11,\ n_4 = 60,$ with minimum value of the MSE is 5517.9.

Solution by Lagrange multiplier technique

Lagrange multiplier technique may also be used to determine the optimum value of n_h. For this technique the Lagrange function ϕ is defined for proposed estimator which is given as

$$L = \phi\left(n_h, \lambda\right) = \sum_{h=1}^{L} \overline{Y}_h^2 \frac{1}{n_h} \left(C_{yh}^2 + B_h^2 C_{mh}^2 - 2B_h C_{ymh}\right) + \lambda \left(\sum_{h=1}^{L} c_h n_h - C_0\right), \tag{5.2.10}$$

where λ is a Lagrange multiplier.

$$\frac{\delta L}{\delta n_h} = \frac{\delta}{\delta n_h}\left(\sum_{h=1}^{L} \overline{Y}_h^2 \frac{1}{n_h} \left(C_{yh}^2 + B_h^2 C_{mh}^2 - 2B_h C_{ymh}\right) + \lambda \left(\sum_{h=1}^{L} c_h n_h - C_0\right)\right) = 0$$

gives

$$n_h = \sqrt{\frac{\overline{Y}_h^2 \left(C_{yh}^2 + B_h^2 C_{mh}^2 - 2B_h C_{ymh}\right)}{\lambda c_h}} \tag{5.2.11}$$

and

$$\frac{\delta L}{\delta \lambda} = \frac{\delta}{\delta \lambda}\left(\overline{Y}_h^2 \frac{1}{n_h} \left(C_{yh}^2 + B^2 C_{mh}^2 - 2B_h C_{ymh}\right) + \lambda \left(\sum_{h=1}^{L} c_h n_h - C_0\right)\right) = 0$$

gives

$$\sqrt{\lambda} = \frac{\sqrt{\overline{Y}_h^2 \left(C_{yh}^2 + B_h^2 C_{mh}^2 - 2B_h C_{ymh}\right) c_h}}{C_0}. \tag{5.2.12}$$

After putting the value of λ in equation (5.2.11), the value of n_h is obtained as

$$n_h = \frac{C_0 \sqrt{\frac{\overline{Y}_h^2 \left(C_{yh}^2 + B_h^2 C_{mh}^2 - 2BC_{ymh}\right)}{c_h}}}{\sum_{h=1}^{4} \sqrt{\overline{Y}^2 \left(C_{yh}^2 + B_h^2 C_{mh}^2 - 2B_h C_{ymh}\right) c_h}}. \tag{5.2.13}$$

By using real data values, the rounded off solution of the above problem is given as

$n_1 = 4,\ n_2 = 16,\ n_3 = 12,\ n_4 = 59,$ with minimum value of the MSE is 5528.854.

Similarly, these techniques may be applied to all discussed estimators and the values of n_h : $h = 1, 2, 3$ and 4 are obtained to calculate MSE for each estimator to provide a comparative study for getting efficient estimation. These values are reported in Tables 5.5 and 5.6.

Table 5.5: The MSE and PRE of proposed and competing estimators for real data obtained by integer programming technique with linear cost function

Population $(N = 448,\ h = 4)$						
	Allocations					
Estimators	n_1	n_2	n_3	n_4	MSE	PRE
$\mathbf{t_{pr1(st)}}$	**4**	**16**	**11**	**60**	**5517.9**	**100**
$t_{s(st)}$	2	12	25	52	23465	425.252
$t_{0(st)}$	4	6	26	54	73791	1337.30
$t_{1(st)}$	3	4	18	62	127370	2308.31
$t_{2(st)}$	3	4	18	62	126830	2298.52
$t_{3(st)}$	3	4	18	62	126830	2298.52
$t_{4(st)}$	3	4	18	62	127010	2301.78
$t_{5(st)}$	2	3	18	63	127100	2303.41
$t_{6(st)}$	3	4	18	62	127090	2303.23
$t_{7(st)}$	3	4	18	62	127200	2305.22
$t_{8(st)}$	3	4	18	62	126810	2298.16
$t_{9(st)}$	3	4	18	62	127260	2306.31
$t_{10(st)}$	2	3	18	63	127100	2303.41
$t_{11(st)}$	3	4	18	62	127220	2305.59
$t_{12(st)}$	3	4	18	62	126430	2291.27
$t_{cr(st)}$	2	2	11	69	15609	282.879
$t_{p(st)}$	3	5	21	59	380750	6900.27
$t_{pe(st)}$	4	5	23	57	196600	3562.95

Table 5.6: The MSE and PRE of proposed and competing estimators for real data obtained by Lagrange multiplier technique with linear cost function

Population $(N = 448,\ h = 4)$						
	Allocations					
Estimators	n_1	n_2	n_3	n_4	MSE	PRE
$\mathbf{t_{pr1(st)}}$	**4**	**16**	**12**	**59**	**5528.854**	**100**
$t_{s(st)}$	3	11	25	52	23423.96	423.667
$t_{0(st)}$	3	5	26	55	73800.80	1334.83
$t_{1(st)}$	3	4	18	62	127368.4	2303.70
$t_{2(st)}$	3	4	18	62	126826.7	2293.90
$t_{3(st)}$	3	4	18	62	126831.3	2293.98
$t_{4(st)}$	3	4	18	62	127169.9	2300.11
$t_{5(st)}$	3	4	18	62	126886.1	2294.98
$t_{6(st)}$	3	4	18	62	127135.3	2299.48
$t_{7(st)}$	3	4	18	62	127196.4	2300.59
$t_{8(st)}$	3	4	18	62	126807.0	2293.55
$t_{9(st)}$	3	4	18	62	127212.3	2300.88
$t_{10(st)}$	3	4	18	62	126816.5	2293.72
$t_{11(st)}$	3	4	18	62	127217.8	2300.97
$t_{12(st)}$	3	4	18	62	126432.4	2286.77
$t_{cr(st)}$	1	2	12	68	15656.06	283.170
$t_{p(st)}$	3	4	22	59	379889.3	6871.03
$t_{pe(st)}$	3	5	23	57	197436.8	3571.03

Simulated data set

A set of 2000 observations for X and Y is generated by using bivariate normal distribution with parameters $(\mu_{xh}, \sigma^2_{xh}, \rho_h)$ and $(\mu_{yh}, \sigma^2_{yh}, \rho_h)$, respectively to carry out simulation study for study variable and auxiliary variable and divided into four strata. For illustration purposes discussed two methodologies which are an integer programming problem and other Lagrange multiplier technique. The summary of simulated data is given in Tables 5.7 and 5.8.

Table 5.7: Obtained results from simulated data from bivariate distribution for study variable (y) and auxiliary variable (x)

h	N_h	$\sum_{h=1}^{L} N_{yh}$	$\sum_{h=1}^{L} N_{xh}$	$\overline{Y}_h$	$\overline{X}_h$	M_{yh}	M_{xh}	S_{yh}^2	S_{xh}^2	c_h
1	1000	109907.7	110916.0	109.9077	110.9160	109.8915	110.923	3.832228	3.942392	2
2	500	150041.2	150057.7	300.0825	300.1155	300.0719	300.1007	4.019246	4.007368	3
3	250	137528.3	137521.1	550.1133	550.0845	550.2232	549.9333	4.139545	4.132717	4
4	250	187528.3	187521.1	750.1133	750.0845	749.9333	749.9333	4.139545	4.143025	5

Table 5.8: Obtained results from simulated data from bivariate distribution for study variable (y) and auxiliary variable (x) (Continue...)

h	S_{yxh}	$S_{m_yh}^2$	$S_{m_xh}^2$	S_{ymh}	S_{xmh}	β_{1xh}	β_{2xh}	G_h	QD_h	ρ_h
1	3.721820	3.979499	3.979512	3.979233	4.015963	-0.014222	3.045211	2.237242	1.3779	0.957524
2	3.903950	4.366777	8.558881	4.366684	8.558662	-0.104107	3.192007	2.272623	1.4331	0.972753
3	3.902696	2.953283	3.611615	2.941148	3.588620	0.0550509	2.905436	2.287887	1.3719	0.943562
4	4.063475	2.953243	3.611575	2.941102	3.588620	-0.093249	2.816044	2.290911	1.3941	0.981211

Integer programming technique

The minimum MSE with fixed linear cost function is obtained for simulated data set by using the following NLPP

$$
\begin{aligned}
\text{Minimize} \quad & \frac{3.832202}{n_1}+\frac{4.01924}{n_2}+\frac{4.139326}{n_3}+\frac{4.139544}{n_4} \\
\text{subject to} \quad & 2n_1+3n_2+4n_3+5n_4 \leq 400 \\
& 2 \leq n_1 \leq N_1 \\
& 2 \leq n_2 \leq N_2 \\
& 2 \leq n_3 \leq N_3 \\
& 2 \leq n_4 \leq N_4 \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ n_4 \text{ are integers.}
\end{aligned}
\tag{5.2.14}
$$

The solution of the above problem is obtained as

$n_1 = 37$, $n_2 = 31$, $n_3 = 27$, $n_4 = 25$ with minimum value of the MSE which equals to 0.5520.

Lagrange multiplier technique

For the proposed estimator $(t_{pr1(st)})$, the MSE is minimized for fixed linear cost function by using Lagrange multiplier technique as discussed in Section 5.2.5. For the simulated data set the values of n_h; $h = 1,2,3,4$, by using (5.2.13), and MSE are obtained as

$n_1 = 37,\ n_2 = 31,\ n_3 = 27,\ n_4 = 24,$ with minimum value of the MSE which equals to 0.552116.

To make a comparative study, both techniques are applied to all discussed estimators and values of n_h; $h = 1,2,3,4$ are obtained to calculate the values of MSE of all estimators for efficient estimation. These values are calculated and given in Tables 5.9 and 5.10.

Table 5.9: The MSE and PRE of proposed and other estimators for simulated data set obtained by integer programming techniques with linear cost function

Population $(N = 2000,\ h = 4)$						
	Allocations					
Estimators	n_1	n_2	n_3	n_4	MSE	PRE
$\mathbf{t_{pr1(st)}}$	**37**	**31**	**27**	**25**	**0.5520**	**100**
$t_{s(st)}$	39	33	27	23	1.0223	185.199
$t_{0(st)}$	37	31	27	25	0.5521	100.018
$t_{1(st)}$	39	30	28	24	0.5911	107.083
$t_{2(st)}$	39	30	28	24	0.5910	107.065
$t_{3(st)}$	36	32	28	24	0.5801	105.091
$t_{4(st)}$	39	30	28	24	0.5911	107.083
$t_{5(st)}$	39	30	28	24	0.5876	106.449
$t_{6(st)}$	39	30	28	24	0.5910	107.065
$t_{7(st)}$	39	30	28	24	0.5899	106.866
$t_{8(st)}$	36	32	28	24	0.5796	105.000
$t_{9(st)}$	39	30	28	24	0.5911	107.083
$t_{10(st)}$	40	35	25	23	0.5877	106.467
$t_{11(st)}$	36	32	28	24	0.5829	105.598
$t_{12(st)}$	37	31	27	25	0.5901	106.902
$t_{p(st)}$	37	31	27	25	2.1825	395.380
$t_{pe(st)}$	37	31	27	25	1.2216	221.304

Table 5.10: The MSE and PRE of different estimators with respect to $t_{pr1(st)}$ for simulated data by Lagrange multiplier technique with linear cost function.

Population ($N = 2000,\ h = 4$)						
	Allocations					
Estimators	n_1	n_2	n_3	n_4	MSE	PRE
$\mathbf{t_{pr1(st)}}$	**37**	**31**	**27**	**24**	**0.552116**	**100**
$t_{s(st)}$	39	33	26	24	1.019755	184.699
$t_{0(st)}$	37	31	27	25	0.552125	100.002
$t_{1(st)}$	38	31	28	24	0.593895	107.567
$t_{2(st)}$	38	31	28	24	0.589313	106.737
$t_{3(st)}$	37	31	28	24	0.581235	105.274
$t_{4(st)}$	38	31	28	24	0.589339	106.742
$t_{5(st)}$	37	31	28	24	0.588761	106.637
$t_{6(st)}$	38	31	28	24	0.593851	107.559
$t_{7(st)}$	38	31	28	24	0.588188	106.533
$t_{8(st)}$	37	31	28	24	0.580809	105.197
$t_{9(st)}$	38	31	28	24	0.589415	106.755
$t_{10(st)}$	40	34	26	23	0.583121	105.616
$t_{11(st)}$	37	31	28	24	0.584033	105.781
$t_{12(st)}$	37	31	27	25	0.590104	106.880
$t_{p(st)}$	37	31	27	25	2.170241	393.077
$t_{pe(st)}$	37	31	27	25	1.221595	221.257

5.3 Efficient estimation with nonlinear cost function

In this section, a ratio estimator for population median is suggested instead of a population mean. The MSE of the suggested estimator is minimized for fixed nonlinear cost function by formulating the problem as NLPP. The solution of the NLPP is also worked out by using appropriate nonlinear programming technique.

5.3.1 Proposed estimator ($t_{pr2(st)}$)

Motivated by Subramani (2016), a ratio type estimator of population mean using the available information on population median is constructed. The proposed estimator is given as,

$$t_{pr2(st)} = \sum_{h=1}^{L} \bar{y}_h \left[1 - K_h log\left(\frac{m_h}{M_h}\right)\right], \tag{5.3.1}$$

where K_h is a constant and is obtained by minimizing the MSE $\left(t_{pr2(st)}\right)$. To examine the properties of the proposed estimator. The proposed estimator may be expressed as

$$\begin{aligned}
t_{pr2(st)} &= \sum_{h=1}^{L} \bar{y}_h \left[1 - K_h log\left(\frac{m_h}{M_h}\right)\right] \\
&= \sum_{h=1}^{L} \overline{Y}_h (1+e_0)\left[1 - K_h log(1+e_1)\right] \\
&= \sum_{h=1}^{L} \overline{Y}_h (1+e_0)\left[1 - K_h\left(e_1 - e_1^2/2 + \ldots\right)\right] \\
&= \sum_{h=1}^{L} \overline{Y}_h \left[(1+e_0) - K_h (1+e_0)\left(e_1 - e_1^2/2 + \ldots\right)\right].
\end{aligned}$$

Ignoring the higher order terms more than two, it may be written as

$$t_{pr2(st)} = \sum_{h=1}^{L} \overline{Y}_h \left[(1+e_0) - K_h\left(e_1 - e_1^2/2 + e_0 e_1\right)\right].$$

Now,

$$\left(t_{pr2(st)} - \overline{Y}_h\right) = \sum_{h=1}^{L} \overline{Y}_h \left[e_0 - K_h\left(e_1 - e_1^2/2 + e_0 e_1\right)\right]. \tag{5.3.2}$$

Taking expectations on both sides of equation (5.3.2), the bias of the proposed estimator may be obtained as

$$B\left(t_{pr2(st)}\right) = E\left[t_{pr2(st)} - \overline{Y}_h\right]$$

$$B\left(t_{pr2(st)}\right) = \sum_{h=1}^{L} \overline{Y}_h \left[K_h\left(\frac{1}{2}\lambda_h C_{mh}^2 - \frac{Bias(m_h)}{M_h} - \lambda_h C_{ymh}\right)\right].$$

From equation (5.3.2), up to an approximation of order one, as given as

$$\left(t_{pr2(st)} - \overline{Y}_h\right)^2 = \sum_{h=1}^{L} \overline{Y}_h^2 \left[e_0 - K_h e_1\right]^2. \tag{5.3.3}$$

Taking expectations on both sides of the equation (5.3.3), the MSE of the proposed estimator may be obtained as

$$
\begin{aligned}
E\left(t_{pr2(st)}-\overline{Y}_h\right)^2 &= \sum_{h=1}^{L}\overline{Y}_h^2\left[E\left(e_0^2\right)+K_h^2E\left(e_1^2\right)-2K_hE\left(e_0e_1\right)\right]\\
&= \sum_{h=1}^{L}\overline{Y}_h^2\left[\lambda_hC_{yh}^2+K_h^2\lambda_hC_{mh}^2-2\lambda_hK_hC_{ymh}\right]\\
&= \sum_{h=1}^{L}\overline{Y}_h^2\lambda_h\left[C_{yh}^2+K_h^2C_{mh}^2-2K_hC_{ymh}\right]
\end{aligned}
$$

$$
MSE\left(t_{pr2(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)\left[C_{yh}^2+K_h^2C_{mh}^2-2K_hC_{ymh}\right] \tag{5.3.4}
$$

Ignoring fpc in equation (5.3.4), we have

$$
MSE\left(t_{pr2(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h^2\frac{1}{n_h}\left[C_{yh}^2+K_h^2C_{mh}^2-2K_hC_{ymh}\right]. \tag{5.3.5}
$$

To estimate the K_h which minimizes the $MSE\left(t_{pr2(st)}\right)$ is obtained by minimizing the $MSE\left(t_{pr2(st)}\right)$ using the principle of maxima and minima. Hence, $\frac{MSE\left(t_{pr2(st)}\right)}{\partial K_h}=0$ gives

$$
K_h=\sum_{h=1}^{L}\frac{C_{ymh}}{C_{mh}^2}.
$$

Now, the minimum MSE will be

$$
MSE\left(t_{pr2(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h^2\frac{1}{n_h}\left[C_{yh}^2-\frac{C_{ymh}^2}{C_{mh}^2}\right]. \tag{5.3.6}
$$

5.3.2 Some special cases of proposed estimator

Case -I: When $K_h=1$ in equation (5.3.1). The proposed estimator becomes

$$
t_{pr2(st)}=\sum_{h=1}^{L}\overline{y}_h\left[1-log\left(\frac{m_h}{M_h}\right)\right],
$$

which is obtained by equation (5.3.1). The bias and MSE will be given as

$$
B\left(t_{pr2(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h\left[\left(\frac{1}{2}\lambda_hC_{mh}^2-\frac{Bias(m_h)}{M_h}-\lambda_hC_{ymh}\right)\right] \tag{5.3.7}
$$

and

$$MSE\left(t_{pr2(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h^2\frac{1}{n_h}\left[C_{yh}^2+C_{ymh}-2C_{ymh}^2\right]. \tag{5.3.8}$$

Case -II: When $K_h=b_{ymh}$ in equation (5.3.1), the proposed estimator is reduced to equation (5.2.1) and given as

$$t_{pr2(st)}=\sum_{h=1}^{L}\overline{y}_h\left[1-b_{ymh}log\left(\frac{m_h}{M_h}\right)\right],$$

where $b_{ymh}=\sum_{h=1}^{L}\frac{s_{ymh}}{s_{mh}^2}$ and the bias is given as

$$B\left(t_{pr2(st)}\right)=\overline{Y}_hB_h\left(\frac{\lambda_h\mu_{21h}}{\sigma_{ymh}M_h}+\frac{1}{2}\lambda_hC_{mh}^2-\frac{\lambda_h\mu_{12h}}{\sigma_{mh}^2M_h}-\lambda_hC_{ymh}-\frac{Bias(m_h)}{M_h}\right). \tag{5.3.9}$$

The equation (5.3.9) may be rewritten as

$$B\left(t_{pr2(st)}\right)=\overline{Y}_hB_h\left(A_h+\frac{1}{2}\lambda_hC_{mh}^2-I_h-\lambda_hC_{ymh}-M_h^*\right), \tag{5.3.10}$$

where

$$E\left(e_1e_2\right)=\lambda_h\frac{\mu_{21h}}{\sigma_{ymh}M_h}=A_h,\ E\left(e_1e_3\right)=\lambda_h\frac{\mu_{12h}}{\sigma_{mh}^2M_h}=I_h,\ \frac{\overline{M}_h-M_h}{M_h}=M_h^*.$$

Therefore, the MSE will be given as

$$MSE\left(t_{pr2(st)}\right)=\sum_{h=1}^{L}\overline{Y}_h^2\frac{1}{n_h}\left[C_{yh}^2+B_h^2C_{mh}^2-2B_hC_{ymh}\right]. \tag{5.3.11}$$

5.3.3 Formulation of problem by using nonlinear cost function

If the strata's travel costs to approach the sampling units are significant, then the cost function remains no linear. Beardwood *et al.* (1959) constructed that the length between n scattered points is proportional to $\sqrt{n}$ and the travel cost may be proportional to the distance traveled. Hence the cost function may be given as

$$C_0 = \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h},$$

where $t_h \sqrt{n_h}$ is the travel cost within h^{th} stratum; $h = 1, 2, \ldots, L$.

In this situation, the minimum MSE for the fixed nonlinear cost function may be obtained by solving the following NLPP which may be formulated as

$$\begin{aligned}
&\text{Minimize} && MSE\left(t_{pr2(st)}\right) \\
&\text{subject to} && \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
& && 2 \leq n_h \leq N_h \\
&\text{and} && n_h \text{ are integers; } h = 1, 2, \ldots, L.
\end{aligned} \tag{5.3.12}$$

5.3.4 Estimators under review

Many estimators of population mean have been given by different authors in the literature for improved estimation. Some estimators have been developed using ratio and product exponential type estimators under large sample approximations. Section 5.2.3 has shown the different estimators of population mean using auxiliary variable. These estimators are given by under simple random sampling which are converted into stratified random sampling in Section 5.2.3.

5.3.5 Efficiency comparison

Here the comparison of the suggested estimator with some other existing estimators mentioned in Section 5.2.3, is discussed. The suggested estimator perform well if it satisfies the condition $MSE\left(t_*\right) - MSE\left(t_{pr2(st)}\right) > 0$, this condition is applied to all other estimators and the differences are calculated and tabulated in Table 5.11. $MSE\left(t_*\right)$ denotes the MSE of the compared estimators.

Table 5.11: Estimators efficiency comparison table for nonlinear cost function

Estimators	Condition	Constant
$t_{0(st)}$	$\sum_{h=1}^{L}\left(\frac{C_{ymh}^2}{C_{mh}^2}\right)>0$	
$t_{cr(st)}$	$\sum_{h=1}^{L}\left(C_{xh}^2-2\rho_h C_{yh}C_{xh}+\frac{C_{ymh}^2}{C_{mh}^2}\right)>0$	
$t_{s(st)}$	$\sum_{h=1}^{L}\left(R_h^2C_{mh}^2-2R_hC_{ymh}+\frac{C_{ymh}^2}{C_{mh}^2}\right)>0$	$R_h=\frac{\overline{Y}_h}{M_h}$
$t_{1(st)}$	$\sum_{h=1}^{L}\left(R_{1h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{1h}=\frac{\overline{Y}_h}{\overline{X}_h}$
$t_{2(st)}$	$\sum_{h=1}^{L}\left(R_{2h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{2h}=\frac{\overline{Y}_h}{\overline{X}_h+C_{xh}}$
$t_{3(st)}$	$\sum_{h=1}^{L}\left(R_{3h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{3h}=\frac{\overline{Y}_h}{\overline{X}_h+\beta_{2h}}$
$t_{4(st)}$	$\sum_{h=1}^{L}\left(R_{4h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{4h}=\frac{\overline{Y}_h\beta_{2h}}{\overline{X}_h\beta_{2h}+C_{xh}}$
$t_{5(st)}$	$\sum_{h=1}^{L}\left(R_{5h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{5h}=\frac{\overline{Y}_hC_{xh}}{\overline{X}_hC_{xh}+\beta_{2h}}$
$t_{6(st)}$	$\sum_{h=1}^{L}\left(R_{6h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{6h}=\frac{\overline{Y}_hC_{xh}}{\overline{X}_hC_{xh}+\rho_h}$
$t_{7(st)}$	$\sum_{h=1}^{L}\left(R_{7h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{7h}=\frac{\overline{Y}_h\rho_h}{\overline{X}_h\rho_h+C_{xh}}$
$t_{8(st)}$	$\sum_{h=1}^{L}\left(R_{8h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{8h}=\frac{\overline{Y}_h\beta_{2h}}{\overline{X}_h\beta_{2h}+\rho_h}$
$t_{9(st)}$	$\sum_{h=1}^{L}\left(R_{9h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{9h}=\frac{\overline{Y}_h\rho_h}{\overline{X}_h\rho_h+\beta_{2h}}$
$t_{10(st)}$	$\sum_{h=1}^{L}\left(R_{10h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{10h}=\frac{\overline{Y}_h}{\overline{X}_h+\beta_{1h}}$
$t_{11(st)}$	$\sum_{h=1}^{L}\left(R_{11h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{11h}=\frac{\overline{Y}_h\beta_{1h}}{\overline{X}_h\beta_{1h}+\beta_{2h}}$
$t_{12(st)}$	$\sum_{h=1}^{L}\left(R_{12h}^2\overline{X}_h^2C_{xh}^2-\overline{Y}_h^2\left(C_{yh}^2\rho_h^2-\frac{C_{ymh}^2}{C_{mh}^2}\right)\right)>0$	$R_{12h}=\frac{\overline{Y}_h}{\overline{X}_h+G_h}$
$t_{pe(st)}$	$\sum_{h=1}^{L}\left(C_{xh}^2\left(1+4K_h^{*}\right)+4\frac{C_{ymh}^2}{C_{mh}^2}\right)>0$	$K_h^{*}=\rho_h\frac{C_{xh}}{C_{yh}}$
$t_{p(st)}$	$\sum_{h=1}^{L}\left(C_{xh}^2\left(1+2K_h^{**}\right)+\frac{C_{ymh}^2}{C_{mh}^2}\right)>0$	$K_h^{**}=\rho_h\frac{C_{xh}}{C_{yh}}$
$t_{pr1(st)}$	$\sum_{h=1}^{L}\left(B_h^2C_{mh}^2-2B_hC_{ymh}+\frac{C_{ymh}^2}{C_{mh}^2}\right)>0$	$B_h=\frac{S_{ymh}}{S_{mh}^2}$

5.3.6 Numerical illustration based on real and simulated data

The following numerical example solved by the integer programming approach for NLPP. For numerical study, we use the data set given in Section 5.2.5. Numerical illustration has been carried out using integer programming using MATLAB. The total amount available for conducting the survey is assumed to be $C = 600$ units with an expected overhead cost $c_0 = 50$ units. This gives $C_0 = C - c_0 = 550$ units.

Integer programming technique with real data set

We obtained the minimum MSE with fixed nonlinear cost function using real data. Then the NLPP may be expressed as

$$\begin{aligned}
\text{Minimize} \quad & \frac{516.3346704}{n_1} + \frac{11099.23291}{n_2} + \frac{7523.589804}{n_3} + \frac{236331.1066}{n_4} \\
\text{subject to} \quad & 2n_1 + 3n_2 + 4n_3 + 5n_4 + \sqrt{n_1} + 1.5\sqrt{n_2} + 2\sqrt{n_3} + 2.5\sqrt{n_4} \leq 550 \\
& 2 \leq n_1 \leq N_1 \\
& 2 \leq n_2 \leq N_2 \\
& 2 \leq n_3 \leq N_3 \\
& 2 \leq n_4 \leq N_4 \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ n_4 \text{ are integers.}
\end{aligned} \tag{5.3.13}$$

The solution of NLPP (5.3.13) may be obtained as

$n_1 = 5,\ n_2 = 22,\ n_3 = 15,\ n_4 = 75,$ with minimum value of the MSE is 4260.4.

Table 5.12: The MSE and PRE of proposed and other estimators for real data set with nonlinear cost function

Population ($N = 2000, h = 4$)						
	Allocations					
Estimators	n_1	n_2	n_3	n_4	MSE	PRE
$\mathbf{t}_{\mathbf{pr2}(\mathbf{st})}$	**5**	**22**	**15**	**75**	**4260.4**	**100**
$t_{s(st)}$	3	14	33	66	18342	430.5230
$t_{0(st)}$	4	7	33	70	57658	1353.347
$t_{1(st)}$	3	4	25	79	99295	2330.650
$t_{2(st)}$	3	4	25	79	99041	2324.688
$t_{3(st)}$	2	4	23	81	99399	2333.091
$t_{4(st)}$	3	4	25	79	99311	2331.025
$t_{5(st)}$	3	4	25	79	99027	2324.359
$t_{6(st)}$	3	4	25	79	99317	2331.166
$t_{7(st)}$	3	4	25	79	99025	2324.312
$t_{8(st)}$	3	4	25	79	99377	2332.574
$t_{9(st)}$	3	4	25	79	99010	2332.960
$t_{10(st)}$	3	4	25	79	99339	2331.682
$t_{11(st)}$	3	4	25	79	98724	2317.247
$t_{12(st)}$	3	4	25	79	60391	1417.496
$t_{cr(st)}$	2	2	15	89	12077	283.4710
$t_{p(st)}$	4	6	29	74	297530	6983.617
$t_{pe(st)}$	4	6	29	74	153480	8936.954
$t_{pr1(st)}$	5	22	15	75	4318.2	101.3570

Integer programming technique with simulated data set

We obtained the minimum MSE with fixed nonlinear cost function. Then the NLPP may be reexpressed as

$$\begin{aligned}
\text{Minimize} \quad & \frac{3.760141}{n_1} + \frac{3.990191}{n_2} + \frac{4.128908}{n_3} + \frac{4.131600}{n_4} \\
\text{subject to} \quad & 2n_1 + 3n_2 + 4n_3 + 5n_4 + \sqrt{n_1} + 1.5\sqrt{n_2} + 2\sqrt{n_3} + 2.5\sqrt{n_4} \leq 550 \\
& 2 \leq n_1 \leq N_1 \\
& 2 \leq n_2 \leq N_2 \\
& 2 \leq n_3 \leq N_3 \\
& 2 \leq n_4 \leq N_4 \\
\text{and} \quad & n_1,\ n_2,\ n_3,\ n_4 \text{ are integers.}
\end{aligned}$$

(5.3.14)

The solution of the NLPP (5.3.14) may be obtained as

$n_1 = 46$, $n_2 = 40$, $n_3 = 35$, $n_4 = 31$ with minimum value of the MSE which equal to 0.4315.

Table 5.13: The MSE and PRE of proposed and other estimators for simulated data set by using nonlinear cost function

Population ($N = 2000,\ h = 4$)						
	Allocations					
Estimators	n_1	n_2	n_3	n_4	MSE	PRE
$\mathbf{t_{pr2(st)}}$	**46**	**40**	**35**	**31**	**0.4315**	**100**
$t_{s(st)}$	50	42	33	30	0.8047	186.489
$t_{0(st)}$	51	40	34	30	0.4354	100.904
$t_{1(st)}$	51	40	34	30	0.4659	107.972
$t_{2(st)}$	51	40	34	30	0.4658	107.949
$t_{3(st)}$	48	41	36	29	0.4573	105.979
$t_{4(st)}$	51	40	34	30	0.4659	107.972
$t_{7(st)}$	51	40	34	30	0.4658	107.949
$t_{8(st)}$	48	41	36	29	0.4650	107.764
$t_{9(st)}$	48	41	36	29	0.4570	105.910
$t_{10(st)}$	51	40	34	30	0.4659	107.972
$t_{11(st)}$	51	43	33	29	0.4626	107.207
$t_{12(st)}$	48	41	36	29	0.4594	106.466
$t_{p(st)}$	51	40	34	30	1.7212	398.888
$t_{pe(st)}$	51	40	34	30	0.9634	223.268
$t_{pr1(st)}$	51	40	34	30	0.4354	100.904

5.3.7 Result

In this chapter the proposed estimators minimize the MSE of the estimators of population parameters. In stratified random sampling, the ratio estimators have been developed using the information of population median for the population mean of the study variable. The proposed estimators are compared with other existing estimators of population parameters with a linear and nonlinear cost function. The theoretical conditions are verified through the numerical examples from real population and the data simulated from bivariate normal distribution. From Tables 5.5, 5.6, 5.9 and 5.10, it can be concluded that proposed estimator $t_{pr1(st)}$ is better than the others competing estimators of Subramani estimator (2016), Unbiased estimator $t_{0(st)}$ and other estimators. From Figures 5.1, 5.2, 5.3 and 5.4, the PRE indicates that the proposed estimator $t_{pr1(st)}$ outperforms the other comparative estimators.

Thus, it may be observed that proposed estimator $t_{pr2(st)}$'s MSE lesser than that of competing existing estimators as given in Tables 5.12 and 5.13. From Figures 5.5 and 5.6, the PRE indicates that the proposed estimator $t_{pr2(st)}$ outperforms the other comparative estimators. Also the proposed estimator $t_{pr2(st)}$ is more efficient than other estimators when they are compared with non linear cost function.

5.4 Conclusion

The theoretical conditions are verified through the numerical examples from the real population and the data simulated from a bivariate normal distribution for proposed estimators $t_{pr1(st)}$ and $t_{pr2(st)}$. Thus, it may be recommended that the proposed estimators may be used for efficient estimation when auxiliary variable is given for both linear and nonlinear cost functions. Hence the aim of achichievity maximum efficiency has fulfilled.

Figure 5.1: PRE of $t_{pr1(st)}$ and other comparative estimators using a real data set for integer programming problem

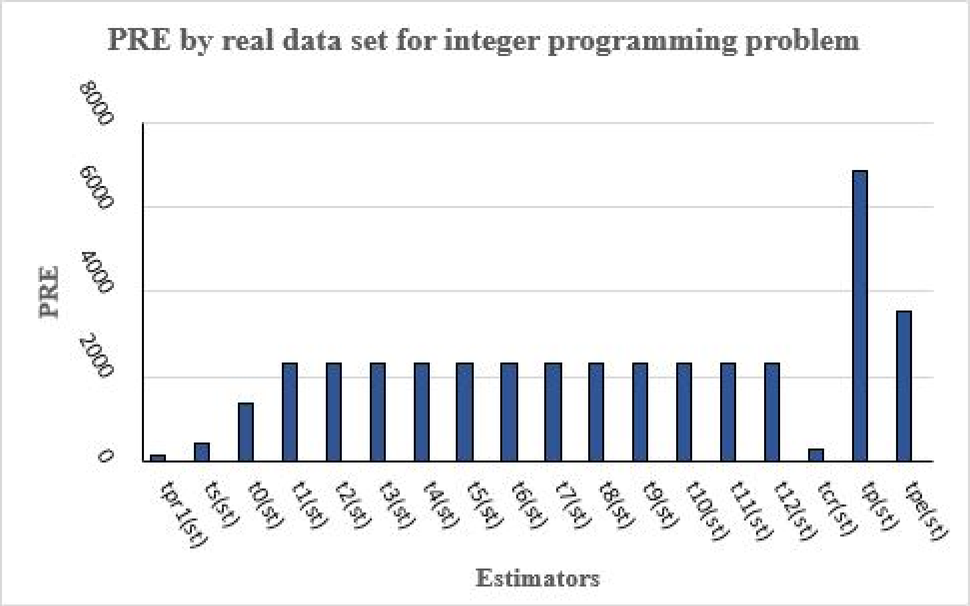


Figure 5.2: PRE of $t_{pr1(st)}$ and other comparative estimators using a real data set for Lagrange multiplier method

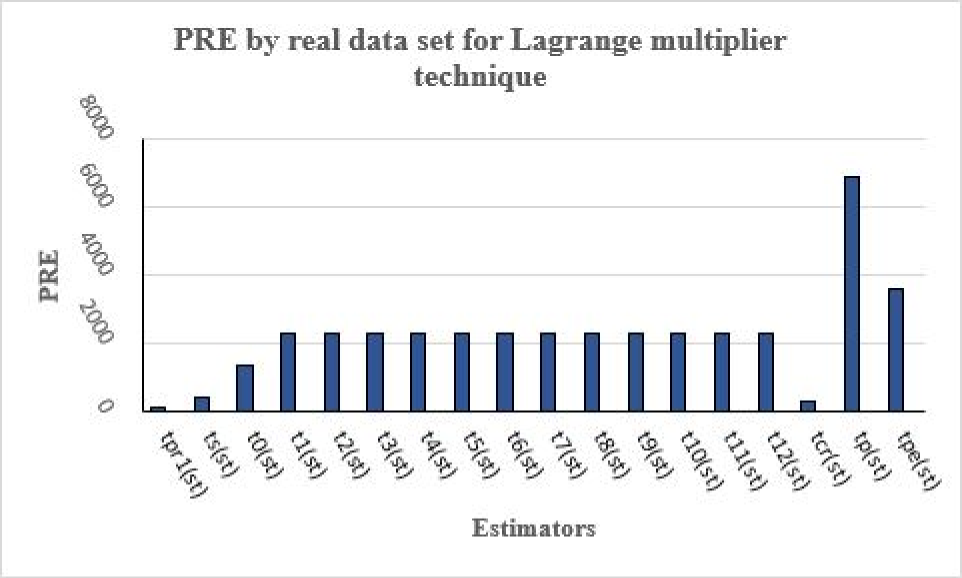


Figure 5.3: PRE of $t_{pr1(st)}$ and other comparative estimators using a simulated data set for integer programming problem

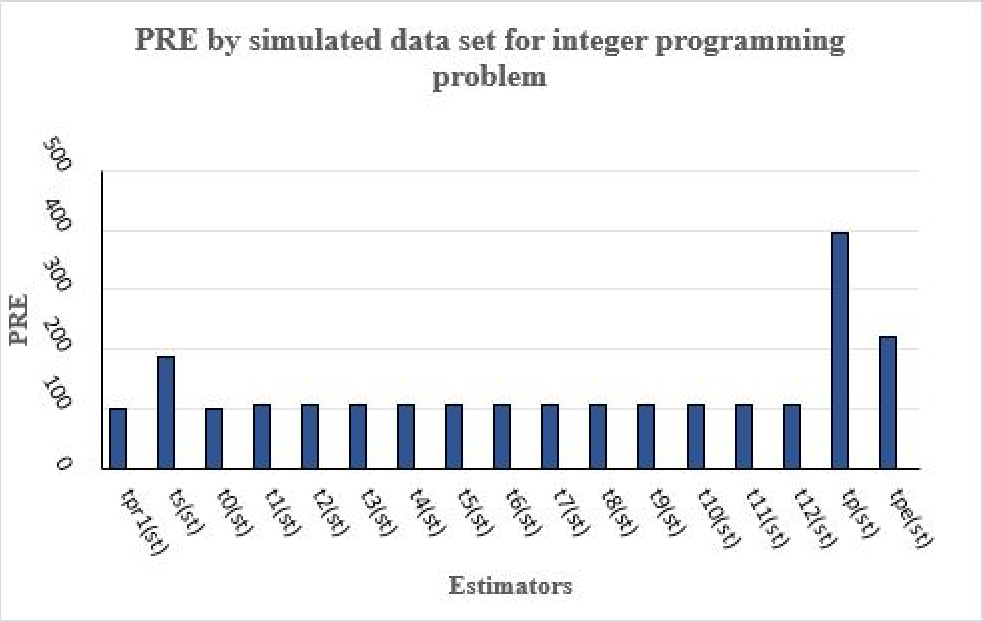


Figure 5.4: PRE of $t_{pr1(st)}$ and other comparative estimators using a simulated data set for Lagrange multiplier method

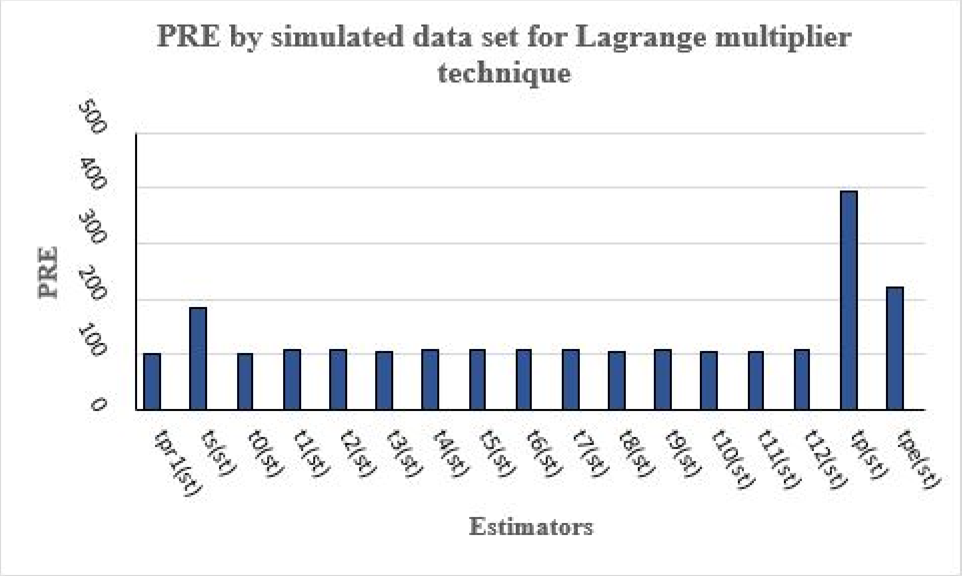


Figure 5.5: PRE of $t_{pr2(st)}$ and other comparative estimators using a real data set for integer programming problem

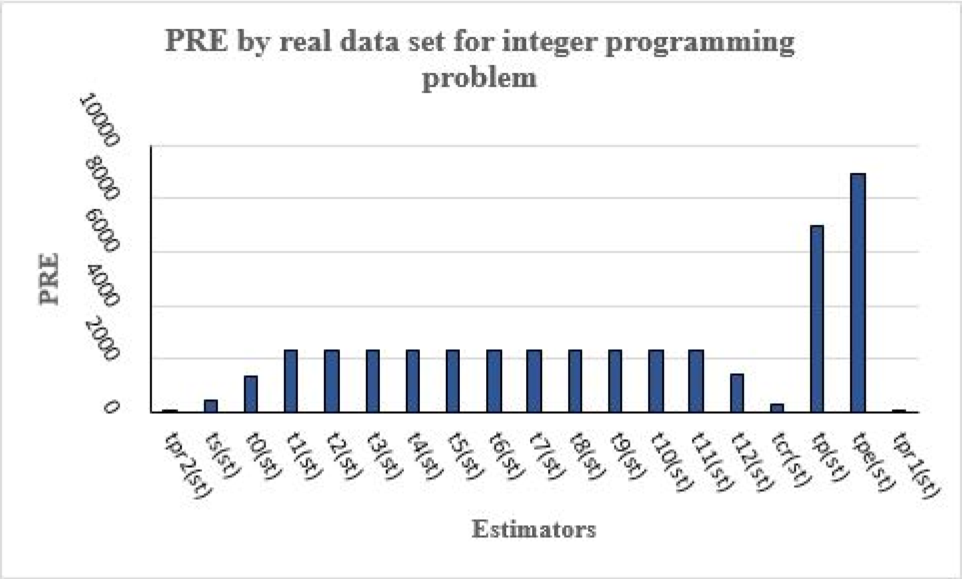
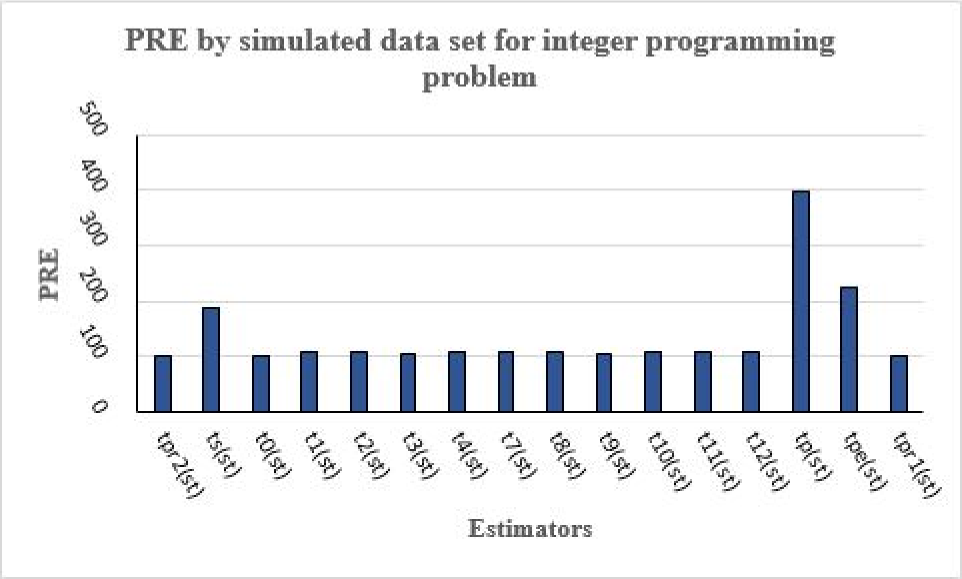


Figure 5.6: PRE of $t_{pr2(st)}$ and other comparative estimators using a simulated data set for integer programming problem



CHAPTER 6

On availability and cost function for maintenance time of continuous operating series-parallel systems using particle swarm optimization

6.1 Introduction

System reliability and availability play an essential role in many functional process, as like industrial systems, power plants, cloud technology structures, telecommunication networks, manufacturing systems, etc. The availability of the system is considered as an important factor for optimization design and preferably for high reliable components. A higher reliability structural design of the system or system components are required to improve the reliability of complicated systems at a higher level. This aim of the problem is to increase the system's availability under some constraints, such as time, weight, cost, etc. Therefore, several studies have been carried out in this area. Cox (1962) has provided the conditions for a finite optimum solution and Barlow *et al.* (1965) found the best solution to minimize the expected costs. Wang (1992) focused on the series-parallel system and Painton and Campbell (1995) dealt with personal computer design in reliability optimization. Klutke *et al.* (1996) derived the availability by an exogenous random environment for an inspected system. Also, they defined a relationship between remaining life, deterioration and repair by the markovian method. One of the most influential factors in availability evaluation is the failure rates of components. In the literature of availability, there are two types of the failure rates of components are considered. One is the constant failure rate and other is the time-dependent failure rate, which are based on lifetime distributions; eg. exponential distribution, Weibull distribution, etc. When the failure rates of components are constant, it is straightforward to obtain the reliability function through statistical and mathematical relations. According to the periodic inspection policy, Kiessler *et al.* (2002) proposed the nonself announcing failures in which a Markov chain governs the rate of deterioration for lifetime distribution. Cui and Xie (2005) carried two models, one is considered when the system is taken to be as good as new after the end of the evaluation or repair, while the other is considered when no maintenance or corrective action is required at the time of the inspection if the system is still running. Also, the state of the system is assumed to be the same as before the inspection. Under the presumption of random repair or replacement time, both models are presented. Several authors have been worked on repairable series-parallel system such as Bris *et al.* (2003), EI-Damcese and Temraz (2010), Ali *et al.* (2011) and Hu *et al.* (2012), etc. After that, Kumar *et al.* (2013) and Kumar and Garg (2016) have been used the semi Markov approach to analyze the steady-state availability of repairable mechanical systems with opportunistic maintenance. Erkoyuncu *et al.* (2017) examined a process to estimate the performance which was based on the support contract costs and attributed to corrective maintenance. Ivezic *et al.* (2019) defined availability by the time coefficient, which includes the time state with a fuzzy expert model. After that Salah and Temraz (2019) used Markov models for availability and reliability function when the failure, imperfect repair and replacement rates are available in general case. Also, they presented a retrial system with mixed standby and unreliable repair facility for availability. In this strategy, Salah and Temraz (2019) obtained mean time to failure of the system by using the Laplace transform technique. Moreover, Mellal *et al.* (2020) has pointed out the cuckoo optimization technique for repairable systems to finding out minimize the cost. They considered the constraint of the availability requirement and involved three aspects as component failure rates, repair rates and redundancy allocation. Further, Gao and Wang (2021) worked on the steady-state availability of the system for general time distributions. Zheng *et al.* (2021) proposed the availability measures for smart electric power grid systems and analyzed parametric sensitivities by using binary decision diagrams for fault trees.

Many authors have been published the work on optimization problems with different techniques. Sung and Cho (2000), Ben-Daya *et al.* (2000) and Osaki (2002) discussed various reliability and maintenance models and shown their optimal approaches. Nourelfath *et al.* (2012) proposed a combined method based on Markov processes, genetic algorithm and universal moment generating function to calculate the availability of the multistate system. Garg *et al.* (2014) provided a methodology to solve the multiobjective reliability optimization model. In their study, the model parameters are considered imprecise in triangular interval data. They converted the uncertain multiobjective optimization model to a deterministic form and used PSO and genetic algorithm to solve these problems. Adhikary *et al.* (2015) used a multiobjective genetic algorithm to solve a series-parallel system with a preventive maintenance (PM) scheduling model that does not provide PM with an off-working time. This chapter proposed availability and cost model for series-parallel systems with components periodically inspected and managed, subject to some maintenance strategy. Our aim is to optimize the maintenance policy for each part of a program and maximizing the availability limit's cost function. The solution procedures are explained by PSO technique. A comparative study is also included by considering some other optimization techniques.

Many techniques are available to solve optimization problems, but PSO technique gives the appropriate, convergent and feasible solutions, as compared to other techniques for this problem. It is computationally inexpensive in terms of time, memory and speed. Due to it's flexibility, we are using PSO technique instead of other techniques. PSO is considered as a potential competitor to other promising techniques as genetic algorithm and goal programming techniques etc. This chapter proposed availability and cost model for series-parallel systems with components periodically inspected and managed, subject to some maintenance strategy. The objective is to optimize the maintenance policy for each part of a system and optimizing the availability limit's cost function. The solution procedures are utilizing by PSO technique. A comparative study has been included by considering some other optimization techniques.

6.2 Problem formulation

6.2.1 Cost function

The cost model is motivated by Nakagawa (2005). This model aims to reduce the average cost per-unit time and the cost of repaired or replaced units and failed units. Suppose each unit has an identical failure distribution of $F(t)$ with a finite mean and the cost of $c_2(< c_1)$ is accumulated for each unit that is transmitted without failure. In addition, let $N_1(T)$ represent the number of failures occurring during $(0,t]$ and $N_2(T)$ represents the number of non failed units occurring within $(0,t]$. So, the estimated cost within $(0,t]$ is given by

$$c_1E(N_1(T)) + c_2E(N_2(T)) = c_1M(T) + c_2.$$

The period of one cycle is considered from one replacement to the next replacement. The time and cost pairs are distributed independently and identically for each cycle and both

Table 6.1: Literature review

Authors	Rates	Description	Solving Method
Cox (1962)	Recurrence times	Probabilistic models of failure and strategies of replacement	Poisson process
Barlow *et al.* (1965)	Constant failure rate	Mathematical model for reliability	Renewal theory process
Wang (1992)	Failures, phased mission systems	Availability demonstration and estimation	Petri nets and dynamic FTA
Painton and Campbell (1995)	Constant failure rate	Computer design model for reliability optimizing	genetic algorithm
Klutke *et al.* (1996)	Constant failure rate	Stationary availability	Markovian method
Sung and Cho (2000)	Constant repair time	Reliability optimization of a series system	Branch and bound algorithm
Sarkar and Sarkar (2000)		Availability of a regularly inspected system under maintenance	Limiting average availability
Kiessler *et al.* (2002)		Non self announcing deficiencies for the rate of degradation	Markov chain control
Bris *et al.* (2003)	Birnbaum importance factor	Minimized under given availability constraint	genetic algorithm
Cui and Xie (2005)	Constant repair rate	Instantaneous and the steady-state availability	
El-Damcese and Temraz (2010)	Markov Models	Availability for repairable parallel systems	Continuous time markov chain
Ali *et al.* (2011)	Constraints as random for repair and replace time	System availability with probabilistic maintenance time	Chance constrained programming
Hu *et al.* (2012)		Equivalence of the availability of multiple designs of a repairable series parallel device	Warm and cold duplication methods
Kumar *et al.* (2013)	Constant rate	Steady-state availability with repairable mechanical systems	Markov model genetic algorithm
Adhikary *et al.* (2015)	Continuous operating series systems	Availability and preventive maintenance scheduling	Multiobjective genetic algorithm
Kaur and Singh (2016)		Minimizes the execution time and maximizes the reliability	BAT algorithm
Ardakan *et al.* (2017)	Time dependent	Reliability of the system	Benchmark problem
Erkoyuncu *et al.* (2017)		Maintenance and availability at the level of equipment form	Benchmarking
Safaei *et al.* (2018)	Three types of failures	Availability and cost functions	
Ivezic *et al.* (2019)		Fuzzy expert model for availability evaluation	Fuzzy model
Salah and Temraz (2019)	General distribution	Availability and reliability of a parallel system	Lagrange multiplier, penalty function methods
Hu *et al.* (2020)	Constant failure rate	Steady state availability of repairable series-parallel system	Markov theory
Mellal *et al.* (2020)	Redundancy allocation, failure, and repair rates	Availability and cost of repairable systems	Efficient cuckoo optimization algorithm
Wang *et al.* (2020)	General time	Availability for general time distribution	Artificial bee colony
Li *et al.* (2020)	Soft and hard failure model	System availability modelling	PSO
Alghazo *et al.* (2020)	Constant failure and repair rate	Availability equivalence and repairable bridge network system	
Gao and Wang (2021)	Constant failure rate	Availability for a stochastic system	Laplace transforms and the Crammer's rule.
Proposed work	Time dependent	Availability and cost function	PSO

have finite mean. So, for an infinite time, the estimated cost per-unit of time is

$$C(T) = \lim_{T\to\infty} \frac{C(T)}{T} = \frac{\text{Expected cost of cycle}}{\text{Mean time of cycle}}.$$

The expected cost rate is

$$C(T) = \frac{1}{T}\left[c_1 M(T) + c_2\right]. \tag{6.2.1}$$

Barlow *et al.* (1965) have established the policy of periodic replacement when period $kT(k = 1, 2, \ldots)$ is still replaced by a cycle but a failure does not replace and thus for failure period the expected cost rate is

$$\frac{c_3}{T}\int_0^{T-T_0} G(t)dt, \tag{6.2.2}$$

where c_3 is downtime cost from a failure to its recognition.

By equations (6.2.1) and (6.2.2) the expected cost rate is

$$C(T_0, T) = \frac{1}{T}\left[c_1 M(T) + c_2 + c_3 \int_0^{T-T_0} G(t)dt\right]. \tag{6.2.3}$$

The equation (6.2.3) is used for age replacement. In this case, the system is regularly replaced at kT and repaired/replaced at the failed unit up to t.

Where, T is the total system time; c_1 be the cost of replacement for a failed unit; c_2 be the cost of the planned replacement.

During the time cycle $E(N_1(T)) = M(T) = \sum_{n=1}^{\infty} F^{(n)}(T)$ represents the mean number of failures over the $(0,t]$ (renewal function) and $F^{(n)}(T)$ convolution of the n-fold lifetime distribution.

$$F^{(n)}(T) = \int_0^t F^{(n-1)}(t-u)dF(u),\ n = 1, 2, \ldots.$$

6.2.2 Cost function for series-parallel system

The number of units in series-parallel system is specific. The independent and identically exponential distribution is used for practical utility. The cost value with a fixed time period is given as

$$C(T_0,T)=\frac{1}{T}\left[c_1M(T)+c_2+c_3\int_0^{T-T_0}G(t)dt\right].$$

Let $F(t)=1-\exp(-\lambda t)$, then

$$M(T)=\sum_{n=1}^{\infty}F^{(n)}(T)=\sum_{n=1}^{\infty}\left[1-\sum_{i=0}^{n-1}\frac{(\lambda t)^i}{i}\right]\exp(-\lambda t).$$

When the unit is replaced only at failure at $T\to\infty$, then $M(T)=T\lambda$.

And value of

$$\int_0^{T-T_0}G(t)dt=\int_0^{T-T_0}\exp(-\mu t)dt,$$

or

$$\int_0^{T-T_0}G(t)dt=\frac{1-\exp(T_0\mu-T\mu)}{\mu}.$$

Then,

$$C(T_0,T)=\frac{1}{T}\left[Tc_1\lambda_i+c_2+c_3\frac{1-\exp(T_0\mu_i-T\mu_i)}{\mu_i}\right].$$

The system consists of n subsystems, connected in series and each subsystem i has m_i components, connected in parallel for $i=1,2,\ldots,n$. Then the cost of the series-parallel system is given as

$$C_s(T_0,T)=\sum_{i=1}^{n}\left[\left(\frac{1}{T}\left[Tc_{1(i)}\lambda_{(i)}+c_{2(i)}+c_{3(i)}\frac{1-\exp\big(T_0\mu_{(i)}-T\mu_{(i)}\big)}{\mu_{(i)}}\right]\right)\right]m_{(i)}. \tag{6.2.4}$$

6.2.3 Availability function

The n subsystems are considered that are connected in series, for $i = 1,2,\ldots,n$ and each i subsystem has m_i parallel connected components. The parallel subsystem works by using the standard series-parallel configuration when at least one of its components are operated and the whole system operates if and only if all subsystems operate.

The components are independent in each subsystem i $(i = 1,2,\ldots,n)$ and distributed identically and independently for failure and repair rate. The X failure rate is considered to be independent and has an equal $F(t)$ distribution with finite mean and the Y repair rate is also independent and has an equal finite mean distribution of $G(t)$. Let $A(T)$ be the availability of the subsystem at time T. Then,

$$A(T) = \frac{E(X_k)}{E(X_k)+E(Y_k)}.$$

Many authors are worked on optimum PM such as Barlow *et al.* (1965), Nakagawa (2005) and others. This model is similar to Nakagawa (2005), which describes the assumptions of system repair and failure problem. If a device fails, it undergoes under maintenance immediately and is restored to the operational state after repaired. Repair time is divided into two parts; one is before time T having Y_1 distribution. The repair time is independent and has a finite mean $G_1(t)$. If a unit's operating time is already known and its failure rate rises over time, it might be prudent to maintain it at the time T preventively until its operating time failure. Other repair time is after time T having distribution of Y_2. Time distribution of Y_2 to PM completion time distribution is $G_2(t)$ with a finite mean, which could be lower than the repair time of Y_1. A new unit begins to work at $t = 0$. We describe one process from the beginning of service until PM or repair is complete. Then the loss of one cycle is meanwhile.

$$E\left(Y_1 I_{(X<T)} + Y_2 I_{(X\geq T)}\right) = \frac{F(T)}{\mu_1} + \frac{\overline{F}(T)}{\mu_2},$$

where X_k and Y_k $(k = 1,2,\ldots)$ as referring to uptime and downtime and $A(T)$ is the likelihood that the device will work at the time of T, respectively. Hence the availability of the state is

$$A(T) = \frac{\int_0^T \overline{F}(T)dt}{\int_0^T \overline{F}(T)dt + \frac{F(T)}{\mu_1} + \frac{\overline{F}(T)}{\mu_2}}.$$

Now, this model is defined as a series-parallel system. Let A_{ij} be the availability of the component j $(j = 1,2,\ldots,m_i)$ in subsystem i $(i = 1,2,\ldots,n)$ and let A_i be the availability of the subsystem i. That is A_{ij} and A_i can be expressed for series-parallel system, such as

$$A_i = 1 - \prod_{j=1}^{m_i} \left(1 - A_{ij}\right).$$

The system consists of n subsystems connected in series and each i subsystem has components of m_i, connected in parallel to $i = 1,2,\ldots,n$. Subsystems are related in sequence, then the system availability is

$$A_s(T) = \prod_{i=1}^{n} 1 - \left(1 - A_{ij}\right)^{m_i}.$$

Suppose failure and repair time follows an exponential distribution with different parameters. Then the availability of the series-parallel system is given as

$$A_s(T) = \prod_{i=1}^{n} \left[1 - \left(1 - \frac{\frac{1-\exp(-\lambda_i T)}{T}}{\frac{1-\exp(-\lambda_i T)}{T} + \frac{1-\exp(-\mu_{1i} T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i} T)}{\mu_{2i}}} \right)^{m_i} \right],$$

or

$$A_s(T) = \prod_{i=1}^{n} \left[1 - \left(\frac{\frac{1-\exp(-\mu_{1i} T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i} T)}{\mu_{2i}}}{\frac{1-\exp(-\lambda_i T)}{T} + \frac{1-\exp(-\mu_{1i} T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i} T)}{\mu_{2i}}} \right)^{m_i} \right].$$

It has been assumed that the time negligible for replaced at periodic times and at any time, there is an unlimited supply of units available for replacement.

6.2.4 Optimization model

To solve such problems, the optimization model of a system may be formulated. So, as to maximize availability and minimize cost function during given time period. It may be formulated as follows:

$$\begin{aligned}
&\text{Maximize} && A_s(T)\\
&\text{minimize} && C_s(T_0,T)\\
&\text{subject to} && A_s(T) \geq WRV(A_0)\\
& && C_s(T_0,T) \leq C_{max}\\
& && T_s \leq T_{max}\\
&\text{and} && T_{max} \geq 0,\ A_0 \geq 0,\ C_{max} \geq 0,\ C_s(T_0,T) \geq 0,
\end{aligned} \tag{6.2.5}$$

where $A_s(T)$ is the system availability at time T; $C_s(T_0,T)$ is the cost value at time T; WRV is the worst reliability value of the system; C_{max} is the maximum allowable system cost; T_s is the system time, and T_{max} is the total maximum time.

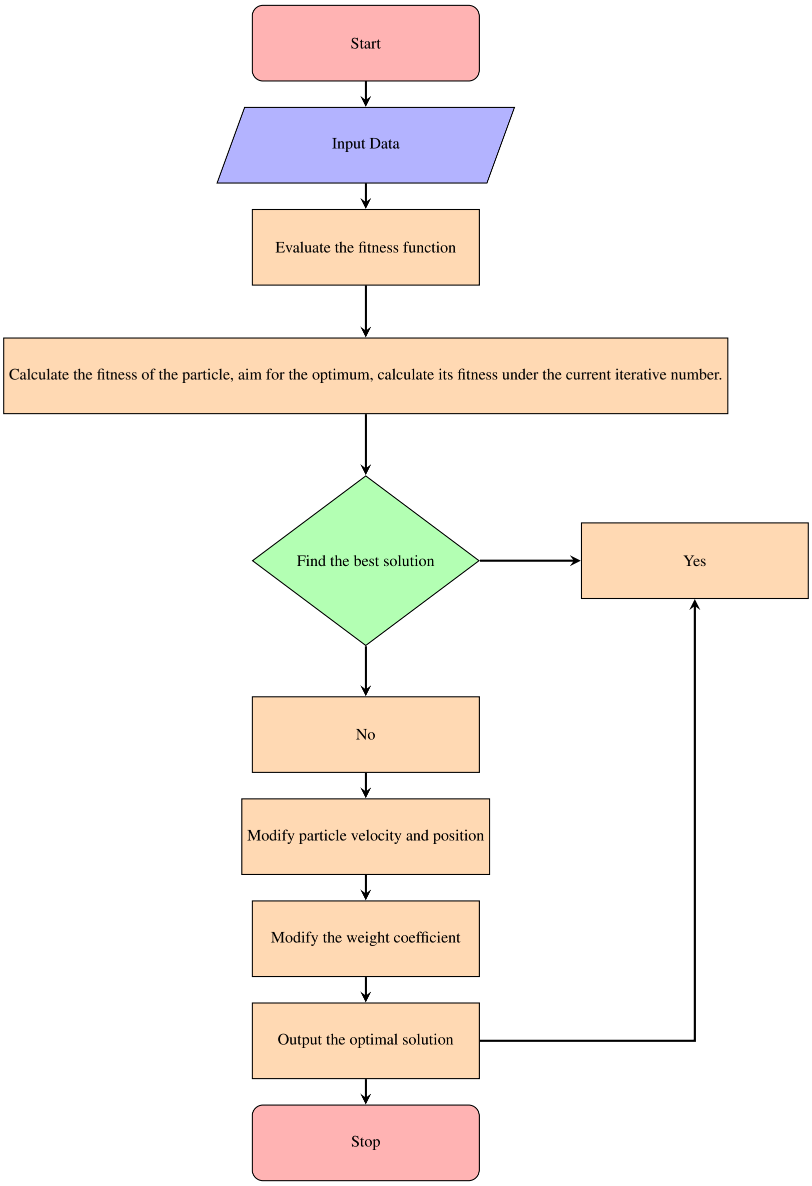
6.3 Solution methods

This section has considered the problem that can be solved using the PSO technique as well as some other existing optimization methods, which are mentioned below.

6.3.1 Particle swarm optimization

Recall that, as we explained in the introduction, the PSO technique is an optimization technique. The PSO is introduced by Kennedy and Eberhart (1995) a population oriented stochastic optimization technique. The flow chart of PSO technique is shown in Figure 6.1.

Figure 6.1: PSO flow chart



6.3.2 Genetic algorithm

In introduction section it is discussed that genetic algorithm is the most widely used optimization technique and first implemented at Michigan University by Holland (1975). In this algorithm, the better population is combined according to the principle of survival of the best and this solution will be repeated more often in the next generation depending on the suitability of each solution. This process will reach an optimal solution.

Algorithm steps

Step 1: Create a random population, including n chromosome or initial solution.

Step 2: Establish in the population the fitness role of each chromosome.

Step 3: Building a new population-based on the selection of parent chromosomes by selective methods such as roulette wheel, match, random, competitive, etc.

Mentioning an absolute value for the likelihood of a crossover operator and then conducting a combination procedure on parents to create offspring and assuming a specific value for the mutation operator's. Using this procedure into establish a new chromosome shift one or more genes from parent's chromosome.

Step 4: Replacing new offspring in the new population.

6.3.3 Goal programming problem

Goal programming is an efficient and feasible methodology which can be applicable to a variety of targets in decision making problems. Charnes *et al.* (1968) developed goal programming and appeared to be a commanding methodology to overcome decision-making problems with multi-criteria. The goal is to minimize the variation between the successes of objectives. For all the p functions, the problem can be reported separately as

$$\begin{aligned} \text{Minimize} \quad & Z_1,\ Z_2,\ \ldots, Z_p \\ \text{subject to} \quad & A_s(T) \geq WRV(A_0) \\ & C_s(T_0,T) \leq C_{max} \\ & T_s \leq T_{max} \\ \text{and} \quad & T_{max} \geq 0,\ A_0 \geq 0,\ C_{max} \geq 0,\ C_s(T_0,T) \geq 0, \end{aligned} \tag{6.3.1}$$

where

$$Z_1 = C_s(T_0,T) = \frac{1}{T}\left[c_1 M(T) + c_2 + c_3 \int_0^{T-T_0} G(t)dt\right]$$

and

$$Z_2 = A_s(T) = \prod_{i=1}^{n} \left[1 - \left(\frac{\frac{1-\exp(-\mu_{1i}T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i}T)}{\mu_{2i}}}{\frac{1-\exp(-\lambda_i T)}{T} + \frac{1-\exp(-\mu_{1i}T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i}T)}{\mu_{2i}}} \right)^{m_i} \right].$$

By using on NLPP, find out the optimal value of Z_j^* of Z_j is given as

$$\begin{aligned} \text{Minimize} \quad & Z_j,\ j = 1,2,\dots,p \\ \text{subject to} \quad & A_s(T) \geq WRV(A_0) \\ & C_s(T_0,T) \leq C_{max} \\ & T_s \leq T_{max} \\ \text{and} \quad & T_{max} \geq 0,\ A_0 \geq 0,\ C_{max} \geq 0,\ C_s(T_0,T) \geq 0. \end{aligned} \tag{6.3.2}$$

Obviously, $Z_j \geq Z_j^*$ and $Z_j - Z_j^* \geq 0;\ j = 1,2,\dots,p$ would raise the cost because the individual optimal value are not used for the j^{th} characteristics. According to the use of compromise allocation, find the importance of the availability and cost function for each characteristic, instead of the individual optimum allocation. Value of m_i does not exceed x_j, where x_j are the unknown goal variables $(x_j,\ j = 1,2,\dots,p)$.

To obtain these goals, m_i must satisfy,

$$Z_j - Z_j^* \leq x_j;\ j = 1,2,\dots,p.$$

The value $\sum_{j=1}^{p} x_j$ will give us the total increases in objective values by not using the individual allocations. To solve the following goal programming problem

$$\begin{aligned} \text{Minimize} \quad & \sum_{j=1}^{p} x_j \\ \text{subject to} \quad & A_s(T) - x_1 \geq Z_1^* \\ & C_s(T_0,T) - x_2 \leq Z_2^* \\ & T_s \leq T_{max} \\ \text{and} \quad & T_{max} \geq 0,\ A_0 \geq 0,\ C_{max} \geq 0,\ C_s(T_0,T) \geq 0. \end{aligned}$$

6.4 Numerical study

For discussing the above techniques numerically, the data are simulated by using exponential distribution using R software and information regarding the parameters has been summaries in Table 6.2 with cost and time values. There is need to define optimal values of several components and cost value in a series-parallel system. The problem is considered as a multiobjective optimization model. It has two objective functions: first is to maximize the availability and second is to minimize the cost function. To solve such complex problem, it can be divided into two groups:

(i) Weighted sum approach by assigning weights to each feature to turn the multiobjective issue into a single objective problem.

(ii) Optimize goal function with some constraints. The objective functions are transformed as follows to a single objective function.

$$
\begin{aligned}
&\text{Minimize } f(A_s, C_s) = 0.5\sum_{i=1}^{n}\left[\left(\frac{1}{T}\left[Tc_{1(i)}\lambda_{(i)} + c_{2(i)} + c_{3(i)}\frac{1-\exp\big(T_0\mu_{(i)} - T\mu_{(i)}\big)}{\mu_{(i)}}\right]\right)\right]m_{(i)} \\
&\qquad - 0.5\prod_{i=1}^{n}\left[1 - \left(\frac{\frac{1-\exp(-\mu_{1i}T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i}T)}{\mu_{2i}}}{\frac{1-\exp(-\lambda_i T)}{T} + \frac{1-\exp(-\mu_{1i}T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i}T)}{\mu_{2i}}}\right)^{m_i}\right] \\
&\text{subject to} \\
&\qquad \prod_{i=1}^{n}\left[1 - \left(\frac{\frac{1-\exp(-\mu_{1i}T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i}T)}{\mu_{2i}}}{\frac{1-\exp(-\lambda_i T)}{T} + \frac{1-\exp(-\mu_{1i}T)}{\mu_{1i}} + \frac{1-\exp(-\mu_{2i}T)}{\mu_{2i}}}\right)^{m_i}\right] \geq WRV(A_0) \\
&\qquad \sum_{i=1}^{n}\left[\left(\frac{1}{T}\left[Tc_{1(i)}\lambda_{(i)} + c_{2(i)} + c_{3(i)}\frac{1-\exp\big(T_0\mu_{(i)} - T\mu_{(i)}\big)}{\mu_{(i)}}\right]\right)\right]m_{(i)} \leq C_{max} \\
&\qquad T_s \leq T_{max} \\
&\text{and} \qquad T_{max} \geq 0,\ A_0 \geq 0,\ C_{max} \geq 0.
\end{aligned}
\tag{6.4.1}
$$

On using values of Table 6.2, we have solved the model (6.4.1) numerically and used PSO technique to find out optimum solution. Also, we have solved the model (6.2.5), using different optimization techniques, defined in Section 6.3 and obtained the optimum solution which summaries in Table 6.3. On the basis of these results, three different existing optimization techniques are considered to solve the numerical illustration. On comparing the results, it can be observed that the PSO technique has outperformed compare to other techniques in terms of availability and cost value. Table 6.3 shows that all other techniques are not advisable than the PSO algorithm, even goal programming problem has an equal availability value but violates the cost constraint. Whereas genetic algorithm provides less cost value than others and gives less availability. From Figure 6.2, it is easy to see that as time increases, availability value decreases.

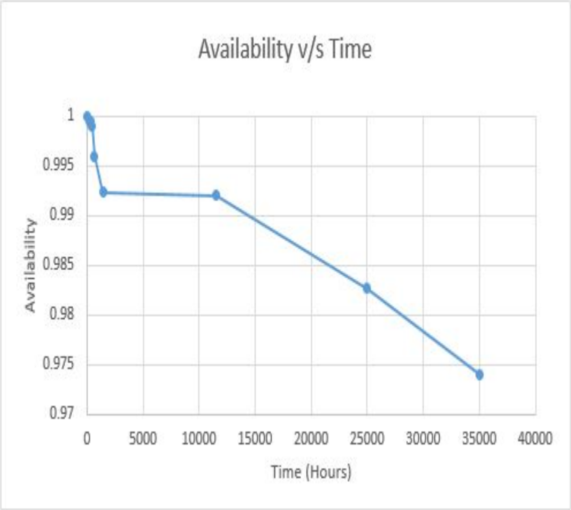
Table 6.2: Data simulation by R software from exponential distribution for different parameters

i	λ_i	μ_i	μ_{1i}	μ_{2i}	c_{1i}	c_{2i}	c_{3i}	T (Hours)
1	1.78881239	5.4037288890	0.3756002	1.849384	2	1	3	1350
2	0.01090341	0.6290124455	0.2803815	1.442109	3	2	4	1350
3	0.09763477	6.4458548851	1.1246886	0.182228	4	3	5	1350
4	1.93698614	1.9918158221	2.7208015	0.546778	5	4	6	1350
5	0.27380822	0.2120789000	2.6440963	1.093491	6	5	7	1350

Table 6.3: Results for the series parallel system for availability $A_0 = 0.85$ and $C_s = 150$

Methods	$(m_1,\ m_2,\ m_3,\ m_4,\ m_5)$	A_s	C_s
Particle swarm optimization	(1, 5, 1, 1, 4)	0.9923	146.2646
Genetic algorithm	(2, 1, 1, 4, 2)	0.9827	136.6990
Goal programming problem	(1, 3, 3, 2, 3)	0.9923	153.8694

Figure 6.2: Availability v/s Time by PSO technique



6.5 Statistical analysis

The t- test on genetic algorithm and goal programming with PSO technique are applied. Assume the null hypothesis is that there is no difference in their population means and populations have equal variances at the significance level of $\alpha = 0.05$. It is carried out by the pooled t -test for null hypothesis as given in Tables 6.4 and 6.5. It may be easily seen that the testing for equality of variance and values of t-statistic are greater than t-critical values. Also, p-value is less than the significance level $\alpha = 0.05$. Thus, it may be concluded that two types of means differ significantly. Since variance function value is minimum for proposed PSO technique, so it may be considered as best in comparison to other techniques and hence, this difference is statistically significant.

Table 6.4: t-test: two sample assuming equal variances for cost function

	Genetic algorithm	Goal programming	PSO technique
Mean	190.8586	152.3006	149.8492
Variance	755.3194	26.93697	3.43604
Observations	25	25	25
Pooled variance	379.0042	14.81299	
Hypothesized mean difference	0	0	
Degree of freedom	48	48	
t statistic	7.339641	1.70581	
P(T<=t) one-tail	1.11×10^{-9}	0.047254	
t critical one-tail	1.677224	1.677224	

Table 6.5: t-test: two sample assuming equal variances for availability function

	Genetic algorithm	Goal programming	PSO technique
Mean	0.99630	0.988516	0.989245
Variance	5.86×10^{-6}	5.53×10^{-6}	2.59×10^{-6}
Observations	25	25	25
Pooled variance	4.23×10^{-6}	4.06×10^{-6}	
Hypothesized mean difference	0	0	
Degree of freedom	48	48	
t statistic	12.19585	1.87285	
P(T<=t) one-tail	1.3×10^{-16}	0.033593	
t critical one-tail	1.677224	1.677224	

6.6 Conclusion

This chapter has proposed the multiobjective system availability optimization problem for the continuous operating series-parallel system. For effective planning of scheduling, both the availability and cost functions are to be considered simultaneously. The case study shows that the model increases availability while decreases maintenance costs. Also, the present model gives the best results when the problem is taken as multiobjective formulation. Hence, on the basis of comparisons, made in Section 6.4, the study justifies that the proposed model may improve availability and reduce maintenance costs by using the PSO technique, which is not granted with other techniques at the points of availability and costs. Comparatively, the PSO technique demonstrates the effective and efficient result for global solutions with other optimization approaches. From statistical analysis, point a view it may be easily seen that the variance of design availability and cost by PSO are minimum as compared to others techniques. Thus, the PSO technique may be used by using pooled t -test statistic.

Bibliography

Abid, M., Abbas, N., Sherwani, R. A. K., and Nazir, H. Z. (2016). Improved ratio estimators for the population mean using non-conventional measure of dispersion. *Pakistan Journal of Statistics and Operations Research*, 12 (2), 353–367.

Adhikary, D. D., Bose, G. K., Jana, D. K., Bose, D., and Mitra, S. (2015). Availability and cost-centered preventive maintenance scheduling of continuous operating series systems using multi-objective genetic algorithm: A case study. *Quality Engineering*, 28 (3), 1–6.

Aeronautical, R. I. (1994). *Reliability engineering*. Prentice Hall, Englewood Cliffs, NJ.

Ahmed, M. S., Raman, M. S., and Hossain, M. I. (2000). Some competitive estimators of finite population variance using multivariate auxiliary information. *International Journal of Information and Management Sciences*, 11 (1), 49–54.

Ahsan, M. J. (1975-1976). A procedure for the problem of optimum allocation in multivariate stratified random sampling stratified random sampling. *Aligarh Bulletin of Mathematics*, 5-6, 37–42.

Ahsan, M. J. (1978). Allocation problem in multivariate stratified random sampling. *Journal of the Indian Statistical Association*, 16 (3), 1–5.

Ahsan, M. J. and Khan, S. U. (1982). Optimum allocation in multivariate stratified random sampling with overhead cost. *Metrika: International Journal for Theoretical and Applied Statistics*, 29 (1), 71–78.

Alam, M. N. (2016). Particle swarm optimization: algorithm and its codes in MATLAB, 1–10. DOI: `10.13140/RG.2.1.4985.3206`.

Alghazo, J. M., Mustafa, A., and El-Faheem, A. A. (2020). Availability equivalence analysis for the simulation of repairable bridge network system. *Complexity*. DOI: `https://doi.org/10.1155/2020/4907895`.

Ali, I. and Hasan, S. S. (2014). Integer fuzzy programming approach in bi-objective selective maintenance allocation problem. *Journal of Mathematical Modelling and Algorithms in Operations Research*, 13 (2), 113–124.

Ali, I., Khan, M. F., Raghav, Y. S., and Bari, A. (2011). Allocation of repairable and replaceable components for a system availability using selective maintenance with probabilistic maintenance time constraints. *American Journal of Operations Research*, 1 (3), 147–154.

Ali, I., Raghav, Y. S., and Bari, A. (2013a). Compromise allocation in multivariate stratified surveys with stochastic quadratic cost function. *Journal of Statistical Computation and Simulation*, 83 (5), 962–976.

Ali, I., Raghav, Y. S., Khan, M. F., and Bari, A. (2013b). Selective maintenance in system reliability with random costs of repairing and replacing the components. *Communication in Statistics- Simulation and Computation*, 42 (9), 2026–2039.

Ali, I., Shafiullah., and Haseen, S. (2015). Geometric programming approach in multivariate stratified sample surveys in case of non-response. *Electronic Journal of Applied Statistical Analysis*, 8 (1), 28–43.

Ansari, A. H., Najmussehar., and Ahsan, M. J. (2009). On multiple response stratified random sampling design. *International Journal of Statistical Sciences*, 1 (1), 45–54.

Ansari, A. H., Varshney, R., and Ahsan, M. (2018). Compromise mixed allocation in multivariate stratified sampling using dynamic programming technique. *Journal of Advanced Statistics*, 3 (4), 71–79.

Ardakan, M. A., Mirzaei, Z., Hamadani, A. Z., and Elsayed, E. A. (2017). Reliability optimization by considering time-dependent reliability for components. *Quality and Reliability Engineering International*, 33 (8), 1641–1654.

Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12 (1), 159–164.

Bakhshi, Z. H., Khan, M. F., and Ahmad, Q. S. (2010). Optimal sample numbers in multivariate stratified sampling with a probabilistic cost constraint. *International Journal of Mathematics and Applied Statistics*, 1 (2), 111–120.

Barlow, R. E., Proschan, F., and Hunter, L. C. (1965). *Mathematical theory of reliability*. New York: John Wiley & Sons.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55 (4), 299–327.

Bellman, R. E. and Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management Science*, 17 (4), 141–164.

Ben-Daya, M., Duffuaa, S. O., and Raouf, A. (2000). *Maintenance, modeling and optimization*. Boston: Kluwer Academic Publishers.

Bowley, A. L. (1926). Measurements of precision attained in sampling. *Bulletin of the International Statistical Institute*, 22, 1–62.

Bris, R., Châtelet, E., and Yalaoui, F. (2003). New method to minimize the preventive maintenance cost of series-parallel systems. *Reliability Engineering & System Safety*, 82, 247–255.

Bühler, W. and Deutler, T. (1975). Optimal stratification and grouping by dynamic programming. *Metrika*, 22 (1), 161–175.

Charnes, A. and Cooper, W. W. (1963). Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research*, 11 (1), 18–39.

Charnes, A., Cooper, W. W., and Ferguson, R. O. (1955). Optimal estimation of executive compensation by linear programming. *Management Science*, 1 (2), 138–151.

Charnes, A., Cooper, W. W., DeVoe, J. K., Learner, D. B., and Reinecke, W. (1968). A Goal programming model for media planning. *Management Science*, 14, 423–430.

Chatterjee, S. (1966). *A programming algorithm and its statistical applications*. Office of Naval Research, Technical Report 1, Department of Statistics, Harvard University, Cambridge.

Chatterjee, S. (1967). A note on optimum allocation. *Scandinavian Actuarial Journal*, 50 (1), 40–44.

Chatterjee, S. (1968). Multivariate stratified surveys. *Journal of the American Statistical Association*, 63 (322), 530–534.

Chern, M. and Jan, R. (1986). Reliability optimization problems with multiple constraints. *IEEE Transactions on Reliability*, 35 (4), 431–436.

Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science*, 30 (2), 262–275.

Cochran, W. G. (1963). *Sampling techniques*. John Wiley, New York, 2nd edition.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley, New York, 3rd edition.

Cox, D. R. (1962). Renewal theory. *London, Methuen; New York, Wiley*, 93 (4), 550–556.

Cui, L. and Xie, M. (2005). Availability of a periodically inspected system with random repair or replacement times. *Journal of Statistical Planning and Inference*, 131, 89–100.

Dalenius, T. (1950). The problem of optimum stratification. *Scandinavian Actuarial Journal*, 33 (3–4), 203–213.

Dalenius, T. (1957). *Sampling in Sweden: contributions to themethods and theories of sample survey practice*. Stockholm: Almqvist OchWiksell.

Dalenius, T. and Gurney, M. (1951). The problem of optimum stratifiaction. II. *Scandinavian Actuarial Journal*, 34 (1–2), 133–148.

Dalenius, T. and Hodges, J. L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54 (285), 88–101.

Danish, F. (2018). A mathematical programming approach for obtaining optimum strata boundaries using two auxiliary variables under proportional allocation. *Statistics in Transition, New Series*, 19 (3), 507–526.

Danish, F., Rizvi, S. E. H., Sharma, M. K., and Jeelani, M. I. (2017). Optimum stratification using mathematical programming approach: a review. *Journal of Statistics Applications & Probability Letters*, 4 (3), 123–129.

Danish, F., Rizvi, S. E. H., Sharma, M. K., Dwivedi, S., Kumar, B., and Kumar, S. (2019a). A mathematical programming approach in optimum stratification under Neyman allocation for two stratifying variables. *Journal of Reliability and Statistical Studies*, 12 (2), 173–185.

Danish, F., Rizvi, S. E. H., Sharma, M. K., Jeelani, M. I., Kumar, B., and Dar, Q. F. (2019b). Optimum stratification for two stratifying variables. *Revista Investigacion Operacional*, 40 (5), 562–573.

Dantzig, G. B. (1953). The multivariate sampling problem. *Skandinavisk Aktuarietidskrift*, 36, 92–102.

Dantzig, G. B. (1977). A prosedure for integer solution to some allocation problems. *Pure and Applied Mathematika Sciences*, 5, 25–32.

Das, A. K. and Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya C*, 40, 139–148.

Daw, R. H. and Pearson, E. S. (1972). Studies in the history of probablity and statistics. XXX. Abraham De Moivre's 1733 derivation of the normal curve: a bibliographical note. *Biometrika*, 59 (3), 677–680.

Díaz-García, J. A. and Cortez, L. U. (2006). Optimum allocation in multivariate stratified sampling: multio-bjective programming. *Comunicación Técnica No I-06-07/28-03-206 (PE/CIMAT)*.

Díaz-García, J. A. and Garay Tapia, M. M. (2007). Optimum allocation in stratified surveys: stochastic programming. *Computational Statistics and Data Analysis*, 51, 3016–3026.

Díaz-García, J. A. and Ramos-Quiroga, R. (2014). Optimum allocation in multivariate stratified random sampling: a modified Prékopa's approach. *Journal of Mathematical Modelling and Algorithms*, 13 (3), 315–330.

EI-Damcese, M. A. and Temraz, N. S. (2010). Analysis of availability and reliability for repairable parallel systems with different failures rates. *Asian Journal of Information Technology*, 9 (4), 231–237.

Ekman, G. (1959). An approximation useful in univariate stratification. *The Annals of Mathematical Statistics*, 30 (1), 219–229.

El-badry, M. A. (1956). A sampling procedure for mailed questionnaires. *Journal of the American Statistical Association*, 51 (274), 209–227.

Elsisi, M. (2019a). Design of neural network predictive controller based on imperialist competitive algorithm for automatic voltage regulator. *Neural Computing and Applications*, 31 (9), 5017–5027.

Elsisi, M. (2019b). Future search algorithm for optimization. *Evolutionary Intelligence*, 12 (1), 21–31.

Elsisi, M. (2020). New variable structure control based on different meta-heuristics algorithms for frequency regulation considering nonlinearities effects. *International Transactions on Electrical Energy Systems*, 30 (7), 5017–5027.

Elsisi, M. and Soliman, M. (2021). Optimal design of robust resilient automatic voltage regulators. *ISA transactions*, 108, 257–268.

Ericson, W. A. (1970). On the posterior mean and variance of a population mean. *Journal of American Statistical Association*, 65 (330), 649–652.

Erkoyuncu, J. A., Khan, S., Eiroa, A. L., Butler, N., Rushton, K., and Brocklebank, S. (2017). Perspectives on trading cost and availability for corrective maintenance at the equipment type level. *Reliability Engineering and System Safety*, 168, 53–69.

Fakhrzad, M. B. and Goodarzian, F. (2019). A Fuzzy multi-objective programming approach to develop a green closed-loop supply chain network design problem under uncertainty: Modifications of imperialist competitive algorithm. *RAIRO Operations Research*, 53 (3), 963–990.

Fallahnezhad, M. S. and Najafian, E. (2017). A model of preventive maintenance for parallel, series, and single-item replacement systems based on statistical analysis. *Communications in Statistics - Simulation and Computation*, 46 (7), 5846–5859.

Fathollahi-Fard, M., A., Ahmadi, A., Goodarzian, F., and Cheikhrouhou, N. (2020). A bi-objective home healthcare routing and scheduling problem considering patients' satisfaction in a fuzzy environment. *Applied Soft Computing*, 93. DOI: `https://doi.org/10.1016/j.asoc.2020.106385`.

Fatima, U., Varshney, R., Najmussehar, S., and Ahsan, M. J. (2014). On compromise mixed allocation in multivariate stratified sampling with random parameters. *Journal of Mathematical Modelling and Algorithms in Operations Research*, 13 (4), 523–536.

Fishburn, P. C. (1974). Lexicographic orders, utilities and decision rules: a survey. *Management Science*, 20 (11), 1442–1471.

Fréchet, M. (1928). "Sur la loi de probabilité de l'écart maximum". sn. Annales De Lla societe Polonaise De Mathematique.

Gao, S. and Wang, J. (2021). Reliability and availability analysis of a retrial system with mixed standbys and an unreliable repair facility. *Reliability Engineering and System Safety*, 205.

Garg, H., Rani, M., Sharma, S. P., and Vishwakarma, Y. (2014). Intuitionistic fuzzy optimization technique for solving multi-objective reliability optimization problems in interval environment. *Expert Systems with Applications*, 41 (7), 3157–3167.

Ghnimi, S., Gasmi, S., and Nasr, A. (2017). Reliability parameters estimation for parallel systems under imperfect repair. *Metrika*, 80, 273–288.

Ghufran, S., Khowaja, S., and Ahsan, M. J. (2012). Optimum multivariate stratified sampling designs with travel cost: a multiobjective integer nonlinear programming approach. *Communications in Statistics - Simulation and Computation*, 41 (5), 598–610.

Goodarzian, F. and Hosseini-Nasab, H. (2021). Applying a fuzzy multi-objective model for a production distribution network design problem by using a novel self-adoptive evolutionary algorithm. *International Journal of Systems Science: Operations & Logistics*, 8 (1), 1–22.

Gruber, A., Yanovski, S., and Ben-Gal, I. (2013). Condition based maintenance via simulation and a targeted bayesian network metamodel. *Quality Engineering*, 25 (4), 370–384.

Gunning, P. and Horgan, J. M. (2004). A new algorithm for the construction of stratum boundaries in skewed populations. *Survey Methodology*, 30 (2), 159–166.

Gupta, N., Ali, I., and Bari, A. (2013). Fuzzy goal programming approach in selective maintenance reliability model. *Pakistan Journal of Statistics and Operation Research*, 9 (3), 321–331.

Gupta, N., Haseen, S., and Bari, A. (2016). Reliability optimization problems with multiple constraints under fuzziness. *Journal of Industrial Engineering International*, 12 (4), 459–467.

Gupta, S. and Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 37 (1), 57–67.

Hallinan, A. J. (1993). A review of the Weibull distribution. *Journal of Quality Technology*, 25 (2), 85–93.

Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41 (236), 517–529.

Hansen, M. H., Hurwitz, W. H., and Madow, W. G. (1953). *Sampling survey methods and theory.* John Wiley, New York.

Haq, A., Ali, I., and Varshney, R. (2020). Compromise allocation problem in multivariate stratified sampling with flexible fuzzy goals. *Journal of Statistical Computation and Simulation*, 90 (9), 1557–1569.

Hartley, H. O. and Hocking, R. R. (1963). Convexing programming by tangential approximation. *Management Science*, 9 (4), 600–612.

Haseen, S., Iftekhar, S., Ahsan, M. J., and Bari, A. (2012). A fuzzy approach for solving double sampling design in presence of nonresponse. *International Journal of Environmental Science and Technology*, 4 (6), 2542–2551.

Haseen, S., Shafiullah., Ali, I., and Bari, A. (2015). Three stage stochastic multivariate stratified sample survey. *Pakistan Journal of Statistics and Operation Research*, 11 (3), 347–359.

Haseen, S., Ali, I., and Bari, A. (2016). Multiobjective stochastic multivariate stratified sampling in presence of nonresponse. *Communications in Statistics - Simulation and Computation*, 45 (8), 2810–2826.

Hedlin, D. (2000). A procedure for stratification by an extended Ekman rule. *Journal of Official Statistics*, 16 (1), 15–29.

Hidiroglou, M. A. and Kozak, M. (2018). Stratification of skewed populations: A comparison of optimisation-based versus approximate methods. *International Statistical Review*, 86 (1), 87–105.

Hidiroglou, M. A. and Srinath, K. P. (1993). Problems associated with designing subannual business surveys. *Journal of Business & Economic Statistics*, 11 (4), 397–405.

Hogg, R. V. and Craig, A. T. (1978). *Introduction to mathematical statistics*. 4th edition, New York: Macmillan Publishing Co., Inc.

Holland, J. H. (1975). *Adaptation in natural and artifical systems*. University of Michigan Press, Ann Arbor.

Horgan, J. M. (2006). Stratification of skewed populations: a review. *International Statistical Review*, 74 (1), 67–76.

Horst, R. (2008). *The Weibull distribution: a handbook*. CRC Press.

Hu, L., Yue, D., and Zhao, D. (2012). Availability equivalence analysis of a repairable series-parallel system. *Mathematical Problems in Engineering*. DOI: `https://doi.org/10.1155/2012/957537`.

Hu, L., Yue, D., and Ma, Z. (2020). Availability analysis of a repairable series-parallel system with redundant dependency. *Journal of Systems Science and Complexity*, 33, 446–460.

Huang, Z. H. (1997). Fuzzy multiobjective optimization decision making of reliability of series system. *Microelectronics Reliability*, 37 (3), 447–449.

Huddleston, H. F., Claypool, P. L., and Hocking, R. R. (1970). Optimum sample allocation to strata using convex programming. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 19 (3), 273–278.

Ignizio, J. P. (1982). On the (re)discovery of fuzzy goal programming. *Decision Sciences*, 13 (2), 331–336.

Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, 78 (381), 117–123.

Ivezic, D., Tanasijevic, M., Jovancic, P., and Djuric, R. (2019). A fuzzy expert model for availability evaluation. *20th International Carpathian Control Conference*, 1–6.

Jahan, N., Khan, M. G. M., and Ahsan, M. J. (2001). Optimum compromise allocation using dynamic programming. *Dhaka University Journal of Science*, 49 (2), 197–201.

Kac, M. (1975). Some reflections of a mathemmatician on the nature and the role of statistics. *Advances in Applied Probability*, 7, 5–11.

Kadilar, C. and Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal*, 45 (2), 218–225.

Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151 (3), 893–902.

Kadilar, C. and Cingi, H. (2006a). Improvement in variance estimation using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 35 (1), 111–115.

Kadilar, C. and Cingi, H. (2006b). Ratio estimators for population variance in simple and stratified random sampling. *Applied Mathematics and Computation*, 173 (2), 1047–1059.

Kakkar, M. K., Bhatti, J., Malhotra, R., Kaur, M., and Goyal, D. (2019). Availability analysis of an industrial system under the provision of replacement of a unit using genetic algorithm. *International Journal of Innovative Technology and Exploring Engineering*, 9 (1), 1236–1241.

Kaur, N. and Singh, S. (2016). A budget-constrained time and reliability optimization BAT algorithm for scheduling workflow applications in clouds. *Procedia Computer Science*, 98, 199–204.

Kennedy, J. and Eberhart, R. C. (1995). Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks*, 4, 1942–1948.

Khalili-Damghani, K., Abtahi, A. R., and Tavana, M. (2014). A decision support system for solving multiobjective redundancy allocation problems. *Quality and Reliability Engineering International*, 30 (8), 1249–1262.

Khan, E. A., Khan, M. G. M., and Ahsan, M. J. (2002). Optimum stratification: a mathematical programming approach. *Calcutta Statistical Association Bulletin*, 52 (1–4), 323–333.

Khan, M. F., Ali, I., and Ahmad, Q. (2011). Chebyshev approximate solution to allocation problem in multiple objective surveys with random costs. *American Journal of Computational Mathematics*, 1 (4), 247–251.

Khan, M. G. M., Ahsan, M. J., and Jahan, N. (1997). Compromise allocation in multivariate stratified sampling: an integer solution. *Naval Research Logistics*, 44 (1), 69–79.

Khan, M. G. M., Najmussehar., and Ahsan, M. J. (2005). Optimum stratification for a exponential study variable under Neyman allocation. *Journal of the Indian Society of Agricultural Statistics*, 59 (2), 146–150.

Khan, M. G. M., Khan, E. A., and Ahsan, M. J. (2008). Optimum allocation in multivariate stratified sampling in presence of nonresponse. *Journal of the Indian Society of Agricultural Statistics*, 62 (1), 42–48.

Khan, M. G. M., Ahmad, S. N., and Khan, S. (2009). Determining the optimum stratum boundaries using mathematical programming. *Journal of Mathematical Modelling and Algorithms*, 8 (4), 409–423.

Khan, M. G. M., Reddy, K. G., and Rao, D. K. (2015). Designing stratified sampling in economic and business surveys. *Journal of Applied Statistics*, 42 (10), 2080–2099.

Khan, S. M. and Ali, I. (2014). Fuzzy geometric programming in multivariate stratified sample surveys in presence of non-response with quadratic cost function. *American Journal of Operations Research*, 4, 173–188.

Khowaja, S., Ghufran, S., and Ahsan, M. J (2013). On the problem of compromise allocation in multiresponse stratified sample surveys. *Communications in Statistics- Simulation and Computation*, 42 (4), 790–799.

Kiessler, P., Klutke, G., and Yang, Y. (2002). Availability of periodically inspected systems subject to Markovian degradation. *Journal of Applied Probability*, 39 (4), 700–711.

Kim, H. and Kim, P. (2017). Reliability redundancy allocation problem considering optimal redundancy strategy using parallel genetic algorithm. *Reliability Engineering and System Safety*, 159 (C), 153–160.

Klutke, G., Wortman, M., and Ayhan, H. (1996). The availability of inspected systems subject to random deterioration. *Probability in the Engineering and Informational Sciences*, 10 (1), 109–118.

Kokan, A. R. (1963). Optimum allocation in multivariate surveys. *Journal of the Royal Statistical Society, Series A*, 126 (4), 557–595.

Kokan, A. R. and Khan, S. U. (1967). Optimum allocation in multivariate surveys: an analytical solution. *Journal of the Royal Statistical Society B*, 29 (1), 115–125.

Kozak, M. (2004). Optimal stratification using random search method in agricultural surveys. *Statistics in Transition*, 6 (5), 797–806.

Kozak, M. (2006). On sample allocation in multivariate surveys. *Communications in Statistics - Simulation and Computation*, 35 (4), 901–910.

Kozak, M. and Wang, H. Y. (2010). On stochastic optimization in sample allocation among strata. *Metron - International Journal of Statistics*, LXVIII (1), 95–103.

Kozak, M., Verma, M. R., and Zielinski, A. (2007). Modern approach to optimum stratification: Review and perspectives. *Statistics in Transition*, 8 (2), 223–250.

Krysan, M., Schuman, H., Scott, L. J., and Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *The Public Opinion Quarterly*, 58 (3), 381–399.

Kumar, A. and Garg, R. K. (2016). Decision support system for maximum availability of series parallel system using particle swarm optimisation. *International Journal of Intelligent Enterprise*, 3 (2), 148–169.

Kumar, G., Jain, V., and Gandhi, O. P. (2013). Availability analysis of repairable mechanical systems using analytical semi-markov approach. *Quality Engineering*, 25 (2), 97–107.

Kuo, W., Prasad, V. R., Tillman, F., and Hwang, C. (2001). *Optimal reliability design: fundamentals and applications*. Cambridge University Press, Cambridge, UK.

Lavallée, P. (1988). Two-way optimal stratification using dynamic programming. *Proceedings of the Section on Survey Research Methods, American Statistical Association, Virginia*, 646–651.

Lavallée, P. and Hidiroglou, M. A. (1988). On the stratification of skewed populations. *Survey Methodology*, 14 (1), 33–43.

Lednick, I. B. and Wieczorkowski, R. (2003). Optimal stratification and sample allocation between subpopulations and strata. *Statistics in Transition*, 6 (2), 287–305.

Li, J., Chen, Y., and Zhang, Y. (2020). System availability modelling and optimization considering multigeneral quality characteristics. *Mathematical Problems in Engineering*, 12, 1–9.

Lientz, B. P. (1974). A stochastic method of allocation of components to maximize reliability. *IEEE Transactions on Reliability*, 23 (2).

Loganathan, M. K. and Gandh, O. P. (2016). Maintenance cost minimization of manufacturing systems using PSO under reliability constraint. *International Journal of Systems Assurance Engineering and Management*, 7, 47–61.

Lu, Y., Lang, M., Sun, Y., and Li, S. (2020). A fuzzy intercontinental road-rail multimodal routing model with time and train capacity uncertainty and fuzzy programming approaches. *IEEE Access*, 8, 27532–27548.

Mahalanobis, P. C. (1952). Some aspects of the design of sample surveys. *Sankhyā: The Indian Journal of Statistics*, 12, 1–17.

Mahapatra, G. S. and Roy, T. K. (2006). Fuzzy multi-objective mathematical programming on reliability optimization model. *Applied Mathematics and Computation*, 174 (1), 643–659.

Mahapatra, G. S. and Roy, T. K. (2009). Reliability evaluation using triangular intuitionistic fuzzy numbers arithmetic operations. *World Academy of Science, Engineering and Technology*, 50, 574–581.

Mahapatra, G. S. and Roy, T. K. (2011). Optimal redundancy allocation in series-parallel system using generalized fuzzy number. *Tamsui Oxford Journal of Information and Mathematical Sciences*, 27 (1), 1–20.

McLeavey, D. W. and McLeavey, J. A. (1976). Optimization of system reliability by branch-and-bound. *IEEE Transactions on Reliability*, 25 (5), 327–329.

Melaku, A. (1986). Asymptotic normality of the optimal allocation in multivariate stratified random sampling. *The Indian Journal of Statistics*, 48 (Series B), 224–232.

Mellal, M. A., Zio, E., and Williams, E. J. (2020). Cost minimization of repairable systems subject to availability constraints using efficient cuckoo optimization algorithm. *Quality and Reliability Engineering International*, 36 (3), 1–13.

Millr, T. J., Skalski, J. R., and Ianelli, J. N (2007). Optimizing a stratified sampling design when faced with multiple objectives. *ICES Journal of Marine Science*, 64 (1), 97–109.

Misra, K. (1972). Reliability optimization of a series-parallel system. Part I: Langrange multiplier approach; Part II: maximum principle approach. *IEEE Transaction on Reliability,* 21 (21), 230–238.

Misra, K. B. (1986). On optimal reliability design: a review. *Journal of System Science*, 12 (4), 5–30.

Moghaddam, R. T., Safari, J., and Sassani, F. (2008). Reliability optimization of series-parallel systems with a choice of redundancy strategies using a genetic algorithm. *Reliability Engineering & System Safety*, 93 (4), 550–556.

Mousavi, S. M., Alikar, N., Avana, M., and Caprio, D. D. (2019). An improved particle swarm optimization model for solving homogeneous discounted series parallel redundancy allocation problems. *Journal of Intelligent Manufacturing*, 30 (3), 1175–1194.

Muhammad, Y. S. and Husain, I. (2017). Trade off between cost and variance for a multiobjective compromise allocation in stratified random sampling. *Communications in Statistics - Theory and Methods*, 46 (6), 2655–2666.

Muhammad, Y. S., Shabir, J., Husain, I., and Abd-el.Moemen, M. (2015). Multiobjective compromise allocation in multivariate stratified sampling using extended lexicographic goal programming with Gamma cost function. *Journal of Mathematical Modelling and Algorithms in Operations Research*, 14 (3), 255–265.

Murthy, M. N. (1964). Product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series A*, 26 (1), 69–74.

Nakagawa, T. (2005). *Maintenance theory of reliability*. Springer Series in Reliability Engineering.

Narasimhan, R. (1980). Goal programming in a fuzzy environment. *Decision Sciences*, 11 (2), 325–336.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7 (4), 308–313.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97 (4), 558–625.

Niemiro, W. (1999). Optimal construction of strata using random search method. *Wiadomosci Statystyczne*, 10, 1–9.

Nourelfath, M., Eric, C., and Nahas, N. (2012). Joint redundancy and imperfect preventive maintenance optimization for series-parallel multi-state degraded systems. *Reliability Engineering and System Safety*, 103, 51–60.

Olufadi, Y. and Kadilar, C. (2014). A study on the chain ratio-type estimator of finite population variance. *Journal of Probability and Statistics*, 26, 69–74.

Osaki, S. (2002). *Stochastic models in reliability and maintainance*. Berlin: Springer Verlag.

Painton, L. and Campbell, J. (1995). Genetic algorithms in optimization of system reliability. *IEEE Transactions on Reliability*, 44 (2), 172–178.

Prasad, B. and Singh, H. P. (1990). Some improved ratio-type estimators of finite population variance in sample surveys. *Communications in Statistics - Theory and Methods*, 19 (3), 1127–1139.

Prasad, V. and Kuo, W. (2000). Reliability optimization of coherent systems. *IEEE Transactions on Reliability*, 49 (3), 323 –330.

Prékopa, A. (1978). *The use of stochastic programming for the solution of the some problems in statistics and probability*. Tech. rep. Mathematical Research Center, Madison.

Raghav, Y. S., Ali, I., and Bari, A. (2014). Multiobjective nonlinear programming problem approach in multivariate stratified sample surveys in case of non-response. *Journal of Statistical Computation and Simulation*, 84 (1), 22–36.

Raghav, Y. S., Khan, M. F., and Khalil, S. (2017). Sample allocation problem in multiobjective multivariate stratified sample survey under two stage randomized response model. *Journal of Mathematical and Computational Science*, 7 (6), 1074–1089.

Rao, M. S. and Naikan, V. N. A. (2014). Reliability analysis of repairable systems using system dynamics modeling and simulation. *Journal of Industrial Engineering International*, 10 (3), 1–10.

Reddy, K. G. and Khan, M. G. M. (2019). Optimal stratification in stratified designs using weibull-distributed auxiliary information. *Communications in Statistics - Theory and Methods*, 48 (12), 3136–3152.

Reddy, K. G., Khan, M. G. M., and Rao, K. D. (2016). A Procedure for computing optimal stratum boundaries and sample sizes for multivariate surveys. *Journal of Software*, 11 (8), 816–832.

Rivest, L. P. (2002). A generalization of the Lavallêe and Hidiroglou algorithm for stratification in business surveys. *Survey Methodology*, 28 (2), 191–198.

Robson, D. S. (1957). Applications of multivariate polykays to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, 52 (280), 511–522.

Rosin, P. and Rammler, E. (1933). Regularities in the distribution of cement particles. *Journal of the Institute of Fuel*, 7, 29–33.

Ross, S. M. (2009). *Introduction to probability and statistics for engineers and scientists*. Associated Press 4th edition, ISBN:978-0-12-370483-2.

Safaei, F., Ahmadi, J., and Gildeh, B. S. (2018). An optimal planned replacement time based on availability and cost functions for a system subject to three types of failures. *Computers & Industrial Engineering*, 124, 77–87.

Salah, N. and Temraz, Y. (2019). Availability and reliability of a parallel system under imperfect repair and replacement: analysis and cost optimization. *International Journal of System Assurance Engineering and Management*, 10 (5), 1002–1009.

Sarkar, J. and Sarkar, S. (2000). Availability of a periodically inspected system under perfect repair. *Journal of Statistical Planning and Inference*, 91 (1), 77–90.

Sasaki, R. O. T. and Shingai, S. (1983). A new technique to optimize system reliability. *IEEE Transactions on Reliability*, 32, 175–182.

Semiz, M. (2004). Determination of compromise integer strata sample sizes using goal programming. *Hacettepe Journal of Mathematics and Statistics*, 33, 91–96.

Sethi, V. K. (1963). A note on optimum stratification of population for estimating the population means. *The Australian Journal of Statistics*, 5 (1), 20–33.

Shafiullah., Khan, M. F., and Ali, I. (2014). Fuzzy geometric programming in multivariate stratified sample surveys in presence of non response with quadratic cost function. *American Journal of Operations Research*, 4 (3), 173–188.

Shafiullah., Haseen, S., and Ali, I. (2015). Geometric programming approach in multivariate stratified sample surveys in case of non-response. *Electronic Journal of Applied Statistical Analysis*, 8 (1), 28–43.

Singh, S. and Kataria, P. (1990). An estimator of finite population variance. *Journal of The Indian Society of Agricultural Statistics*, 42 (2), 186–188.

Singh, S. and Singh, R. (1979). On random non-response in unequal probability sampling. *Sankhya*, 41, 127–137.

Singh, S. and Singh, R. (1985). On random non-response in double sampling with difference estimator. *Communications in Statistics - Simulation and Computation*, 14 (3), 747–757.

Soland, R. M. (1967). Optimum multivariate stratified sampling with prior information. *Scandinavian Actuarial Journal*, 1 (2), 31–39.

Srinath, K. P. (1971). Multiphase sampling in nonresponse problems. *Journal of the American Statistical Association*, 66 (335), 583–586.

Srivastava, S. K. (1967). An estimator using auxiliary information in sample surveys. *Calcutta Statistical Association Bulletin*, 16 (2), 121–132.

Srivastava, S. K. and Jhajj, H. S. (1980). A class of estimators using auxiliary information for estimating finite population variance. *Sankhya C*, 42, 87–96.

Stuart, A. (1954). A simple presentation of optimum sampling results. *Journal of the Royal Statistical Society: Series B*, 16 (2), 101–116.

Subramani, J. (2013). A new modified ratio estimator of population mean when median of the auxiliary variable is known. *Pakistan Journal of Statistics and Operation Research*, 9 (2), 137–145.

Subramani, J. (2016). A new median based ratio estimator for estimation of the finite population mean. *Statistics in Transition, New Series*, 17 (4), 591–604.

Subramani, J. and Kumarapandiyan, G. (2012). Estimation of population mean using coefficient of variation and median of an auxiliary variable. *International Journal of Probability and Statistics*, 1 (4), 111–118.

Subramani, J. and Kumarapandiyan, G. (2013). A new modified ratio estimator of population mean when median of the auxiliary variable is known. *Pakistan Journal of Statistics and Operation Research*, 9 (2), 137–145.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Asok, C. (1984). *Sampling theory of surveys with applications*. Iowa State University Press, Iowa, USA and Indian Society of Agricultural Statistics, New Delhi, India.

Sung, C. S. and Cho, Y. K. (2000). Reliability optimization of a series system with multiple choice and budget constraints. *European Journal of Operational Research*, 127 (1), 159–171.

Sweet, E. M. and Sigman, R. S. (1998). Evaluation of model-assisted procedures for stratifying skewed populations using auxiliary data. *Proceedings of the Section on Survey Research Methods*, 1, 491–496.

Taghizadeh, H. and Hafezi, E. (2012). The investigation of supply chain's reliability measure: a case study. *Journal of Industrial Engineering International*, 8 (22), 1–10.

Tailor, R. and Sharma, B. (2009). A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis. *Statistics in Transition, New Series*, 10 (1), 15–24.

Tillman, I., Hwang, C., and Kuo, W. (1980). *Optimization of system reliability*. Marcel Dekker, New York.

Tschuprow, A. A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 2 (1), 646–683.

Unnithan, V. K. G. (1978). The minimum variance boundary points of stratification. *Sankhya 40C*, 10, 60–72.

Upadhyaya, L. N., Singh, H. P., Chatterjee, S., and Yadav, R. (2011). Improved ratio and product exponential type estimators. *Journal of Statistical Theory and Practice*, 5 (2), 285–302.

Valliant, R., Dorfman, A. H., and Royall, R. M. (2000). *Finite population sampling and inference: A prediction approach*. John Wiley, New York.

Varshney, R., Ahsan, M. J., and Khan, M. G. M. (2011). An optimum multivariate stratified sampling design with nonresponse: a lexicographic goal programming approach. *Journal of Mathematical Modelling and Algorithms*, 10 (4), 393–405.

Varshney, R., Najmussehar., and Ahsan, M. J. (2012). Estimation of more than one parameters in stratified sampling with fixed budget. *Mathematical Methods of Operations Research*, 75 (2), 185–197.

Varshney, R., Khan, M. G. M., Fatima, U., and Ahsan, M. J. (2015). Integer compromise allocation in multivariate stratified surveys. *Annals of Operations Research*, 226 (1), 659–668.

Varshney, R., Gupta, S., and Ali, I. (2017). An optimum multivariate multiobjective stratified sampling design: fuzzy programming approach. *Pakistan Journal of Statistics and Operation Research*, 13 (4), 829–855.

Wagner, H. M. (1959). Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, 54 (285), 206–212.

Wang, H., Wang, W., Zhou, X., Zhao, J., Wang, Y., Xiao, S., and M., Xu. (2020). Artificial bee colony algorithm based on knowledge fusion. *Complex & Intelligent Systems*. DOI: `https://doi.org/10.1007/s40747-020-00171-2`.

Wang, Z. H. (1992). *Reliability engineering theory and practice*. Taipei: Quality Control Society of Republic of China.

Yadav, S. K., Mishra, S. S., and Shukla, A. K. (2014). Improved ratio estimators for population mean based on median using linear combination of population mean and median of an auxiliary variable. *American Journal of Operational Research*, 4 (2), 21–27.

Yadav, S. K., Gupta, S., Misra, S. S., and Shukla, A. K. (2016). Modified ratio and product estimators for estimating population mean in two-phase sampling. *American Journal of Operational Research*, 6 (3), 61–68.

Yan, Z. and Tian, B. (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. *Information Computing and Applications - International Conference*, 106, 103–110.

Yates, F. (1960). *Sampling methods for censuses and surveys*. Charles Griffin and Co. Ltd., London.

Yusuf, I. (2014). Comparative analysis of profit between three dissimilar repairable redundant systems using supporting external device for operation. *Journal of Industrial Engineering International*, 10 (4), 199–207.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338–390.

Zhanga, G., Wang, J., Xue, Y., Zhang, C., Xu, B., Cheng, Y., and Ren, Y. (2020). Comparison of sampling effort allocation strategies in a stratified random survey with multiple objectives. *Aquaculture and Fisheries*, 5 (3), 113–121.

Zheng, J., Okamura, H., Pang, T., and Dohi, T. (2021). Availability importance measures of components in smart electric power grid systems. *Reliability Engineering and System Safety*, 205 (C). DOI: `10.1016/j.ress.2020.107164`.

Zhiyuan, O., Yu, L., Sheng-Jia, R., and Tao, J. (2019). An improved particle swarm optimization algorithm for reliability redundancy allocation problem with mixed redundancy strategy and heterogeneous components. *Reliability Engineering and System Safety*, 181 (C), 62–74.

Zhou, Y., Lin, T. R., Bian, Y., and Ma, L. (2015). An effective approach to reducing strategy space for maintenance optimisation of multi state series parallel systems. *Reliability Engineering & System Safety*, 138, 40–53.

Zimmermann, H. J. (1978). Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems*, 1 (1), 45–55.

Zukhovitsky, S. I. and Avdeyeva, L. I. (1966). *Linear and convex programming*. William H Burt Saunders, Philadelphia.