OPTIMUM STRATIFICATION WITH AUXILIARY INFORMATION USING MATHEMATICAL PROGRAMMING

By

FAIZAN DANISH
(J-14-D-04-BS)

Thesis submitted to Faculty of Basic Sciences
in partial fulfillment of requirements
for the degree of

DOCTOR OF PHILOSOPHY

IN

STATISTICS

Division of Statistics and Computer Science
Faculty of Basic Sciences
Sher-e-Kashmir University of Agricultural Sciences & Technology of Jammu
Main Campus, Chatha, Jammu-180009

2018

CERTIFICATE-I

This is to certify that the thesis entitled "Optimum Stratification with Auxiliary Information using Mathematical Programming" submitted in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Statistics** to the Faculty of Basic Sciences, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu is a record of bonafide research carried out by **Mr. Faizan Danish**, Registration No. **J-14-D-04-BS** under my supervision and guidance. No part of the thesis has been submitted for any other degree or diploma. It is further certified that the help and assistance received during the course of investigation have been duly acknowledged.

18/01/18
(S.E.H. Rizvi)
Major Advisor

Place: Jammu
Date : 18/01/18

Endorsed

Head 18/01/18
Division of Statistics and Computer Science, FBSc
SKUAST-J, Chatha

Date: 18/01/18

CERTIFICATE-II

We, the members of the Advisory Committee of **Mr. Faizan Danish**, Registration No: **J-14-D-04-BS,** a candidate for the degree of **Doctor of Philosophy in Statistics**, have gone through the manuscript of the thesis entitled **"Optimum Stratification with Auxiliary Information using Mathematical Programming"** and recommend that it may be submitted by the student in partial fulfillment of the requirements for the degree.

18/01/18
Major Advisor & Chairman

Place : Jammu
Date : 18-01-18

Advisory Committee Members

Dr. Manish Kumar Sharma
Assoc. Professor
Division of Statistics and Computer Science

Dr. Sudhakar Dwivedi
Assoc. Professor
Division of Agricultural Economics and ABM

On leave

Dr. Bupesh Kumar
Assistant Professor
Division of Plant Breeding and Genetics

Dr. Sanjeev Kumar
Sr. Scientist
Division of Plant Breeding and Genetics
ACRA, Dhiansar , SKUAST-Jammu

CERTIFICATE-III

This is to certify that the thesis entitled **"Optimum Stratification with Auxiliary Information using Mathematical Programming"** submitted by **Mr. Faizan Danish,** Registration No. **J-14-D-04-BS** to the Faculty of Basic Sciences, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, in partial fulfillment of the requirement for the degree of **Doctor of Philosophy,** was examined and approved by the Advisory Committee and External Examiner(s) on **23.02.2018.**

27-2-18

External Examiner
(Dr. Athar Ali Khan)
Professor (Statistics)
Department of Stat. & Operations Research
Aligarh Muslim University, Aligarh (UP)

03/03/18

Dr. S.E.H.Rizvi
Professor (Statistics) and Head
Statistics and Computer Science
Major Advisor

05/03/18

Head
Division of Statistics and Computer Science

05/03/18

Dean, FBSc

ACKNOWLEDGEMENT

In the name of Allah The Most Gracious and The Most Merciful. All bounties Are in the hand of Allah: He grants them To whom He pleases And Allah cares for all, And He knows all things. For His Mercy He specially chooses Whom he pleases: For Allah is the Lord of bounties unbounded.

No one can ever make any achievement all of his alone. It is always difficult to acknowledge so precious a debt as that of learning. It is the debt that could only be paid through gratitude. I bow with gratitude to all the teachers who guided me. Completion of this doctoral dissertation was possible with the support of several people. I would like to express my sincere gratitude to all of them. First of all, I am extremely grateful to my research guide, Dr. S. E. H. Rizvi, Professor and Head, Division of Statistics and Computer Science and Dean, Faculty of Basic Sciences, for his valuable guidance, scholarly inputs and consistent encouragement I received throughout the research work. This feat was possible only because of the unconditional help and support provided by Sir. A person with an amicable and positive disposition, Sir has always made himself available to clarify my doubts despite his busy schedules and I consider it as a great opportunity to do my doctoral programme under his guidance and to learn from his research expertise. I am highly thankful to Dr. Manish Kumar Sharma, Associate Professor, Division of Statistics and Computer Science, Faculty of Basic Sciences for his kind guidance, support and encouragement from time to time with the help of which I could submit the thesis within the stipulated time. I am extremely thankful to my other Advisory committee members who have been very kind enough to extend their help at various phases of this research, whenever I approached them, and I do hereby acknowledge Dr. Sudhakar Dwivedi Associate professor, Division of Agricultural Economics and ABM, Dr. Sanjeev Kumar, Sr. Scientist Division of PBG (Dean's Nominee) and Dr. Bupesh Kumar, Assistant Professor Division of PBG.I whole heartedly thank Dr. M. Iqbal Jeelani Bhat, Assistant Professor, Division of Statistics and Computer Science, Faculty of Basic Sciences for his valuable suggestions and support in statistical data analysis and I thank him for his contributions.

I am extremely thankful to Dr. M.G. M. Khan, Professor, School of Computing, Information and Mathematical Sciences, University of South Pacific, Suva, Fiji for his kind support especially in simulation/numerical illustration part of the thesis and also to

Professor Aquil Ahmad, Department of Statistics and O.R, AMU Aligarh for their help and encouragement when approached.

I express my utmost regards to Hon'ble Vice Chancellor and Director Education of SKUAST-JAMMU, who have been kind and supportive.

I thank all non teaching staff which includes like Sajad Ahmad Bhat, Bandhana, Joginder Lal, Sanjay Sharma, Teke Rajin from division of Statistics and Computer Science and Dean, FBSc office for their kind support and cooperation.

Role of library was inevitable during my research programme and I hereby take privilege to express my gratitude to all the library staff.

I am obliged towards Department of Statistics and Operations research, Aligarh Muslim University, Aligarh (U.P.) for giving me the permission for literature collection.

I am extremely thankful to my father Gull Mohd Ganie and my Mother Gulshana who encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true. I am deeply thankful to my dear elder brother Javaid Hussain Ganaie who always sacrificed his happiness to make my dream while sacrifices his own dreams.I extend my gratitude to my sisters Rukiya Akhter and Rafiyaia Aara who always anxiously waited for my achievements. Once again I thank to my family who always stood like a lighthouse in laminating the pathway of my success and provided me valuable opportunities in building of my career.

I fell short of words expressing my gratitude to all my friends Dr. Bunti Kumar, Archana, Sunali Mahajan, Saquib Khan, Chetna Sakolia, Abdul Majeed, Sajad Ahmad Sofi, Owais Ali Wani, Dr. Meenakshi Raina, Dr. Jebi Sudan, Jahangir Ali, Tajamul Islam, Mansoor Shafi, Anjum Ayoub, Ashwani Kumar, Ankit Dayal. I once again owe it all to Almighty Allah for granting me the wisdom, health and strength to undertake this research task and keeping me steadfast till its completion.

Place: Jammu

Date: March, 2018.

Faizan -Danish.

Faizan Danish

ABSTRACT

Title of the Thesis	:	Optimum Stratification with Auxiliary Information using Mathematical Programming
Name of the student	:	Faizan Danish
Major Subject	:	Statistics
Name and Designation of the Major Advisor	:	Dr. S.E.H.Rizvi Professor (Statistics) & Head Division of Statistics and Computer Science
Degree to be awarded	:	Ph.D. (Statistics)
Name of the University	:	Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu

The method of choosing the best boundaries that makes strata internally homogeneous as far as possible is known as optimum stratification. To achieve this, the strata should be constructed in such a way that the strata variances for the characteristic under study be as small as possible. If the frequency distribution of the study variable is known, the stratification points could be obtained by cutting the range of the distribution at suitable points. However, if the frequency distribution of the study variable is unknown, it may be approximated from the past experience or some prior knowledge obtained at a recent study. In the present investigation entitled "optimum stratification with auxiliary information using mathematical programming", theories have been developed for optimum stratification, when one character of the population is under study, using two auxiliary variables as stratification variables. Minimal equations giving optimum strata boundaries (OSB) have been obtained for stratified random sampling under different methods of allocations, by minimizing the variance of the sample estimates. For all these cases, Cum $\sqrt[3]{D_i\left(x,z\right)}$ rules (i=1,2,3,4) ,where the function $D_i\left(x,z\right)$ takes different forms for different allocations, have been developed. Under the classical optimization technique the empirical studies on uniform, right triangular and exponential distribution have been made for the gain in efficiency that could be achieved through the proposed rules. It has been observed that proposed technique under classical optimization approach performed better than the techniques using single auxiliary variable proposed by Ekman (1959a), Singh (1971) and other workers.

Methods have also been developed for obtaining OSB using mathematical programming technique for different allocations. The problem of determining OSB are formulated as nonlinear programming problem (NLPP),which turn out to be multistage decision problem and are solved using dynamic programming approach. The basic advantage of mathematical programming over the classical optimization is that it can determine OSB efficiently, when the density function of the population is approximately known from the previous studies. Iterative methods required initial solutions and there is no guarantee that they will converge and give the global minimum variance in the absence of a suitably chosen initial solution while as the proposed methods do not require any initial approximatio solution. More importantly, the proposed techniques have a wide scope of application a compared to other methods. In practice, the complete data set of the study variable i unknown, which diminishes the uses of many stratification techniques. In such a situatior the proposed techniques can be used as it requires only the values of parameters of th population, which can easily be available from the past studies. The proposed methods usin the simulated data are presented to illustrate the applications and the computational detail using R and LINGO softwares. In case of general variance, the results are presented togethe with the results of Cum $\sqrt{f}$ method of Dalenius and Hodges (1959), the geometric method by Gunning and Horgan (2004) and the generalized method of Lavallee-Hidiroglou (1988), Khan *et al.* (2008) for computational analysis. It has been observed that proposed method leads to substantial gains in the precision of the estimates. Besides, the proposed method under equal allocation also showed more efficient results than the method proposed by Singh (1977).Under proportional allocation, by empirical study it is found that the method proposed by Thomson (1973) and Khan *et al.* (2005) are inferior to the proposed method. The proposed technique, under Neyman allocation, has been found to be more efficient than the methods given by Singh and Sukhatme (1969), Khan *et al.* (2008), Khan *et al.* (2014) and Nand and Khan (2008).Through the empirical investigations it is to be concluded that the method proposed by Fonolahi and Khan (2014) is having greater variance than proposed method. However, it is to be noted that the variance obtained using mathematical programming technique is less than the variance obtained by the classical optimization technique (Cum $\sqrt{D_i(x,z)}$) of the present investigationfor all the allocations. Thus, it is to be concluded that the mathematical programming technique leads to substantial increase in the precision as compared to classical optimization technique.

Keywords: *Optimum Stratification, Optimum Strata Boundaries, Auxiliary Information, Mathematical Programming Technique*

Signature of Major Advisor

Signature of the Student

CONTENTS

Chapter No.	Topic	Page No.
1.	Introduction	1-4
2.	Review of Literature	5-26
3.	Material and Methods	27-32
4.	Classical Optimization Technique	33-94
5.	Mathematical Programming Technique	95-156
6.	Discussion	157-164
7.	Summary and Conclusions	165-168
	References	i-x
	Vita	

LIST OF TABLES

Table No.	Title	Page No.
	Classical Optimization Technique	
4.11.1	OSB when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\alpha$	68
4.11.2	OSB and Variance when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\alpha$	69
4.11.3	OSB when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\lambda xz$	69
4.11.4	OSB and Variance when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\lambda xz$	70
4.11.5	OSB and Variance when the auxiliary variables are dependent and follow uniform and exponential distributions, respectively, having $\eta(x,z)=\alpha$	71
4.11.6	OSB and Variance when the auxiliary variables are dependent and follow uniform and exponential distributions, respectively, having $\eta(x,z)=\lambda xz$	71
4.14.1	OSB when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, and with $\eta(x)=\alpha_1 \quad and \quad \eta(z)=\alpha_2$	74
4.14.2	OSB and Variance when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, and with $\eta(x)=\alpha_1 \quad and \quad \eta(z)=\alpha_2$	75
4.14.3	OSB when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, with $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$	75
4.14.4	OSB and Variance when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, and with $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$	76

4.14.5	OSB and Variance when the auxiliary variables are independent following right triangular and uniform distributions, respectively, with $\eta(x)=\alpha_1 \quad and \quad \eta(z)=\alpha_2$	76
4.14.6	OSB and Variance when the auxiliary variables are independent following right triangular and uniform distributions, respectively, with $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$	77
4.19.1	OSB when the auxiliary variables X and Z have both standard normal distribution for the case of dependent variables under proportional allocation	85
4.19.2	OSB and Variance when the auxiliary variables are both standard normally distributed for dependent variables under proportional allocation	86
4.19.3	OSB and Variance when the auxiliary variables X and Z are independent following right triangular and exponential distribution, respectively	87
4.19.4	OSB and Variance when the auxiliary variables X and Z are independent following right triangular and exponential distribution, respectively.	87
4.19.5	OSB for the case of dependent auxiliary variables X and Z following standard log-normal and uniform distributions	88
4.19.6	OSB and Variance for the case of dependent auxiliary variables X and Z following standard log-normal and uniform distributions	89
4.22.1	OSB when the auxiliary variables X and Z are independent having right triangular and exponential distribution, respectively	92
4.22.2	OSB and Variance when the auxiliary variables X and Z are independent having right triangular and exponential distribution, respectively	92
4.22.3	OSB and Variance when the auxiliary variables X and Z are independent having right triangular and exponential distribution, respectively	93
4.22.4	OSB and Variance when the auxiliary variables X and Z are independent and follow standardized log-normal and uniform distributions, respectively	94

Mathematical programming Technique		
5.3.1	Stratification points when the auxiliary variables X and Z are independent and follow exponential and right-triangular distributions	105
5.3.2	OSB, Stratum Weight and Variance of the exponential and right triangularly distributed independent auxiliary variables	105
5.3.3	The variance of variables and their total variances	107
5.3.4	OSB when the auxiliary variables X and Z are correlated and are exponentially and right- triangularly distributed	108
5.3.5	Displays OSB and Variance for correlated auxiliary variables having exponentially and right- triangularly distribution	109
5.3.6	OSB and Variance when the auxiliary variables are uniformly and exponentially distributed	110
5.3.7	OSB and Variance for exponential and pareto distributed auxiliary variables	114
5.4.1	OSB when the auxiliary variables X and Z are uniformly and right triangularly distributed	117
5.4.2	OSB and Variance for uniform and right triangular distributed auxiliary variables	118
5.4.3	OSB and Variance for uniform and right triangular distributed auxiliary variables	119
5.5.1	OSB when the auxiliary variables X and Z are independent having right triangular and exponential distribution respectively	123
5.5.2	OSB and Variance when the auxiliary variables X and Z are independent having right triangular and exponential distribution respectively	124
5.5.3	OSB and Variance when the auxiliary variables X and Z are dependent having right triangular and exponential distribution respectively	125
5.5.4	OSB when the auxiliary variables X and Z follow log-normal and uniform distribution respectively	128
5.5.5	OSB, and Variance when the auxiliary variables X and Z follow log-normal and uniform distribution respectively	129
5.6.1	OSB when the auxiliary variables X and Z have uniform and exponential distribution respectively	132

5.6.2	OSB and Variance of proposed method and others when the auxiliary variables X and Z have uniform and exponential distribution respectively	133
5.6.3	OSB and Variance of proposed method and others having uniform and exponential distributed auxiliary variables when they are dependent	134
5.6.4	OSB and Variance of proposed method when the auxiliary variables X and Z have right-triangular and gamma distribution respectively	138
5.6.5	OSB and Variance of proposed method and other existing methods when the auxiliary variables X and Z follow uniform and exponential distribution respectively	141
5.6.6	OSB when the auxiliary variables X and Z have uniform and right triangular distribution respectively	144
5.6.7	OSB and Variance of proposed method having distribution of X and Z as uniform and right triangular distributions, respectively	144
5.7.1	OSB when the auxiliary variables X and Z having right-triangular and exponential distribution respectively	149
5.7.2	OSB and Variance of proposed method and others when the auxiliary variables X and Z are having right-triangular and exponential distributions respectively	149
5.7.3	OSB and Variance of proposed method when the auxiliary variables X and Z are having right triangular and exponential distribution respectively	151
5.7.4	OSB when the auxiliary variables X and Z follow uniform and standard normal distributions respectively	153
5.7.5	OSB and Variance of proposed method and others when the auxiliary variables X and Z follow uniform and standard normal distribution respectively	154
5.7.6	OSB and Variance when the auxiliary variables X and Z are having right triangular and exponential distribution respectively	156

LIST OF FIGURE

Figure No.	Title	Page No.
1.	Curve of number of strata and corresponding variance	105

CHAPTER 1

INTRODUCTION

In statistical investigations the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to the group under study, known as population or universe. The population, which is aggregate of objects, may be finite or infinite. It is obvious that for any statistical investigation complete enumeration of the population (known as census) is generally impracticable. In such cases we take help of sampling because of multiplicity of causes like administrative and financial implications, time factor etc.

Sampling is the procedure of selecting smaller portion of the population (known as sample) that is supposed to be truly representing the patterns of the target population at large. In other words, it consists of selecting sample so that one may construct the estimates of the unknown population parameters on the basis of the sample observations. In a national opinion survey, in which only a sample of the people belonging to the population under study are contacted, the opinions of individuals in the sample are used to estimate the proportions with the various opinions for the overall population. For instance to estimate the prevalence of a rare disease, the sample might consist of a number of medical institutions, each of which has records of the patients treated for that particular disease. To estimate the presence of a rare and endangered plant in a crop, the presence of these plants in the population is estimated based on the pattern of detections from a sample of sites in the study region, etc. The population being investigated may be homogenous or heterogeneous one with respect to characteristics under study .In the latter case, stratified sampling is generally used. In this technique the whole population is divided into homogenous sub-populations, known as strata. These strata are non-overlapping, and together they comprise the whole population. After forming the strata, samples of predetermined sizes are drawn independently from each stratum. For the purpose of selection of sampling units from different strata, one may use different sampling designs in different strata. When the sample within each stratum is drawn with simple random sampling method, the technique is called stratified simple random sampling.

There are many reasons for using the technique of stratification. Many a times, the administrative convenience in carrying out the survey also dictates the use of this technique. The principal idea is of course to increase the precision of the estimates of the population parameters under study. However, here we shall restrict our study to the estimation of population mean or total. Thus, while using stratified sampling one has to select an appropriate stratification technique, which corresponds to minimum variance in order to increase the efficiency of the estimators. The stratification technique which results in minimum possible variance is known as optimum stratification.

The factors that influence the reduction of variance include:

a) choice of stratification variable

b) number of strata

c) determination of strata boundaries

d) allocation of samples.

The present investigation is mainly concerned with the determination of strata boundaries. However, for this purpose other factors would also be taken into consideration. Once the total number of strata and the procedure of allocating sample sizes to different strata is decided, the problem of stratification may be considered to consist of determination of strata boundaries, which constitute the construction of strata. These strata should be constructed in such a way that the sampling variance is minimized. A wrong choice of stratification may result in a considerable increase in variance.

The study of the behaviour of variance with the change in the sample sizes brings out the question of sample allocation to different strata. Several methods of allocation units from different strata are available among which the most commonly used are the methods of proportional and optimum allocations. In proportional allocation, the units from each stratum are taken in proportion to its size. In the method of optimum allocation, the selection of units from each stratum is made in proportion to the product of stratum size and its standard deviation, where it is assumed that the cost of observing a unit is the same in all the strata. Another least used method of allocation is known as equal allocation in which equal number of units is taken from each stratum. For a given sample size, the variance of the estimate of population mean is supposed to be minimum when allocation of sample sizes to different strata is done according to Neyman allocation, assuming of course that the other parameters are kept constant. The rate of decrease in variance is considerably more when the number of strata goes on increasing.

However, a stage is soon reached where the increase in number of strata does not significantly decrease the variance.

Strata may be constructed either on the basis of study variable itself or by using some other variable highly correlated with the variable under study. The variable under study is known as estimation variable, and the variable on the basis of which stratification is done is known as stratification variable. Since, in most of the cases the distribution of the estimation variable is not known, it is not possible to stratify on the basis of estimation variable itself. However, one could stratify on the basis of some auxiliary variable (closely related with the estimation variable) whose density is known in advance.

Dalenius (1950) first considered the problem of optimum stratification for one variable using study variable itself as the basis of stratification. Dalenius and Gurney (1951) developed a technique of obtaining optimum strata boundaries (OSB) using an auxiliary variable closely related to study variable. Singh and Sukhatme (1969) considered the problem of finding approximately optimum strata boundaries (AOSB) on an auxiliary variable for one estimation variable case. The problem of optimum stratification, when two or more characters of the population are under study, seems to be relatively of greater practical importance. Ghosh (1963) has given an outline of the approach to be followed for solving the problem of optimum stratification when two characters are under study. Rizvi *et al.* (2000) tackled the problem of obtaining AOSB for two study variables using one auxiliary variable as the basis of stratification. However, very little work has been done in this regard.

Another advanced method used for obtaining OSB is the optimization method. Investigating for and arriving at the best possible decision in any given circumstances is called optimization. The ultimate aim of all such decisions is to maximize the gain or profit or to minimize cost which could be earned in the given circumstances or to minimize the cost or waste or loss incurred in certain processes. Existence of this method can be traced back to the mid of eighteenth century. The work of Newton, Lagrange and Cauchy in solving certain types of optimization problems arising in geometry and physics by using differential calculus methods and calculus of variations are pioneering. These optimization methods, better known as classical optimization methods, have their own illustrations and cannot be applied successfully to every optimization problem. These are mainly of theoretical interest. The first mathematical programming problem (MPP) was perhaps the problem of optimum allocation of limited resources recognised by economists in the early 1930s.In 1947, after World War II, in 1947 the United States Air Force team

SCOOP (Scientific Computation of Optimum Programs) started as intensive research on some optimum resource allocation problem which led to the development of the famous simplex method by George B. Dantzig for solving a linear programming problem. Kuhn and Tucker (1951) developed the necessary conditions to be satisfied by an optimal solution of an MPP. These conditions known as K-T conditions laid the foundation for great deal of later research and development in nonlinear programming techniques.

In most of the investigations, the stratification variables have been taken as identical to the study variables. But, this is unrealistic and impracticable as the probability density functions of the estimation variables are generally not available. Further, as per literature many investigations have been done using single auxiliary variable as the basis of stratification in order to obtain AOSB. But it is seldom that work is reported for the use of more than one auxiliary variable. Presuming the fact that the efficiency of the estimator may be improved by using more auxiliary information, in the present study tackles the problem of optimum stratification for single study variable using two auxiliary variables as the basis of stratification, by using classical as well as programming technique approaches, with the following objectives:

1. To develop the techniques of optimum stratification for single study variable using two auxiliary variables.
2. To develop the technique of optimum stratification for various allocation procedures.
3. To obtain optimum strata boundaries using mathematical programming approach.
4. To obtain the efficiency of developed methods with the existing methods empirically.

In order to achieve the above objectives, it is assumed that the finite population under consideration is a random sample from an infinite super-population with the same characteristics. This assumption enables us to have the joint density function of (Y, X, Z), where X and Z are stratification variables. The marginal density functions of Y, X and Z are assumed to be continuous, so that various methods of calculus can be used. Further, it is assumed that the regression equation of Y on X and Z is given by Y = C (X, Z) + e where C (X, Z) is a real valued functions of X and Z, and 'e' is error term. Furthermore, the forms of conditional variances of Y for given X and Z are also assumed.

CHAPTER 2

REVIEW OF LITERATURE

This chapter has been devoted to present a review of previous studies related with the present investigation. The basic purpose of this chapter is to understand the methodologies adopted and the trend of conclusions derived in earlier related studies so that a suitable methodological framework could be developed for the present study. Keeping in view the objectives of this study, the review of available literatures have been presented under the following sub-headings:

- Classical Optimization Technique
- Mathematical Programming Technique

Classical Optimization Technique

Dalenius (1950) first considered the problem of optimum stratification for the case of stratified random sampling estimate. By minimizing the variance of the estimate with respect to the strata boundaries he was able to obtain equations, the solutions to which gave optimum points of stratification. The equations were obtained for both optimum and proportional allocation methods. Although theoretically, these equations gave the optimum strata boundaries, their solution presented many difficulties. The parameters involved in these equations themselves were functions of points of stratification and, therefore, could not be evaluated unless these points are known. In deriving these equations, Dalenius assumed both stratification and estimation variables to be the same. Therefore, even if the strata boundaries were known, the values of parameters could not be estimated without having the knowledge of the distribution of estimation variable. The need of having some methods for finding at least approximations to the optimum strata boundaries (OSB) was therefore felt.

Dalenius and Gurney (1951) considered the problem of stratifying the population with respect to an auxiliary variable so as to minimize the variance of the stratified random sampling estimate. They assumed the knowledge of frequency function of the auxiliary variable X in an infinite population and found the best boundary points for dividing this frequency function into two strata. The variable Y, whose population mean is to be estimated from the sample, was assumed to be related to X by the equation $Y = \psi(X) + \eta$, where ψ and η are uncorrelated and E(η) = 0.The division points were established for both optimum and proportional allocations with fixed sample size and also for optimum allocation with cost C_i per unit in the i^{th} stratum. For the case of optimum allocation, it was suggested that $W_h\sigma_h$ is constant gave approximately optimum strata boundaries. For fixed sample size, the division points were also given for finite populations, with illustration which suggested that the results for an infinite population will always be nearly adequate in practice.

Mahalanobis (1952) suggested that when the number of strata is predetermined, say L, a practical method of possible stratification is to stratify the whole population into a set of L strata such that the strata sum are equal. The method was not supported by any theoretical justification. In this method also since the strata totals for the estimation variable could not be obtained in advance, one has to use some highly correlated auxiliary variable. These approaches, therefore, only approximately satisfied Mahalanobis' criteria. The method suffers from another drawback that it is not in-variant under the change of origin while the problem of optimum stratification was invariant.

Hansen *et al.* (1953) demonstrated that Mahalanobis rule would lead to efficient stratification if the coefficient of variations were same for all the strata. Aoyama (1954) while considering various problems in stratified sampling suggested that nearly optimum stratification is obtained if the range of the variable is divided into equal intervals, the number of intervals being equal to the number of strata desired. This rule was suggested for both Neyman and proportional allocations. As this method of constructing the strata is universal i.e. same rule is suggested irrespective of the distribution of the variable in the population, much accuracy cannot be expected to be obtained. Another drawback of the rule is that it can only be applied to variables having finite range. The construction of strata according to this rule is of course very easy in practice.

Kitagawa (1956) considered Mahalanobis suggestion and showed that equal strata total method gave optimum stratification only for the populations in which strata co-efficient of the variation were same for all possible stratifications. Assuming that the whole population consists of an aggregate of some elementary clusters, he showed that under certain conditions equal strata coefficients of variation criterion was approximately satisfied if the elementary clusters were set up in such a way that they followed log-normal distribution. Considering a linear cost function he also suggested the principle of equipartitioning of total cost into each stratum, as a method of stratifying the population.

Block (1958) considered the problem of finding optimum points of stratification for optimum allocation method when the joint density function of the estimation variable and the stratification variable could be thought to be bivariate log-normal with parameters (1_1, σ_1 , 1_2 and σ_2), index 1 referring to the former variable. He found that the optimum stratification point's x_i can be calculated relatively easily if the number of strata is less say 2, 3 or 4.It is, however, necessary that the logarithmic correlation coefficient 'r' as well as the standard deviation σ_1 of the estimation variable Y are known at least as estimates from some earlier investigations. The natural logarithm of the approximation to the optimum points of stratification are given by the relation

$$\log x_i = 1_2 + \sigma_2 \left(h_i + r\sigma_1 \right)$$

where $\{h_i\}$ are the optimum points of stratification for the standard normal distribution. The result is valid for reasonably strong correlation and low sampling rates.

An efficient method of constructing strata under Neyman allocation was given by Dalenius and Hodges (1959) .In this case also, it was assumed that both the estimation and the stratification variables were same. If the estimation variable is Y and f (y) its probability density function with (a, b) as the range of the variable, where (b-a) is finite, the nearly optimum points of stratification { y_h^* } are given by

$$G\left(y_h^*\right) = \frac{h.K}{L}, \quad h= 1,2,...,L-1$$

and y_0 = a , y_L = b where L is the total number of strata desired, $G\left(y\right) = \int_a^y \sqrt{f\left(t\right)}dt$ and K = G(b).The results were obtained under the assumption that the number of strata is very large and the density function of the variable Y in each can be approximated by uniform distribution. Therefore, the results are only asymptotically optimum. It was also shown that when $L \to \infty$,the variance obtained by using the suggested strata boundaries tends to minimum variance, and $LS_0 \to \frac{K^2}{\sqrt{12}}$ where $S_0 = \left(\sum W_h \sigma_h\right) \min$ .When stratifying for the variable with infinite range, it was suggested to suitably truncate the distribution and then apply the suggested method to obtain the points of stratification. It was also proved that with L large this procedure was equivalent to keeping $W_h\sigma_h$ = constant as suggested by Dalenius and Gurney.

Another method of finding approximately optimum strata boundaries (AOSB) for Neyman's allocation in simple random sampling estimate was given by Ekman (1959a).Under certain regularity conditions on the density function f(y) and a finite range of the variable Y, it was shown that the points $\{y_h\}$ satisfying the equations

$$\left(y_h - y_{h-1}\right)W_h = C_L \qquad h= 1 , 2, ...,L$$

where C_L is a constant depending on L, approximately satisfy the minimal equations. For the densities over an infinite range, obviously, the equalities suggested were not applicable. Certain modifications to the above equalities were recommended to deal with such cases. The results were derived under the assumption of large number of strata so that higher power of stratum widths could be neglected. It was also proved that as L $\to\propto$, the AOSB obtained from suggested method tend to optimum points $\{y_h\}$ and the variance of the estimate using approximate stratification approaches the minimum variance. Numerical illustrations for L = 2,3,4,5 and for three densities showed that also for small values of L, the approximations were quite satisfactory. A method of iteration to obtain successively better approximations to the optimum $\{y_h\}$ was also suggested, where approximate $\{y_h\}$ obtained from the above equalities could be taken as the starting values.

Ekman (1959b) in another paper obtained the method of finding AOSB for the case of proportional allocation. In this case both estimation and stratification variables are assumed to be the same. By finding the series expansion of μ_{hy},the stratum mean of Y in the h^{th} stratum, about both lower and upper boundaries of the stratum, he showed that when the number of strata is large the approximations to optimum strata boundaries are obtained by finding the solutions of the system of equations given by

$$\left(y_h - y_{h-1}\right)^2 W_h = \text{constant}, \qquad \text{where } h= 1 , 2 , ..., L.$$

This system of equations is equivalent to the system $\sigma_{hy}^2 W_h$ = constant and, therefore, the points satisfying the latter set of equations can also be taken as modified forms of the equations for dealing with variables having infinite range. For this allocation method also, he has given the method of iteration for obtaining successively better approximations to the optimum strata boundaries.

Another method of finding approximations to the optimum points of stratification for the case of optimum allocation was proposed by Durbin (1959) in a review of Dalenius doctoral thesis. Let F (y) be the cumulative density function of the estimation variable Y. Forming a rectangular distribution

$$r(y) = \frac{F(y_L)}{(y_L - y_o)}$$

over the same range ,the approximations to optimum boundaries are obtained by taking equal intervals on the cumulative of $\frac{1}{2}\left[r(y) - f(y)\right]$.The rule amounts to forming strata by taking equal areas under a frequency distribution with density half way between the original distribution and a rectangular distribution. Durbin developed the rule by considering, for L = 2, the simplest departure from a rectangular distribution, namely a linear function f(y) between f(o) =1-a and f(1) = 1+a ,where ‘a’ is small.

An alternative approach to tackle the problem of optimum stratification was tried by Sethi (1963).In place of finding AOSB by solving some suitably chosen system of equations (other than the minimal equations) and then using the method of iteration to arrive at the optimum dividing points, he thought of preparing readymade tables giving optimum stratification points for certain standard frequency distributions. In actual practice the distributions encountered are all discontinuous. But, if the populations are sufficiently large, they can very often be approximated by some standard continuous distributions. Thus, the tables of optimum stratification points for these standard distributions can profitably be used for the stratification of the actual populations. The distributions considered by him are the standard normal and a set of chi-square distributions. He has also given the method of obtaining optimum points of stratification for various gamma distributions from corresponding chi-square distributions. In his paper he has considered three allocation methods viz. proportional, equal and optimum. Tables of optimum strata boundaries have been given for all the distributions and the allocation methods considered. He also found that the equalization of strata totals as suggested by Mahalanobis (1952) and Hansen *et al.* (1953) do not lead to optimum points of stratification for any of the populations considered. On the other hand, the approximation suggested by Dalenius and Hodges is excellent for both equal and optimum allocations. It was also observed that optimum points of stratification for equal and optimum allocations almost coincide.

The problem of minimizing the variance of stratified simple random sampling estimate of population mean, when total cost of the process is fixed, was first considered by Ekman (1959a, 1960).In this case also it was assumed that the density function of the estimation variable is known for the population which was taken to be infinite so that the density function could be taken to be continuous. If the exact form of the distribution was not known, some analytical form f (y) to the density was described on the basis of some previous knowledge of the distribution of a closely correlated auxiliary variable or the information obtained from a pilot survey. Ekman took the cost of sampling a unit to be a function of the value of Y for that unit. The approximate solutions to the minimal equations were obtained under the assumption that the cost and density functions satisfied certain regularity conditions. Four systems of equations were suggested which on solving gave approximation to the optimum points of stratification. He proved that the variance of the estimate obtained by using the approximate boundaries and the minimum variance were asymptotically equal. He also considered the generalized variance function given by

$$S = \sum_{h=1}^{L} F\left(y_{h-1}, y_h\right)$$

with the function F (y_{h-1} , y_h) of the form

$$F\left(y_{h-1}, y_h\right) = \left(y_{h-1}, y_h\right)^{\lambda-1} \int_{y_{h-1}}^{y_h} f\left(t\right) dt \left[1 + O\left(y_h - y_{h-1}\right)^2\right]$$

where $\lambda > 0, \lambda \neq 1$ and the function f(y) satisfy certain regularity conditions. It was shown for this case that the system of minimal equations, to a certain degree of approximation, could be replaced by another set of equations given by

$F(y_{h-1},y_h)$ = constant, h= 1,2,3,...,L.

The solutions to these equations gave approximations to the exact solutions of the minimal equations. Both these sets of solutions were proved to be asymptotically equivalent. For this general case, a method of iteration was also given to obtain successively better approximations to the optimum points of stratification.

The problem of optimum stratification when two or more characters of the population are under study becomes highly complicated. It is because in the multicharacter situation, the strata are to be defined on the basis of joint variations of the characters and hence they are solid figures. Therefore, the problem of optimum stratification is the problem of optimum determination of both shapes and sizes of the strata. Ghosh (1963) has theoretically solved the problem of optimum stratification when two characters are under investigation and the strata are formed by lines parallel to the axes. Samanta (1965) considered the situation where strata are disjoined regions formed by a set of increasing rectangles. With this rule of stratification, he has solved the problem of optimum stratification in case of proportional allocation by minimizing the generalized variance of the sample estimates. The resulting minimal equations are of course complicated but a rapidly converging iterative procedure is available. A study of the efficiency of this system of subdivision has also been made. As a numerical illustration, a standardised bivariate normal population has been sub-divided into two strata for varying values of the correlation coefficient. The results obtained by him are of limited value since there may exist other rules of stratification leading to considerable decrease in variance of the estimates and in practical situations the stratification variables are generally different than estimation variables. In practice sometimes it may also not possible to neglect finite population correction as has been done by him.

Apart from theoretical investigations into the problem of optimum stratification for unicharacter and multicharacter cases, some empirical studies to compare the efficiency of the various approximations to the optimum points have also been made by several authors.

Cochran (1961) compared four methods of finding approximations to optimum points of stratification when a frequency function f(y) is to be subdivided into L strata. The optimum boundaries are defined as those that give minimum variance for the estimated population mean from a stratified sample of size n, with optimum choices of sample sizes in the individual strata. The four rules are: i) equal intervals on the cumulative of $\sqrt{f(y)}$, ii) to choose the boundaries so that $W_h \mu_{hy}$ = constant ,iii) Ekman's rule of taking $W_h (y_h - y_{h-1})$ = constant and iv) Durbin's rule of taking equal intervals on the cumulative of (r + f) ,where r is the rectangular distribution with same frequency as f.

These rules are compared for L = 2,3 and 4 on eight skew frequency distributions intended to be somewhat representative of those that occur in practice. It was found that rules i) and iii) performed consistently well while rule iv) was found to be satisfactory except on two highly skew distributions. Rule ii) was satisfactory only on four of the eight distributions. For the equal allocation method, he found that the optimum boundaries in this case differed little, if at all, from those for optimum allocation method.

While considering the problem of stratifying on the basis of auxiliary information, it was found that strata constructed by above rules were nearly as good as the best strata that could be set up from the earlier data.

Des Raj (1964) studied the performance of equal size stratification with equal allocation by comparing it with optimum stratification when same allocation of samples in different strata is used. The estimation variable and stratification variables were assumed to be same. By considering four different density functions it was concluded that in these distributions equal size stratification method gave poorer boundaries with increase in the number of strata. Even for exponential populations for which the rule is supposed to be at its best, it is not optimum or near optimum for large L. It produces variance twice of that obtained through optimum stratification. For right triangular the performance is still worse. It was also observed that the lowest stratum made by this procedure was always too large as compared with the stratum in the optimum case. The relative contribution of the lowest stratum to the total variance was maximum and it didn't decrease with an increase in the number of strata of equal aggregate size.

Hess *et al.* (1966) made an investigation into the efficiency of optimum stratification. Comparisons were made among four types of allocations viz. Neyman, equal, proportional to the stratum aggregates of the stratification variable and proportional to the number of sampling units with the corresponding optimum stratification. Gains from stratification were examined for several estimation variables.

Schneeberger (1966) solved the problem of optimum stratification in a very simple way by means of an iterative analogue computer. He considered the case when the samples are proportionally allocated to strata. For the case of triangular distribution with density $f(x) = 0.5\,(1-x)$, $-1 \le x \le 1$, results were obtained for 2, 3, ...,7 strata. The accuracy increased when proposed optimal stratification was employed.

Sethumadhavi (1966) considered five methods of finding the approximations to optimum strata boundaries for the case of optimum allocation method. Apart from the four rules considered by Cochran (1961), she also considered Sethi's iterative method for proportional allocation. She compared the efficiency of these methods on the data collected in the sample survey conducted on temperatee fruit crops in Mahasu district Himachal Pradesh in the year 1965-66.Both estimation and stratification variables were taken to be the same. She found that the four rules considered by Cochran gave nearly identical results. However, beyond two strata Ekman's rule excelled ,and equalization of strata totals,cumulative $\sqrt{f(y)}$ rule, Durbin's rule and Sethi's iterative method followed in that order of performance.

Nilson (1967) adopted the method of least squares for solving the problem of optimum stratification. A necessary condition for obtaining a minimum variance is given and an iterative procedure of finding it is assumed. The method is also generalized for other measures of dispersion than the quadratic deviation. The convergence rate of the iterative procedure is investigated for the case when the study variable has a rectangular distribution and then for an arbitrary continuous distribution.

The problem of general optimum stratification for the objective variable Y based on the concomitant variables using prior information was tackled in more detail by Taga (1967).He pointed out that in case of proportionate sample allocation to each stratum, the above mentioned optimum stratification reduces to the optimum decomposition of the distribution function H (Z) for the random variable Z = η(x) ,where η(x) is the regression function of Y on X. Further, a general method was given by which such an optimal stratification can be asymptotically obtained.

The use of Cum$\sqrt{f}$ rule has been suggested by Serfling (1968) under the assumption that the regression of estimation variable Y on the stratification variable X is linear with uncorrelated homoscedastic errors and nearly perfect correlation. Using this approximation, he was able to choose optimally, for fixed cost, the number of strata to be constructed and the total sample size to be used. Finally, he remarked that for a large budget, optimal stratified random sampling by a covariable X will yield confidence intervals with lengths in proportion $\left[k_x^*\left(1-\rho^2\right)\right]^{1/2}$ to these for optimal simple random sampling, where k_x^* is a parameter.

When the study variable follows a log-normal distribution, optimum strata boundaries were derived by Schaffer (1971) under the assumption that the variable of stratification is the same as variable taken into consideration.

Isil and Taga (1969) considered the problem of optimal stratification for multivariate distributions. For this purpose, they proposed hyper plane stratification, in the case of proportionate allocation, which minimizes the covariance matrix of an estimator $\bar{x}$ for the mean vector μ of a multivariate distribution F(x) in the sense of semi-order in the symmetric matrix space defined as A ≥ B if A - B is non-negative definite. Besides, the optimal stratification is given by the quadratic hyper surface stratification in the case of optimal allocation.

Singh and Sukhatme (1969) considered the problem of optimum stratification on a concomitant variable X in more general form and evolved various methods of finding approximate solutions to the minimal equations for Neyman and Proportional allocations. For this purpose they assumed the regression of the study variable Y on the auxiliary variable X to be of the form Y = C(X) + e such that E(e|x) = 0 and V(e|x) = ∅ (x) for all x in the range (a, b) with (b - a) < ∞.Under this set up they obtained the minimal equations under Neyman allocation as

$$\frac{\left(c\left(x_h\right)-\mu_{hc}\right)^2+\sigma_{hc}^2+\phi\left(x_h\right)+\mu_{h\phi}}{\sqrt{\sigma_{hc}^2+\mu_{h\phi}}}=\frac{\left(c\left(x_h\right)-\mu_{ic}\right)^2+\sigma_{ic}^2+\phi\left(x_h\right)+\mu_{i\phi}}{\sqrt{\sigma_{ic}^2+\mu_{i\phi}}}$$

and, under proportional, allocation ,the minimal equations were

$$c\left(x_h\right)=\frac{\mu_{hc}+\mu_{ic}}{2} \quad \text{, i = h+1 , h = 1,2,3,...,L-1}$$

These system of equations give AOSB in the sense of minimum variance. Among the several approximations to the above system of equations, as developed by them, the one which they judged to use in practice is to make

$$\int_{x_{h-1}}^{x_h}\sqrt[3]{g_1(t)\,f(t)}\,dt=\text{ constant , h = 1,2,...,L.}$$

Where $$g_1\left(t\right)=\frac{\phi^{/2}\left(t\right)+4\phi\left(t\right)c^{/2}\left(t\right)}{\left[\phi\left(t\right)\right]^{3/2}}$$

for Neyman allocation, and for proportional allocation this rule reduces to make

$$\int_{x_{h-1}}^{x_h} \sqrt[3]{c'^2(t) f(t)} dt = \text{ constant, h = 1,2,...,L.}$$

These rules are known as cumulative cube root rules. Through numerical illustrations they observed that these rules give better approximation to the OSB than those based on equal intervals.

Following Singh and Sukhatme (1969), Singh (1971a) made an investigation into the accuracy of various methodology relating to stratification on study variable.

In continuation to their earlier work, Singh and Sukhatme (1972a) developed certain asymptotic properties of the AOSB following the method of Ekman (1959).Singh and Sukhatme (1972b) considered the problem of optimum stratification on an auxiliary variable X, when the units from different strata are selected with PPSWR sampling scheme. Proceeding on the lines of their earlier work they obtained minimal equations giving OSB for Neyman allocation method. Methods to find AOSB have also been proposed.

When information on an auxiliary variable highly correlated with the study variable is available, the use of ratio and regression methods of estimation help in improving the precision of the estimate of population mean/total for the character under study. Singh and Sukhatme (1973) used this concept while considering the problem of optimum stratification and gave some methods of finding AOSB for Neyman allocation. They have shown that the problem of determining OSB with ratio and regression methods of estimation is a particular case of optimum stratification on the auxiliary variable for stratified simple random sampling estimate.

A new method for obtaining AOSB in case of proportional allocation, known as 'cum $\sqrt[3]{f(y)}$ rule', was proposed by Singh (1975a).The proposed method is easy to apply in practice and involves same order of approximations as is involved in Ekman's method. In the same investigation, he also provided an approximation to his proposed rule. Through numerical illustration, he has shown that the AOSB obtained from approximate cum $\sqrt[3]{f(y)}$ rule were quite close to those obtained from the cum $\sqrt[3]{f(y)}$ rule for the rectangular and right triangular distributions. Further, some difference in the boundaries for the exponential distribution was seen which might have occurred because of infinite range of the variable under consideration.

Singh (1975b) once again considered the problem of optimum stratification with varying probabilities of selection, but for the proportional and equal allocation methods. He has also shown empirically that the performance of equal allocation was found to be better than that of proportional allocation and practically equivalent to Neyman allocation.

An alternative method of stratification was proposed by Singh (1975c). He gave a new cum $\sqrt{f(c'^2 + \theta\phi)^{1/2}}$ rule, where $\theta = \frac{12L^2}{(b-a)^2}$ as a generalization to Dalenius cum√f rule. A numerical investigation into the relative efficiency of this rule with respect to the cum√f and cum $\sqrt[3]{p(x)}$ rules has also been made which indicated that the proposed rule and the cum $\sqrt[3]{p(x)}$ rule can, however, be used in more general situations.

Singh and Prakash (1975) considered the problem of optimum stratification on the auxiliary variable X for equal allocation. They proposed a cum f√∅ rule for obtaining AOSB.

Anderson *et al.* (1976) applied the theoretical framework developed by Singh and Sukhatme (1969) to a bivariate normal model to tabulate gain in precision due to stratification. They observed that for small-to-moderate number of strata, the cum√f rule performed better than cum ∛f rule.

If the stratification is carried out on an auxiliary variable, then taking equal intervals on the cum $\sqrt[3]{f}$ gives AOSB which compares favourably in certain situations with those determined by cum√f rule. This was suggested by Thomsen (1976).

The problem of determining AOSB on the auxiliary variable X, under equal allocation, using ratio and regression methods of estimation was considered by Singh (1977). It is shown that equal allocation is equally efficient to Neyman allocation as regards the methods of estimation considered by him.

Thomson (1977) made an attempt to study the effect of stratification by using two variables, say, Y and Z. The method consists in stratifying the population into r strata along X variable by using cun√f_1 method, and constructing s strata along Z variable by an equal partitioning of cum$\sqrt{f_2}$ method. The results indicated that in many practical situations the gain from using two stratifying variables over one is nontrivial.

Schneeberger and Gollar (1979) studied the feasibility of optimal stratification points according to Dalenius. It is well known that the necessary conditions for optimal strata boundaries by optimal allocation, given by Dalenius and Gurney, lead-by sampling fraction q = $\frac{n}{N}$ > 0 to non-feasible solutions, if only one of the conditions $n_h \leq N_h$ (h = 1,2,...,L) is violated. Subsequently, Schneeberger and Drefahl (1980) presented limits of feasible sampling fraction in optimal stratification. They found that for $q > q_c(L)$,where q_c (L) is the critical sampling fraction, the Dalenius Neyman allocation solution is not feasible.

Taguri (1980) has given OSB, minimum variance and efficiencies for five distributions for the cases: (1) The population parameter to be estimated is μ or σ^2 (2) The stratification method is extended to general stratification (GS) in addition to interval stratification (IS), (3) the sample allocation method for each stratum is proportional or Neyman allocation. (4) L = 2, 3, 4 or 5.He has numerically ascertained the empirical results, “GOS tends to coincide OIS”.

Jarque (1981) considered the use of cluster analysis for solving the problem of optimum stratification in multivariate sampling, a case study in which the states of Mexico are stratified with respect to nine socio-economic variables.

Taguri (1982a) made the study of so called “problem of robustness in optimal stratification”. He provided some tables and some suggestion which make theoretical results obtained in earlier studies be applicable in actual sampling problems. In the sequel, he reported some more findings (1982b, 1982c, 1982d) regarding this aspect. He pointed out the importance of the evaluation of the robustness of the optimum stratification method with respect to changes of (i) the distribution, (ii) the sample sizes in respective strata, and (iii) stratification points. These results have also been shown empirically.

By reducing the problem of optimum stratification to a nonlinear programme, Schneeberger and Drefahl (1982) made an investigation into the gain in precision by optimum stratification in dependence on sampling fraction.

Wang and Aggarwal (1984) while considering the problem of stratification under a particular pareto distribution, also considered the case when stratification and estimation variables are different but related by a single regression model.

Yadav and Singh (1984) considered the problem of finding OSB when sample sizes to different strata are allocated in proportion to strata totals of the auxiliary variable. Because of the implicit nature of the minimal equations obtained in this case, methods of obtaining their approximate solutions have been presented. A limiting expression for the variance of the estimate of population mean, as the number of strata become large, has also been obtained.

Iachan (1985) considered the problem of optimum stratification when stratification is on auxiliary variables and obtained the AOSB for Shellfish surveys. Sai and Taguri (1989) discussed this problem for equal and Neyman allocations and applied it to "The current Statistics of Commerce in Japan" which showed great improvement of the precision in estimation of population mean.

Pla, L (1991) showed the use of principal components analysis is to determine strata boundaries in multivariate sampling improves estimation of the vector mean. He utilized the first principal component as the stratification variate. The proposed method reduced the generalized and total variances, outperforms the univariate or bivariate procedures for total and linear variances of the mean vector. Not only this but also the examples with real data were analysed.

Mahajan *et al.* (1994) considered the problem of optimum stratification in case of sensitive variable for which the data are collected by scrambled randomised response technique. In this regard, they proposed a rule of finding AOSB.

Mandowara and Gupta (1994) made an attempt to obtain points of stratification for two or more stage designs with equal primary stage units and subsequent stage units. Stratification on the auxiliary variable when the study variable is closely related to the auxiliary variable has been made. The determination of OSB in these cases has been illustrated with the help of some known specific distributions.

Wywial (1995) discussed the problem of optimal stratification of population on the basis of auxiliary variables using clustering method. Optimum stratification is done in such a way that the mean square prediction error was minimal and it led to minimization of the spectral radius of the variance-covariance matrix of the auxiliary variables.

Rizvi *et al*. (2000) considered the optimum stratification for two characters using proportional method of allocation, by taking an auxiliary variable as stratification variable. By minimizing the generalized variance of the sample means of the study variables, minimal equations have been obtained under proportional method of allocation. Due to implicit nature of these equations, a cum $\sqrt[3]{R(x)}$ rule has been proposed for obtaining approximately optimum strata boundaries. Empirical studies have also been made on certain density functions.

Rivest (2002) suggested some iterative procedures to determine OSB. The algorithms require an initial approximation solution to strata and also there is no guarantee that the algorithm which we are used will provide the global minimum in the absence of a suitable approximate initial solution and the variance functions have more than one local minima.

Gunning and Horgan (2004) discussed a simple and practicable algorithm for constructing strata boundaries in such a way that the coefficient of variation is equal in each stratum which is derived for positively skewed populations. The new algorithm is shown to compare favourably with the cumulative square root frequency method given by Dalenius and Hodges (1957) and the Lavallee and Hidiroglou (1988) approximation methods for estimating the OSB.

Rizvi *et al*. (2004) extended the Singh's (1975a) problem for proportional allocation, when two variates are under study. A Cum $\sqrt[3]{R_3(x)}$ rule for obtaining AOSB has been provided. It has been shown theoretically as well as empirically that the use of stratification has inverse effect on the relative efficiency of probability proportional to size with replacement (PPSWR) as compared to unstratified PPSWR method when proportional method of allocation is envisaged. Further comparison showed that with increase in the number of strata the stratified simple random sampling is equally efficient as PPSWR.

Kozak *et al*. (2007) discussed a modern approach to stratification of a finite population and presented a general picture of univariate and multivariate stratification. They addressed issues such as strata geometry, an optimization function and constraints for it, dimensionality of stratification, approximate univarite stratification, the choice of an optimization method, to perform stratification, initial parameters to be employed in optimization-based stratification, and other population and stratification attributes such as subdivision of a population into domains, domain oriented approach and a take-all stratum.

Horgan (2010) tracked the progress of various methods used for obtaining optimum strata boundaries and asked where we are now and where to go from here in which he mentioned that modification of the geometric algorithm are necessary to address.

Mehta and Mandowara (2012) considered the problem of determining optimum strata boundaries for cluster sampling design considering unequal sizes of clusters. The minimal equations giving optimum strata boundaries by minimizing the variance of the estimator of the population mean. Sampling in each stratum being carried out independently by simple random sampling without replacement. These minimal equations were difficult to solve exactly and thus the approximate solution solutions to these minimal equations had been obtained for three allocation methods namely proportional, equal and Neyman allocation.

Mathew *et al.* (2013) made an attempt in the context of interest in precision. They made use of prior knowledge of the population and tried to put the population into series of homogeneous groups so as to increase the precision. The study was designed to investigate the efficiency of neyman allocation procedure over equal and proportional allocations.

Hidiroglou and Kozak (2017) have developed a comprehensive solution for computing stratification boundaries either in the context of fixed sample size or fixed coefficient of variation, given an allocation scheme. They showed that the two scenarios are equivalent: that is, given a fixed sample size, the boundaries will be the same as those achieved using the anticipated coefficient of variation, resulting from the fixed sample size. The same holds true if we start off with a fixed coefficient of variation. They have reviewed computing stratification boundaries either in the context of fixed sample size (*n-scenario*) or fixed coefficient of variation (*c-scenario*), given an allocation scheme. From this study and others (Baillargeon & Rivest, 2009, 2011) it follows that Kozak (2004) algorithm greatly reduces chances of convergence problems reported for the Sethi algorithm used in Lavallée & Hidiroglou (1988) algorithm, offers the best way to stratify univariate populations. It is worth stressing that this algorithm is *not* computationally intensive because it is a simple random search algorithm. The only problem is that it may find a local optimum instead of the global optimum. The main focus of the study was that national statistical agencies should use optimisation-based stratification and use approximate rules only as a source of starting points for the optimisation-based methods.

Mathematical Programming Technique

The first ever mathematical programming problem (MPP) was perhaps the problem of optimum allocation of limited resources recognized by economists in early 1930s. After World War II, in 1947 the United States Air Force team SCOOP (Scientific Computation of Optimum Programs) started intensive research on some optimum resource allocation problem which led to the development of the famous simplex method by George B. Dantzig for solving a linear programming problem.

Kuhn and Tucker (1951) developed the necessary conditions (which become sufficient also under special circumstances) to be satisfied by an optimal solution of an MPP. These conditions, known as K-T conditions, laid the foundation for great deal of later research and development in nonlinear programming techniques. Till date no single technique is available which can provide an optimal solution to every Non-linear programming problem (NLPP) like simplex method for linear programming problem (LPP). However, different methods are available for some special types of NLPPs. In 1959 Beale gave a method for solving convex quadratic programming problem (CQPP). One of the most powerful techniques for solving an NLPP is to transform it, by some means, into a form which permits the use of simplex method of LPP. Using K-T conditions, Wolfe in 1959 transformed the CQPP into an equivalent LPP to which simplex method could be applied with some additional restrictions on the vectors entering the basis at various iterations. A similar method was developed by Panne and Whinston (1964a, 1964b, 1966). Some other techniques for solving quadratic programming problems are due to Lemke(1962), Graves(1967),letcher (1971), Aggarwal(1974a,1974b), Finkbeiner and Kail (1978), Arshad *et al*. (1981), Khan *et al*. (1983) etc. Among other NLPP methods there are Gradient Methods and Gradient Projection Methods. Like simplex method of LPP these are iterative procedures in which at each step we move from one feasible solution to another in such a way that the value of the objective function is improved. Rosen (1960, 1961), Kelley (1960), Goldfarb (1969), Du *et al*. (1990), Lai *et al*. (1993) etc. gave gradient projection methods for nonlinear programming with linear and nonlinear constraints.

Des Raj (1956) determined the optimum probabilities of selection in sampling without replacement. The methods for the approximate solution to the separable programming problems are found in the works of Charnes and Cooper (1957), Markowitz and Manne (1957), Dantzig *et al*. (l958) etc. The principle of optimality enunciated by Bellman (1957) paved the way for the development of the dynamic programming technique which has been applied successfully for solving certain special types of MPPs. The dynamic programming became rich and more applicable by the contributions of several authors. Since dynamic programming technique has been used as the main tool in the present investigation for solving nonlinear programming problems arising in sampling, in the next section it is discussed in detail.

The problems in which decisions are to be made sequentially at different stages of solutions are called multistage decision problems. Many multistage decision problems can be formulated as an MPP. The dynamic programming technique is a computational procedure which is well suited for solving MPPs that may be treated as a multistage decision problem. The basic idea which led to the development of the dynamic programming technique was given by Richard Bellman in early 1950s.This is the approach used by various authors for obtaining optimum strata boundaries.

Buhler and Deutler in (1975) studied the problem of optimum stratification through dynamic programming. They found that by a suitable transformation a global optimal solution could be obtained through dynamic programming. The results have been illustrated empirically for discrete and continuous variables.

A comparative study of solution procedures as regards the determination of optimum stratification was made by Deutler in 1976.He considered two groups of solution methods including procedure of operations research. In both the groups it has been discussed which procedure are capable of finding the global optimum.

Khan *et al*. (1994) proposed a technique for determining the optimum number of strata using mathematical programming approach.

The problem of determining the optimum strata boundaries, when the main study variable is used as a stratification variable and a stratified sample, using Neyman allocation is to be selected to estimate the population mean, has been formulated as an MPP and explained by Khan *et al*. (2002).It has been shown that with some modification the MPP may be converted into a multistage decision problem that could be solved using dynamic programming technique.

Khan *et al.* (2003) examined the problem of determining an optimum compromise allocation in multivariate stratified random sampling, when the population means of several characteristics are to be estimated. Formulating the problem of allocation as an all integer nonlinear programming problem, they developed a solution procedure using a dynamic programming technique. The compromise allocation discussed is optimal in the sense that it minimizes a weighted sum of the sampling variances of the estimates of the population means of various characteristics under study.

Khan *et al*. (2005) studied the problem of optimum stratification and formulated as an MPP assuming exponential frequency distribution of the main study variable. The stratum boundaries are optimum in the sense that they minimize the sampling variance of the stratified sample mean under Neyman allocation. The formulated MPP found to be separable with respect to the decision variables and was treated as multistage decision problem. A solution procedure has also been developed using dynamic programming.

The problem of finding OSB has been considered as the problem of determining optimum strata width (OSW) by Khan *et al*. (2008).The problem has been formulated as MPP, which minimizes the variance of the estimated population parameter under Neyman allocation subject to the restriction that the sum of the widths of all the strata is equal to the total range of the distribution. The distribution of the study variable has been considered as continuous with triangular and standard normal density functions. The formulated MPPs which turn out to be multistage decision problems, can then be solved using dynamic programming technique proposed by Buhler and Deutler (1975).Comparative study revealed that proposed technique is better than Dalenius and Hodges (1959) method.

Khan *et al*. (2008) found that in order to make the strata internally homogenous, the strata should be constructed in such a way that the strata variance for the characteristic under study be as small as possible. This could be achieved effectively by having the known distribution of the main study variable and create strata by cutting the range of the distribution at suitable points. If the frequency distribution of the study variable is unknown, it may be approximated from the past experience or some prior knowledge obtained at a recent study. They considered the problem of finding optimum strata boundaries as the problem of determining optimum strata width.

Khan *et al*. (2009) proposed the method of choosing the best boundaries that make strata internally homogenous given some sample allocation, known as optimum allocation. In this paper the problem of OSB has been discussed when strata are formed based on a single auxiliary variable with a varying measurement cost per unit strata the auxiliary variable considered in the problem is a size variable that holds a common model for the population. The OSB were achieved effectively by assuming a suitable distribution of the auxiliary variable and creating strata by cutting the range of the distribution at suitable points. The problem of finding the OSB, which minimizes the variance of the estimated population mean under a weighted stratified balanced sampling, has been formulated as an MPP.

Sebnem (2011) concluded that having stratified many populations with different characteristics revealed that Kozak's random search method and Keskinturk and Sebnem's genetic algorithm give more or less the same results both from the point of coefficient of variations and boundaries. The results obtained there showed that either genetic algorithm or Kozak's random search method could be efficiently applied in order to obtain near optimum boundaries in stratified sampling.

Khan *et al*. (2011) developed a method of choosing the boundaries that have high level of precision. They took the problem of finding the OSB for a skewed population with standard log-normal distribution. The problem is then redefined as the problem of determining optimum strata width and is formulated as MPP that seeks minimization of the variance of the estimated population parameter under neyman allocation subject to the constraint that sum of the widths of all the strata is equal to the total range of the distribution.

Rao *et al*. (2012) proposed a technique of determining the multivariate calibrated estimator to improve the survey estimates when more than one auxiliary variable is available. The problem of determining optimum calibrated weights is formulated as an MPP, which is solved using Lagrange multiplier technique. The numerical illustration presented in the paper for computation details reveals that the proposed estimator performs better than the usual estimator.

Fonolahi and Khan (2014) proposed a technique to determine the optimum strata boundaries when the measurement cost per unit varies across the strata. The problem has been formulated as an MPP and solved to obtain the strata width, which is then used to obtain the OSB. The study variable is assumed to be exponentially distributed.

Rao *et al.* (2014) developed a technique of solving a combined problem of determining OSB and optimum strata size of each stratum, when the population under study is skewed and the study variable has a pareto frequency distribution. The problem is formulated as an MPP, which has been solved by using a dynamic programming technique.

The problem of construction optimum stratification for two study variables based on auxiliary variable that follow respectively a uniform and a right-triangular distribution has been discussed by Khan *et al.* (2014a).The problem of determining the OSB has been formulated as NLPP ,which turned out to be multistage decision problems and were solved using dynamic programming techniques. The comparative study has been done on simulated data of two sets and the results were obtained by cum√f method of Dalenius and Hodges (1959), geometric method by Gunning and Horgan (2004),the generalized method of Lavallee and Hidiriglous (1988) and proposed method, and it was found that construction of strata using auxiliary variable for the populations with uniform and right-triangular distributions, leads to substantial gains in precision of the estimates while using the proposed technique.

Khan *et al.* (2014b) proposed a technique by using auxiliary information for determining the optimum strata boundaries of the population that has uniformly distributed auxiliary variable. The problem has been formulated as NLPP that seek minimization of the variance of the estimated population parameter under Neyman allocation. The NLPP is then solved by developing a solution procedure using a dynamic programming technique.

In a stratified sampling design for economic surveys based on auxiliary information has been developed by Khan *et al.* (2015a), which could be used for constructing optimum stratification and determining optimum sample allocation to maximize the precision in estimate. They made comparison of the proposed method with Dalenius and Hodges (1959), Gunning and Horgan (2004) and generalized (1988) method with Kozak’s algorithm. It concluded that the construction of strata and determination of sample allocation using auxiliary variable of the populations with Gamma distribution, leads to substantial gains in the precision of the estimates while using proposed technique than others taken in comparison.

The problem of finding the OSB and the optimum sample sizes within the stratum for a skewed population with log-normal distribution has been studied by Khan *et al.* (2015b).The problem of determining the OSB has been redefined as the problem of determining optimum strata width (OSW) and is formulated as a NLPP that seeks minimization of the variance of the estimated population mean under Neyman allocation subject to the constraint that the sum of the widths of all the strata is equal to the range of the distribution. The formulated NLPP turns out to be a multistage decision problem that can be solved by dynamic programming technique. Furthermore, a comparison studied is carried out using 10 artificial populations to compare efficiency of the proposed method with the cum√f , geometric and L-H methods. The results in the study reveal that the proposed method and the L-H method are more efficient than cum√f and geometric methods in minimizing the variance of the estimate of the population mean.

Danish *et al.* (2017a) discussed the way of obtaining optimum strata boundaries when the cost of every unit varies in the whole strata. The problem has been formulated as non-linear programming problem which is solved by using Bellman's principle of optimality. They used the uniform distribution to support their numerical illustration.

Danish *et al.* (2017b) discussed that the strata may be constructed either on the basis of study variable itself or by using some other variable(s) closely related with study variable known as auxiliary variable. Over the past many years several computational methods have been developed to obtain stratification points. They review the contribution toward obtaining optimum strata boundaries using mathematical programming. For several methods the comparative study has been made.

Danish and Rizvi (2017) studied problem of finding OSB by taking into consideration as the problem of optimum strata width (OSW), using MPP by dynamic programming technique, when the study variable is uniformly distributed. The empirical study has been made where it is revealed that with the increase in the number of strata to a fixed number the precision of the method goes on increasing. Also the proposed method proves better than the method given by Singh and Sukhatme (1969).

CHAPTER 3

MATERIALS AND METHODS

The methods of sample surveys are generally required for those situations where in one is interested to draw inferences about the population characteristic(s) under study, on the basis of a part on an aggregate. The population being investigated may be homogenous or heterogeneous one with respect to the characteristics under study. In the latter case, stratified random sampling is used in which whole population is divided into subpopulations, known as strata. These strata are non-overlapping, and together they comprise the whole population. The process of dividing the heterogeneous population into homogenous strata, which comprise the whole population, is known as stratification and the stratification that results the minimum variance is known as optimum stratification. The present investigation is mainly concerned with the determination of strata boundaries. However, for this purpose other factors will also be taken into consideration. Once it is decided about the total number of strata and the procedure of allocating sample sizes to different strata, the problem of stratification may be considered to consist of determination of strata boundaries, which constitutes the construction of strata. Strata may be constructed either on the basis of study variable itself or by using some other variable closely related with the variable under study. Therefore, in the present study the problem of optimum stratification is considered by taking some appropriate auxiliary variables as stratification variable. In this study, we are taking two auxiliary variables as the basis of stratification with one study variables. The study variable Y has been linearly related to two auxiliary variables X and Z.

Notations

Let there be a finite population consisting of N units, for which it is required to estimate the total or mean for the characteristic Y under study, using simple random sampling technique. In order to have this, we divide the whole population into $L \times M$ strata on the basis of two auxiliary variables say, X and Z, such that the number of units in the $(h , k)^{th}$ stratum is N_{hk}. A sample of size 'n' is to be drawn from the whole population and suppose that the allocation of sample size to the $(h , k)^{th}$ stratum is n_{hk} $(h = 1,2,...,L ; k=1,2,...,M)$. The value of population unit in the $(h , k)^{th}$ stratum be denoted by y_{hki} (i = 1 , 2 , 3,…, N_{hk}).Since the study variable is denoted by 'Y'. The unbiased estimate of population mean $\overline{Y}$ is

$$\overline{y}_{st} = \sum_{h=1}^{L}\sum_{k=1}^{M} W_{hk}\overline{y}_{hk}$$

Where W_{hk} denotes the stratum weight for the (h , k)th .

For stratified simple random sampling, the sample estimate $\overline{y}_{st}$ is unbiased and its sampling variance is given as below:

$$V\left(\overline{y}_{st}\right) = \sum_{h}\sum_{k}(1 - f_{hk})\frac{W_{hk}^{2}\sigma_{hky}^{2}}{n_{hk}}$$

where f_{hk} denotes the sampling fraction in the $(h,k)^{th}$ stratum and

$$\sigma_{hky}^{2} = \frac{1}{N_{hk}}\sum_{i=1}^{N_{hk}}\left(y_{hki} - \overline{y}_{hk}\right)^{2}$$

Let

$$\sum_{h=1}^{L}\sum_{k=1}^{M} N_{hk} = N \quad ; \quad \sum_{h=1}^{L}\sum_{k=1}^{M} n_{hk} = n$$

$$Y = \sum_{h=1}^{L}\sum_{k=1}^{M}\sum_{i=1}^{N_{hk}} y_{hki} \quad ; \quad \overline{y}_{st} = \sum_{h=1}^{L}\sum_{k=1}^{M} w_{hk}\overline{y}_{hk}$$

$$w_{hk} = \frac{N_{hk}}{N} \quad ; \quad \overline{y}_{hk} = \frac{1}{n_{hk}}\sum_{i=1}^{n_{hk}} y_{hki}$$

$$f_{hk} = \frac{n_{hk}}{N_{hk}} \quad ;$$

However, under proportional allocation variance of the sample estimate $\overline{y}_{st}$ is given as

$$V\left(\overline{y}_{st}\right)_{prop} = \sum_{h}\sum_{k}(1 - f)\frac{W_{hk}\sigma_{hky}^{2}}{n}$$

Under equal allocation

$$V\left(\overline{y}_{st}\right)_{eq} = \frac{1}{nN^{2}}\sum_{h}\sum_{k} N_{hk}\left(LMN_{hk} - n\right)\frac{\sigma_{hky}^{2}}{n}$$

Under Neyman allocation

$$V\left(\bar{y}_{st}\right)_{Ney} = \frac{1}{n}\sum_h\sum_k\left(W_{hk}\sigma_{hky}\right)^2 - \sum_h\sum_k\frac{W_{hk}^2\sigma_{hky}^2}{N}$$

Under optimum allocation

$$V\left(\bar{y}_{st}\right)_{opt} = \frac{1}{n}\sum_h\sum_k\left(W_{hk}\sigma_{hky}\sqrt{C_{hk}}\right)\left(\sum_h\sum_k\frac{W_{hk}\sigma_{hky}}{\sqrt{C_{hk}}}\right) - \sum_h\sum_k\frac{W_{hk}^2\sigma_{hky}^2}{N}$$

where C_{hk} be the cost of observing the variable y in the $(h,k)^{th}$ stratum. Let the regression model of Y on X and Z be given as

Y = C (X, Z) + e

Where C (X, Z) is a function of X and Z and 'e' is error term such that $E(e\,|\,x,z) = 0$ and $V(e\,|\,x,z) = \eta(x,z) > 0$, $\forall$ $x \in (a,b)$ $z\varepsilon(c,d)$, (b – a) < ∞, (c – d) < ∞

Let the joint density function of (Y, X, Z) in the super population is f(y, x, z), joint marginal of X and Z is f(x, z) and the marginal density function of X and Z are f(x) and f(z), respectively. Then for the $(h,k)^{th}$ stratum, we have

$$W_{hk} = \int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} f(x,z)\,\partial x\partial z \qquad \text{and} \qquad \mu_{hky} = \mu_{hkc} = \frac{1}{W_{hk}}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} c(x,z)\,f(x,z)\,\partial x\partial z$$

Where $(x_{h-1}, x_h, z_{k-1}, z_k)$ is the boundary points of $(h,k)^{th}$ and $\mu_{hk\eta}$ is the expected value of the function $\eta(x,z)$ in the $(h,k)^{th}$ stratum and σ_{hkc}^2 is given as

$$\sigma_{hkc}^2 = \frac{1}{W_{hk}}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} c^2(x,z)\,f(x,z)\,\partial x\partial z - \left(\mu_{hkc}\right)^2$$

Let ρ denotes the correlation coefficient between X and Z and is equal to

$$\rho = \frac{\sigma_{XZ}}{\sigma_X\sigma_Z}$$

The comparison for the efficiency of the proposed methods and the existing methods have been done with the help of simulation data, simulated through R software one of the programme for generating data is

```
# Random numbers for normal variate
# sample size
n=1000
u=rnorm(n, mean=0, sd=1)
```

```
View(u)
```

However, for certain comparisons with the proposed methods the programme made in R software for executing the rule given by some authors like Dalenius and Hodges (1959),Gunning and Horgan (2004),Khan *et al.* (2008).

R Commands:

```
require(stratification)
rm(list=ls(all=TRUE))
### Sethi's algorithm versus Kozak's algorithm
# LACK OF CONVERGENCE
# Here is an example of numerical difficulties met with
Sethi but not with Kozak
UScities<-seq(1,500)
Sethi <- strata.LH(x=UScities, CV=0.01, Ls=3,
alloc=c(0.35,0.35,0), takenone=0, takeall=1,
rh=1, model="loglinear", model.control=list(beta=1,
sig2=0.5, ph=0.85),
algo="Sethi", algo.control=list(maxiter=20))
Sethi
Sethi$iter.detail[1:5,]
# Kozak's algorithm with arithmetic initial boundaries
# (default initial boundaries for Sethi's algorithm)
Kozak<-strata.LH(x=UScities, initbh=c(18,27), CV=0.01, Ls=3,
alloc=c(0.35,0.35,0),
takenone=0, takeall=1, rh=1, model="loglinear",
model.control=list(beta=1, sig2=0.5, ph=0.85), algo="Kozak")
Kozak
Kozak$iter.detail[Kozak$iter.detail[,"run"]==Kozak$run.min[1
],]
# Looking at the iteration history for the optimization with
Sethi and Kozak,
# we see that the initial boundaries are very close from the
optimal ones.
# Kozak reaches very quickly a minimum. However, Sethi
increases n instead of
# minimizing it and afterwards it oscillates between two
sets of boundaries
# without converging.
# LOCAL MINIMUM
# In this example, Sethi's algorithm obviously reaches a
local minimum since Kozak
```

```
# proposes a much smaller n.
############################################################
################
Sethi<-strata.LH(x=UScities, CV=0.01, Ls=4,
alloc=c(0.5,0,0), takenone=0, takeall=1,
rh=0.85, model="loglinear", model.control=list(beta=1.1,
sig2=0, ph=1),
algo="Sethi")
Sethi
Kozak<-strata.LH(x=UScities, CV=0.01, Ls=4,
alloc=c(0.5,0,0), takenone=0, takeall=1,
rh=0.85, model="loglinear", model.control=list(beta=1.1,
sig2=0, ph=1),
algo="Kozak")
Kozak
#########################################################
### Take-none stratum
# As illustrated in the following example (presented in
Baillargeon and Rivest 2011),
# it is sometimes beneficial to include a take-none stratum
in the stratified design
# (possibly with a bias penalty lower than 1).
MRTS<-seq(1,200)
notn <- strata.LH(x=MRTS, CV=0.1, Ls=3, alloc=c(0.5,0,0.5))
notn
tn1 <- strata.LH(x=MRTS, CV=0.1, Ls=3, alloc=c(0.5,0,0.5),
takenone=1)
tn1
tn0.5 <- strata.LH(x=MRTS, CV=0.1, Ls=3, alloc=c(0.5,0,0.5),
takenone=1, bias.penalty=0.5)
tn0.5
# Note: Sethi does not converge here. This occurs often with
a take-none stratum.
tn1.Sethi <- strata.LH(x=MRTS, CV=0.1, Ls=3,
alloc=c(0.5,0,0.5), takenone=1, algo="Sethi")
tn1.Sethi
#############################
```

Geometric Method

```
strata.cumrootf(x, n = NULL, CV = NULL, Ls = 3, certain =
NULL,
alloc = list(q1 = 0.5, q2 = 0, q3 = 0.5), rh = rep(1, Ls),
model = c("none", "loglinear", "linear", "random"),
model.control = list(), nclass = NULL)
strata.geo(x, n = NULL, CV = NULL, Ls = 3, certain=NULL,
```

```
alloc = list(q1 = 0.5, q2 = 0, q3 = 0.5), rh = rep(1, Ls),
model = c("none", "loglinear", "linear", "random"),
model.control = list())
```

##############

One of the basic programmes used for solving functions in Mathematica is

Integrate[Exp[-0.8294((x^2)+(z^2)-1.26zx)],{x,0,∞},{z,0, ∞}]

#########

Integrate[(0.609943(z^(1/2)x^(-3/2))+20.507599(z)^(-1/2)(x)^(-1/2))(Exp[-z+1])/x,{x,1.05,1.1},{z,1.05,1.1}]

CHAPTER 4

CLASSICAL OPTIMIZATION TECHNIQUE

4.1 Stratified Simple Random Sampling –The Unbiased Estimate

Let there be a finite population consisting of N units, for which it is required to estimate the total or mean for the characteristic Y under study, using simple random sampling technique. In order to have this, we divide the whole population into L × M strata on the basis of two auxiliary variables say, X and Z, such that the number of units in the $(h , k)^{th}$ stratum is N_{hk} so that

$$\sum_{h=1}^{L}\sum_{k=1}^{M} N_{hk} = N$$

A sample of size 'n' is to be drawn from the whole population and suppose that the allocation of sample size to the $(h , k)^{th}$ stratum is n_{hk} such that

$$\sum_{h=1}^{L}\sum_{k=1}^{M} n_{hk} = n$$

The value of population unit in the $(h , k)^{th}$ stratum be denoted by y_{hki} (i = 1 , 2 , 3,…, N_{hk}) and then the population total is

$$Y = \sum_{h=1}^{L}\sum_{k=1}^{M}\sum_{i=1}^{N_{hk}} y_{hki}$$

Since the study variable is denoted by 'Y'. The unbiased estimate of population mean $\overline{Y}$ is

$$\overline{y}_{st} = \sum_{h=1}^{L}\sum_{k=1}^{M} W_{hk}\overline{y}_{hk}$$

Where $W_{hk} = \frac{N_{hk}}{N}$ denotes the stratum weight for the $(h , k)^{th}$ and $\overline{y}_{hk} = \frac{1}{n_{hk}}\sum_{i=1}^{n_{hk}} y_{hki}$

For stratified simple random sampling, the sample estimate $\overline{y}_{st}$ is unbiased and its sampling variance is given as below:

$$V\left(\overline{y}_{st}\right) = \sum_{h}\sum_{k}\left(1 - f_{hk}\right)\frac{W_{hk}^2\sigma_{hky}^2}{n_{hk}}$$

where $f_{hk} = \frac{n_{hk}}{N_{hk}}$ denotes the sampling fraction in the $(h,k)^{th}$ stratum.

If the finite population correction (f.p.c) is ignored, the variance of the estimate is given by

$$V\left(\bar{y}_{st}\right)=\sum_{h}\sum_{k}\frac{W_{hk}^{2}\sigma_{hky}^{2}}{n_{hk}}$$

σ_{hky}^{2} represents the population variance for the character Y in the $\left(h,k\right)^{th}$ stratum and is defined as

$$\sigma_{hky}^{2}=\frac{1}{N_{hk}}\sum_{i=1}^{N_{hk}}\left(y_{hki}-\bar{y}_{hk}\right)^{2}$$

$\bar{y}_{hk}$ being the population mean of all the N_{hk} units in the $\left(h,k\right)^{th}$ stratum.

4.2 Sample allocation methods

The problem of allocation in stratified sampling concerns the choice of the sample size n_{hk} in the respective strata with fixed total sample size. Here, we shall take two methods of allocation viz. Proportional allocation and Optimum allocation with constant cost of observing a unit of each stratum.

4.2.1 Proportional allocation method

In this method, the sample size n_{hk} in the $\left(h,k\right)^{th}$ stratum is taken in proportion to the total number of units in that stratum i,e. $n_{hk} \propto N_{hk}$ and $\sum_{h=1}^{L}\sum_{k=1}^{M}n_{hk}=n$,we get

$$n_{hk}=\frac{nN_{hk}}{N}=nW_{hk}$$

Therefore, the variance of the estimator $\bar{y}_{st}$,under proportional method becomes

$$V\left(\bar{y}_{st}\right)_{p}=\frac{1}{n}\sum_{h}\sum_{k}W_{hk}\sigma_{hky}^{2} \qquad (4.2.1.1)$$

4.2.2 Method of optimum allocation

In this method, the sample sizes n_{hk} are determined in such a way that for the given total sample size (which amounts to fixed total cost when the cost of observing a unit in each stratum is same) .The variance of the estimate $\bar{y}_{st}$ is minimized. Thus, we have to minimize

$$V\left(\bar{y}_{st}\right)=\sum_{h}\sum_{k}\frac{W_{hk}^{2}\sigma_{hky}^{2}}{n_{hk}}$$

Subject to the constraint

$$\sum_{h=1}^{L}\sum_{k=1}^{M} n_{hk} = n \qquad (4.2.2.1)$$

We select n_{hk} and the Langrangean multiplier 'λ' so as to minimize the function

$$\psi(n_{hk},\lambda) = \sum_h \sum_k \frac{W_{hk}^2 \sigma_{hky}^2}{n_{hk}} + \lambda\left(\sum_h \sum_k n_{hk} - n\right)$$

Differentiating it with respect to n_{hk} and equating to zero, we get

$$\frac{\partial}{\partial n_{hk}}\psi(n_{hk},\lambda) = -\frac{W_{hk}^2 \sigma_{hky}^2}{n_{hk}^2} + \lambda = 0$$

$$\Rightarrow n_{hk}^2 = \frac{W_{hk}^2 \sigma_{hky}^2}{\lambda}$$

This gives $n_{hk} = \dfrac{W_{hk}\sigma_{hky}}{\sqrt{\lambda}}$

Using it in the constraint (4.2.2.1), we have

$$n = \sum_h \sum_k \frac{W_{hk}\sigma_{hky}}{\sqrt{\lambda}}$$

$$\Rightarrow \frac{1}{\sqrt{\lambda}} = \frac{n}{\sum_h \sum_k W_{hk}\sigma_{hky}}$$

Therefore,

$$n_{hk} = n\frac{W_{hk}\sigma_{hky}}{\sum_h \sum_k W_{hk}\sigma_{hky}}$$

Thus, for optimum method of allocation the variance of the estimate becomes

$$V\left(\bar{y}_{st}\right)_{opt} = \frac{\left(\sum_h \sum_k W_{hk}\sigma_{hky}\right)^2}{n} \qquad (4.2.2.2)$$

4.3 Optimum strata boundaries for character Y

Now let us consider the problem of determining the optimum strata boundaries (OSB) which correspond to the minimum variance. We assume that the finite population of N units in a random sample from the infinite super population with same population characteristics as those of the finite population. This assumption enables us to deal with the continuous functions. If the density function of the estimation variable Y in super population is known, the optimum points of stratification $\left[y_h, y_k\right]$ for the cases of optimum and proportional allocation methods are given by the solutions of the following two systems of equations respectively.

$$\frac{\sigma_{hky}^2 + \left(y_{hk} - \mu_{hky}\right)^2}{\sigma_{hky}^2} = \frac{\sigma_{ijy}^2 + \left(y_{hk} - \mu_{ijy}\right)^2}{\sigma_{ijy}^2}$$

and

$$y_{hk} = \frac{\left(\mu_{hky} + \mu_{ijy}\right)}{2}$$

y_{hk} is the common boundary points for $\left(h,k\right)^{th}$ and $\left(i,j\right)^{th}$ strata,
i=h+1, h= 1,2,…,L-1
j=k+1, k= 1,2,…,K-1
and μ_{hky} = population mean for Y in (h , k)th stratum.

The minimal equations as given above were obtained by Dalenius (1950) for the case where the estimation and stratification variable are same. But, in most of the cases the density function of Y is not known and, therefore, it is always desirable to obtain the equations giving optimum points of stratification for the auxiliary variable which are highly related to the study variable. This is possible because density functions of the auxiliary variables are generally known. In the present investigation two auxiliary variables (X and Z) have been taken as the basis of stratification. Now we shall obtain the equations, the solution to which will give the optimum points of stratification for the auxiliary variables X and Z. The variance corresponding to these optimum strata is also minimum. For doing this, we shall assume the knowledge of the regression of Y on X and Z and also knowledge of conditional variance function $V\left(y \mid x, z\right)$.

4.4 Variance expression

Let the regression model of Y on X and Z be given as

$$Y = C(X, Z) + e$$

where C (X, Z) is a function of X and Z and 'e' is error term such that

$E(e|x,z)=0$ and $V(e|x,z)=\eta(x,z)>0, \forall\ x\varepsilon(a,b)\ z\varepsilon(c,d)$, (b – a) < ∞ ,(c – d) < ∞

If the joint density function of (Y, X, Z) in the super population is f(y, x, z), joint marginal of X and Z is f(x, z) and the marginal density function of X and Z are f(x) and f(z), respectively, then under above regression model, we have

$W_{hk} = \int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} f(x,z)\partial x\partial z$ is the weight of the $(h,k)^{th}$ stratum.

$\mu_{hky} = \mu_{hkc} = \frac{1}{W_{hk}}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} c(x,z)\, f(x,z)\partial x\partial z$ denotes the mean of the $(h,k)^{th}$ stratum

and $$\sigma^2_{hky} = \sigma^2_{hkc} + \mu_{hk\eta}$$

Where $(x_{h-1}, x_h, z_{k-1}, z_k)$ is the boundary points of $(h,k)^{th}$ stratum and $\mu_{hk\eta}$ is the expected value of the function $\eta(x,z)$ in the $(h,k)^{th}$ stratum and σ^2_{hkc} is given as

$$\sigma^2_{hkc} = \frac{1}{W_{hk}}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} c^2(x,z)\, f(x,z)\partial x\partial z - (\mu_{hkc})^2$$

Using these relations, the variance for the estimate $\bar{y}_{st}$ as given in (4.2.1.1) and (4.2.2.2) can be expressed in terms of the population parameters of the function $C(x,z)$ and $\eta(x,z)$.The variance expression for the case of optimum, proportional and equal allocation are therefore given by

$$V(\bar{y}_{st})_{opt} = \frac{\left(\sum_h\sum_k W_{hk}\sqrt{\sigma^2_{hkc} + \mu_{hk\eta}}\right)^2}{n} \tag{4.4.1}$$

$$V(\bar{y}_{st})_{prop} = \frac{\left[\sum_h\sum_k W_{hk}\left(\sigma^2_{hkc} + \mu_{hk\eta}\right)\right]}{n} \tag{4.4.2}$$

and $$V\left(\bar{y}_{st}\right)_{eql} = \frac{\sum_h \sum_k W_{hk}^2\left(\sigma_{hkc}^2 + \mu_{hk\eta}\right)}{n} \tag{4.4.3}$$

4.5 Minimal equations

Let $[x_h, z_k]$ denotes the set of optimum points of stratification on the range (a, b) and (c, d) of X and Z , respectively, then corresponding to these strata boundaries as determined by the optimum points of stratification $[x_h, z_k]$ the variance of the estimate $\bar{y}_{st}$ is minimum. These points $[x_h, z_k]$ are the solutions of the minimal equations which are obtained by equating to zero the partial derivatives of $V\left(\bar{y}_{st}\right)$ with respect to x_h *and* z_k .Before deriving the minimal equations ,let us first find out the expression for some partial derivatives which will be helpful in obtaining the equations. We have for the $(h,k)^{th}$ stratum

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f(x,z)\,\partial x \partial z$$

Differentiate it w.r.t. x_h and z_k, we get

$$\frac{\partial}{\partial x_h} W_{hk} = \int_{z_{k-1}}^{z_k} f(x_h, z)\,\partial z$$

and

$$\frac{\partial}{\partial z_k} W_{hk} = \int_{x_{h-1}}^{x_h} f(x, z_k)\,\partial x$$

Also

$$\mu_{hk\eta} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \eta(x,z) f(x,z)\,\partial x \partial z$$

By differentiating it w.r.t. x_h and z_k, we get

$$\frac{\partial}{\partial x_h} \mu_{hk\eta} = \int_{z_{k-1}}^{z_k} \frac{f(x_h, z)}{W_{hk}} \left[\eta(x_h, z) - \mu_{hk\eta}\right] \partial z$$

$$\frac{\partial}{\partial z_k} \mu_{hk\eta} = \int_{x_{h-1}}^{x_h} \frac{f(x, z_k)}{W_{hk}} \left[\eta(x, z_k) - \mu_{hk\eta}\right] \partial x$$

Similarly

$$\frac{\partial}{\partial x_h}\mu_{hkc} = \int_{z_{k-1}}^{z_k} \frac{f(x_h,z)}{w_{hk}}\left[c(x_h,z)-\mu_{hkc}\right]\partial z$$

$$\frac{\partial}{\partial z_k}\mu_{hkc} = \int_{x_{h-1}}^{x_h} \frac{f(x,z_{hk})}{w_{hk}}\left[c(x,z_k)-\mu_{hkc}\right]\partial x$$

$$\frac{\partial}{\partial x_h}\sigma^2_{hkc} = \frac{1}{W_{hk}}\int_{z_{k-1}}^{z_k}\left\{f(x_h,z)\left[c(x_h,z)-\mu_{hkc}\right]^2-\sigma^2_{hkc}\right\}\partial z$$

$$\frac{\partial}{\partial z_k}\sigma^2_{hkc} = \frac{1}{W_{hk}}\int_{x_{h-1}}^{x_h}\left\{f(x,z_k)\left[c(x,z_k)-\mu_{hkc}\right]^2-\sigma^2_{hkc}\right\}\partial x$$

In the same way for the $(i,j)^{th}$ stratum, where i=h+1 and j=k+1, the expression for the corresponding partial derivatives with respect to its lower boundaries x_h *and* z_k are obtained as

$$\frac{\partial}{\partial x_h}W_{ik} = -\int_{z_{k-1}}^{z_k} f(x_h,z)\partial z$$

$$\frac{\partial}{\partial z_k}W_{hj} = -\int_{x_{h-1}}^{x_h} f(x,z_k)\partial x$$

$$\frac{\partial}{\partial x_h}\mu_{ik\eta} = -\int_{z_{k-1}}^{z_k} \frac{f(x_h,z)}{w_{ik}}\left[\eta(x_h,z)-\mu_{ik\eta}\right]\partial z$$

$$\frac{\partial}{\partial z_k}\mu_{hj\eta} = -\int_{x_{h-1}}^{x_h} \frac{f(x,z_k)}{w_{hj}}\left[\eta(x,z_k)-\mu_{hj\eta}\right]\partial x$$

$$\frac{\partial}{\partial x_h}\mu_{ikc} = -\int_{z_{k-1}}^{z_k} \frac{f(x_h,z)}{W_{ik}}\left[c(x_h,z)-\mu_{ikc}\right]\partial z$$

$$\frac{\partial}{\partial z_k}\mu_{hjc} = -\int_{x_{h-1}}^{x_h} \frac{f(x,z_k)}{W_{hj}}\left[c(x,z_k)-\mu_{hjc}\right]\partial x$$

$$\frac{\partial}{\partial x_h}\sigma^2_{ikc} = -\frac{1}{W_{ik}}\int\limits_{z_{k-1}}^{z_k}\left\{f(x_h,z)\left[c(x_h,z)-\mu_{ikc}\right]^2-\sigma^2_{ikc}\right\}\partial z$$

$$\frac{\partial}{\partial z_k}\sigma^2_{hjc} = -\frac{1}{W_{hj}}\int\limits_{x_{h-1}}^{x_h}\left\{f(x,z_k)\left[c(x,z_k)-\mu_{hjc}\right]^2-\sigma^2_{hjc}\right\}\partial x$$

Having found the relations, we shall proceed to obtain the minimal equations for the two allocation methods. The case of proportional allocation is considered first.

4.5.1 Proportional allocation

To obtain the minimal equations for this allocation method, we minimize the variance expression given (4.4.2).The minimization of it, is equivalent to minimization of

$$\sum_h\sum_k W_{hk}\sigma^2_{hkc}$$

Since $\sum_h\sum_k W_{hk}\mu_{hk\eta} = \mu_\eta$ which is population parameter and therefore is a fixed constant.

Let $V_p = \sum_h\sum_k W_{hk}\sigma^2_{hkc}$ (4.5.1.1)

Thus to obtain minimal equations, we minimize V_p by on equating the partial derivative of this expression taken with respect of x_h to zero, we get

$$\frac{\partial}{\partial x_h}V_p = \sum_k\left[W_{hk}\frac{\partial}{\partial x_h}\sigma^2_{hkc}+\sigma^2_{hkc}\frac{\partial}{\partial x_h}W_{hk}+W_{ik}\frac{\partial}{\partial x_h}\sigma^2_{ikc}+\sigma^2_{ikc}\frac{\partial}{\partial x_h}W_{ik}\right]=0$$

After further simplification, we get

$$\sum_k\left[W_{hk}\int_{z_{k-1}}^{z_k} f(x_h,z)\frac{\left\{\left[c(x_h,z)-\mu_{hkc}\right]^2-\sigma^2_{hkc}\right\}}{W_{hk}}\right]\partial z+\sigma^2_{hkc}\int_{z_{k-1}}^{z_k} f(x_h,z)\partial z$$

$$=\sum_k\left[W_{ik}\int_{z_{k-1}}^{z_k} f(x_h,z)\frac{\left\{\left[c(x_h,z)-\mu_{ikc}\right]^2-\sigma^2_{ikc}\right\}}{W_{ik}}\right]\partial z+\sigma^2_{ikc}\int_{z_{k-1}}^{z_k} f(x_h,z)\partial z$$

(4.5.1.2)

For obtaining minimal equations we also differentiate V_p partially w.r.t. z_k in similar way, we get

$$\sum_h \left[W_{hk} \int_{x_{h-1}}^{x_h} f(x, z_k) \frac{\left\{ \left[c(x, z_k) - \mu_{hkc} \right]^2 - \sigma_{hkc}^2 \right\}}{W_{hk}} \right] \partial x + \sigma_{hkc}^2 \int_{x_{h-1}}^{x_h} f(x, z_k) \partial x$$

$$= \sum_h \left[W_{hj} \int_{x_{h-1}}^{x_h} f(x, z_k) \frac{\left\{ \left[c(x, z_k) - \mu_{hjc} \right]^2 - \sigma_{hjc}^2 \right\}}{W_{hj}} \right] \partial x + \sigma_{hjc}^2 \int_{x_{h-1}}^{x_h} f(x, z_k) \partial x$$

(4.5.1.3)

However, for obtaining minimal equations we minimize V_p on equating the partial derivative of this expression with respect of x_h and z_k to zero, we get

$$W_{hk} f(x_h, z_k) \frac{\left\{ \left[c(x_h, z_k) - \mu_{hkc} \right]^2 - \sigma_{hkc}^2 \right\}}{W_{hk}} + f(x_h, z_k) \sigma_{hkc}^2$$

$$= W_{ij} f(x_h, z_k) \frac{\left\{ \left[c(x_h, z_k) - \mu_{ijc} \right]^2 - \sigma_{ijc}^2 \right\}}{W_{ij}} + f(x_h, z_k) \sigma_{ijc}^2$$

This gives the equation as

$$c(x_h, z_k) = \frac{\left(\mu_{hkc} + \mu_{ijc} \right)^2}{2}, \qquad \begin{array}{l} i = h+1, h = 1, 2, ..., L-1 \\ j = k+1, k = 1, 2, ..., M-1 \end{array} \tag{4.5.1.4}$$

If the function $\lambda(x, z) = c^{'}(x, z) f(x, z)$ belongs to the class Ω of functions, the solutions to the system of equation (4.5.1.4) give OSB in the sense of minimization of variance $V\left(\overline{y}_{st}\right)_{prop}$ .These equations are also very difficult to solve and, therefore, for

these equations also we shall find methods of obtaining approximation to the exact solutions $[x_h, z_k]$.Further better approximation can be obtained by using some approximate iterative procedures.

4.5.2 Optimum allocation

The minimization of the variance expression as given in (4.4.1) is equivalent to minimization of the expression

$$\sum_h \sum_k W_{hk} \sqrt{\sigma_{hkc}^2 + \mu_{hk\eta}} \tag{4.5.2.1}$$

Equating to zero, the partial derivative of this expression with respect to x_h ,we get

$$\sum_k \left[W_{hk} \frac{\partial}{\partial x_h}\left(\sqrt{h}\right) + \left(\sqrt{h}\right)\frac{\partial}{\partial x_h} W_{hk} + W_{ik} \frac{\partial}{\partial x_h}\left(\sqrt{i}\right) + \left(\sqrt{i}\right)\frac{\partial}{\partial x_h} W_{ik} \right] = 0 \tag{4.5.2.2}$$

where i=h+1, $(h) = \sigma^2_{hkc} + \mu_{hk\eta}$ and $(i) = \sigma^2_{ikc} + \mu_{ik\eta}$

Note that

$$\frac{\partial}{\partial x_h}(h) = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} f(x_h, z)\left\{\left[c(x_h, z) - \mu_{hk\eta}\right]^2 - \sigma^2_{hkc} + \eta(x_h, z) - \mu_{hk\eta}\right\}\partial z \tag{4.5.2.3}$$

Similarly, we have

$$\frac{\partial}{\partial x_h}(i) = -\frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} f(x_h, z)\left\{\left[c(x_h, z) - \mu_{ik\eta}\right]^2 - \sigma^2_{ikc} + \eta(x_h, z) - \mu_{ik\eta}\right\}\partial z \tag{4.5.2.4}$$

Now, using the result obtained in equation (4.5.2.2), on simplifications we get the minimal equations as

$$\sum_k \left[\frac{\int_{z_{k-1}}^{z_k} \left\{ f(x_h, z)\left[c(x_h, z) - \mu_{hkc}\right]^2 + \sigma^2_{hkc} + \eta(x_h, z) - \mu_{hk\eta} \right\}\partial z}{\sqrt{\sigma^2_{hkc} + \mu_{hk\eta}}} \right] = \sum_k \left[\frac{\int_{z_{k-1}}^{z_k} \left\{ f(x_h, z)\left[c(x_h, z) - \mu_{ikc}\right]^2 + \sigma^2_{ikc} + \eta(x_h, z) - \mu_{ik\eta} \right\}\partial z}{\sqrt{\sigma^2_{ikc} + \mu_{ik\eta}}} \right] \tag{4.5.2.5}$$

Similarly once again equating to zero, the partial derivative of the (4.5.2.1) expression w.r.t. z_k, we get

$$\sum_h \left[\frac{\int_{x_{h-1}}^{x_h} \left\{ f(x, z_k)\left[c(x, z_k) - \mu_{hkc}\right]^2 + \sigma^2_{hkc} + \eta(x, z_k) - \mu_{hk\eta} \right\}\partial x}{\sqrt{\sigma^2_{hkc} + \mu_{hk\eta}}} \right] = \sum_h \left[\frac{\int_{x_{h-1}}^{x_h} \left\{ f(x, z_k)\left[c(x, z_k) - \mu_{ikc}\right]^2 + \sigma^2_{ikc} + \eta(x, z_k) - \mu_{ik\eta} \right\}\partial x}{\sqrt{\sigma^2_{ikc} + \mu_{ik\eta}}} \right] \tag{4.5.2.6}$$

However, the partial derivative of the (4.5.2.1) expression w.r.t x_h and z_k , we get

$$\frac{f(x_h, z_k)\left[c(x_h, z_k) - \mu_{hkc}\right]^2 + \sigma^2_{hkc} + \eta(x_h, z_k) - \mu_{hk\eta}}{\sqrt{\sigma^2_{hkc} + \mu_{hk\eta}}} = \frac{f(x_h, z_k)\left[c(x_h, z_k) - \mu_{ijc}\right]^2 + \sigma^2_{ijc} + \eta(x_h, z_k) - \mu_{ij\eta}}{\sqrt{\sigma^2_{ijc} + \mu_{ij\eta}}} \tag{4.5.2.7}$$

where i=h+1,j=k+1,h=1,2,…,L and k=1,2,…,M

Here, it should be noted that the system of equations that we can obtain from (4.5.2.7) gives the stratification points $\left[x_h, z_k\right]$ which correspond to the minimum of the variance $V\left(\bar{y}_{st}\right)_{opt}$ if the function

$$\lambda(x,z) = f(x,z)\frac{\left[4\eta(x,z)c'^{2}(x,z)+\eta'^{2}(x,z)\right]}{\left[\eta(x,z)\right]^{\frac{3}{2}}}$$

belongs to the class of Ω of functions for x in [a, b] and z in [c, d] and if $\lambda(x,z)\in\Omega \qquad \forall x\in[a,b] \quad \& \quad z\in[c,d]$.Then the system of equations (4.5.2.7) gives $\left[x_h, z_k\right]$ that minimizes the variance $V\left(\bar{y}_{st}\right)_{opt}$.

These equations on solving give the optimum points of stratification $\left[x_h, z_k\right]$. As the parameter involved (4.5.2.7) are themselves function of x_h *and* z_k ,the exact solutions are, therefore, very difficult to obtain. Due to this difficulty, it becomes extremely desirable to find some approximate solutions of this system of equations and then use some iteration procedure to obtain better approximations to $\left[x_h, z_k\right]$,if so desired.

4.5.3 Equal allocation

To obtain minimal equations for this allocation method, we minimize the variance expression given in (4.4.3).The minimization of this variance is equivalent to the minimization of the expression

$$\sum_h\sum_k W_{hk}^2\sigma_{hkc}^2$$

Equating to zero the partial derivative of this expression w.r.t. x_h,we get

$$\sum_k\left[W_{hk}^2\frac{\partial}{\partial x_h}\sigma_{hkc}^2+\sigma_{hkc}^2\frac{\partial}{\partial x_h}W_{hk}^2+W_{ik}^2\frac{\partial}{\partial x_h}\sigma_{ikc}^2+\sigma_{ikc}^2\frac{\partial}{\partial x_h}W_{ik}^2\right]=0$$

Using the values obtained in previous equations and after simplification, we get

$$\sum_k\left[W_{hk}^2\int_{z_{k-1}}^{z_k}\frac{1}{W_{hk}}\left\{f(x_h,z)\left[c(x_h,z)-\mu_{hkc}\right]^2+\sigma_{hkc}^2+\eta(x_h,z)-\mu_{hk\eta}\right\}\partial z\right]$$
$$=\sum_k\left[W_{ik}^2\int_{z_{k-1}}^{z_k}\frac{1}{W_{ik}}\left\{f(x_h,z)\left[c(x_h,z)-\mu_{ikc}\right]^2+\sigma_{ikc}^2+\eta(x_h,z)-\mu_{ik\eta}\right\}\partial z\right]$$

Similarly, equating to zero the partial derivative of the same equation w. r. t. z_k, we get

$$\sum_h \left[W_{hk}^2 \int_{x_{h-1}}^{x_h} \frac{1}{W_{hk}} \left\{ f(x,z_k)\left[c(x,z_k)-\mu_{hkc}\right]^2 + \sigma_{hkc}^2 + \eta(x,z_k) - \mu_{hk\eta} \right\} \partial x \right]$$
$$= \sum_h \left[W_{ik}^2 \int_{x_{h-1}}^{x_h} \frac{1}{W_{ik}} \left\{ f(x,z_k)\left[c(x,z_k)-\mu_{ikc}\right]^2 + \sigma_{ikc}^2 + \eta(x,z_k) - \mu_{ik\eta} \right\} \partial x \right]$$

While partially differentiating w.r.t. x_h and z_k, the same equation and equating to zero, we get

$$W_{hk} \left\{ f(x_h,z_k)\left[c(x_h,z_k)-\mu_{hkc}\right]^2 + \sigma_{hkc}^2 + \eta(x_h,z_k) - \mu_{hk\eta} \right\}$$
$$= W_{ij} \left\{ f(x_h,z_k)\left[c(x_h,z_k)-\mu_{ijc}\right]^2 + \sigma_{ijc}^2 + \eta(x_h,z_k) - \mu_{ij\eta} \right\}$$

where i=h+1,j=k+1,h=1,2,…,L and k=1,2,…,M

These equations are also very difficult to solve and, therefore, for this case also we shall find methods of obtaining approximation to the exact solutions $[x_h, z_k]$.Further better approximation can be obtained by using some approximate iterative procedures.

4.6 Some Miscellaneous results

Before we prove the results, let us impose certain regularity conditions on the functions $f(x,z), c(x,z) \quad and \quad \eta(x,z)$.We say that a function $\zeta(x,z)$ belongs to the class Ω if it satisfies

i) $0 < \zeta(x,z)$

ii) $\zeta(x,z) < \infty$

iii) $\zeta(x,z), \zeta'(x,z) \quad and \quad \zeta''(x,z)$ exist and are continuous for all x and z in (a, b) and (c, d), respectively. Where $(b-a) < \infty \quad and \quad (d-c) < \infty$.

We shall assume that the functions $f(x,z) \quad and \quad \eta(x,z)$belong to the class of Ω and the function c(x, z) satisfies the conditions (ii) and (iii).

Before we proceed to prove the results, let us define the symbol 'O' which have been used in the present thesis.

If the two functions $T_1(x,z)$ *and* $T_2(x,z)$ of X and Z, are such that the ratio $T_1(x,z)/T_2(x,z)$ remains bounded as x and z tends to their limits. We write $T_1(x,z) = O(T_2(x,z))$.

Lemma 4.1: If the function $I_{ij}(x,z)$ is defined as

$$I_{ij}(x,z) = \int_{z_1}^{z_2}\int_{x_1}^{x_2}(t_1-x_1)^{i}(t_2-z_1)^{j} f(t_1,t_2)\partial t_1\partial t_2, x_1 < x_2 \,\&\, z_1 < z_2$$

where $f(t_1,t_2)$ is a function of two variables, then

$$I_{ij}(x,z) = \left[\begin{array}{l} \dfrac{k_1^{i+1}k_2^{j+1}}{(i+1)(j+1)}f + \dfrac{k_1^{i+1}k_2^{j+1}}{(i+1)(j+1)}f_x + \dfrac{k_1^{i+1}k_2^{j+2}}{(i+1)(j+2)}f_z \\ +\dfrac{1}{2!}\left[\dfrac{k_1^{i+3}k_2^{j+1}}{(i+3)(j+1)}f_{xx} + 2\dfrac{k_1^{i+2}k_2^{j+2}}{(i+2)(j+2)}f_{xz} + \dfrac{k_1^{i+1}k_2^{j+3}}{(i+1)(j+3)}f_{zz}\right] + O\left(k^{i+j+5}\right) \end{array}\right] \tag{4.6.1}$$

where $f(t_1,t_2) = f, \dfrac{\partial f}{\partial t_1} = f_x, \dfrac{\partial f}{\partial t_2} = f_z, \dfrac{\partial^2 f}{\partial t_1^2} = f_{xx}, \dfrac{\partial^2 f}{\partial t_2^2} = f_{zz}, \dfrac{\partial^2 f}{\partial t_1 \partial t_2} = f_{xz}$ and k_1 an k_2 are the range of 1st and 2nd stratum.

Proof: In order to prove the lemma, if (t_1, t_2) is near (x_1 , z_1) and derivation of 'f' are continuous, then we can expand f(t_1,t_2) with the help of Taylor's series.

Let us define

$$t_1 = x_1 + (t_1 - x_1) \text{ and } t_2 = z_1 + (t_2 - z_1)$$

then $f(t_1,t_2) = f\left(x_1 + (t_1 - x_1), z_1 + (t_2 - z_1)\right)$

Using Taylor's Series expansion for the function of two variables, the expansion of f(t_1,t_2) is given by

$$f(t_1,t_2) = f(x_1,z_1) + (t_1 - x_1)\frac{\partial f}{\partial t_1} + (t_2 - z_1)\frac{\partial f}{\partial t_2} + \frac{(t_1 - x_1)^2}{2!}\frac{\partial^2 f}{\partial t_1^2} + \frac{(t_2 - z_1)^2}{2!}\frac{\partial^2 f}{\partial t_2^2} + \ldots$$ In

this way, utilizing it we get

$$I_{ij}(x,z)=\int_{z_1}^{z_2}\int_{x_1}^{x_2}(t_1-x_1)^i(t_2-z_1)^j\left[\begin{array}{l} f(x_1,z_2)+(t_1-x_1)\dfrac{\partial f}{\partial t_1}+(t_2-z_1)\dfrac{\partial f}{\partial t_2} \\ +\dfrac{(t_1-x_1)^2}{2!}\dfrac{\partial^2 f}{\partial t_1^2}+\dfrac{(t_2-z_1)^2}{2!}\dfrac{\partial^2 f}{\partial t_2^2}+\ldots \end{array}\right]$$

$$I_{ij}(x,z)$$

$$=\left[\begin{array}{l} \dfrac{k_1^{i+1}k_2^{j+1}}{(i+1)(j+1)}f(x_1,z_1)+\dfrac{k_1^{i+2}k_2^{j+1}}{(i+2)(j+1)}\dfrac{\partial f}{\partial t_1}+\dfrac{k_1^{i+1}k_2^{j+2}}{(i+1)(j+2)}\dfrac{\partial f}{\partial t_2} \\ +\dfrac{1}{2!}\dfrac{k_1^{i+3}k_2^{j+1}}{(i+3)(j+1)}\dfrac{\partial^2 f}{\partial t_1^2}+\dfrac{1}{2!}\dfrac{k_1^{i+1}k_2^{j+3}}{(i+1)(j+3)}\dfrac{\partial^2 f}{\partial t_2^2}+\dfrac{1}{2!}\dfrac{2k_1^{i+2}k_2^{j+2}}{(i+2)(j+2)}\dfrac{\partial^2 f}{\partial t_1\partial t_2}+O\left(k^{i+j+5}\right) \end{array}\right] \text{at}$$

$t_1 = x_1$ and $t_2 = z_1$.

where $k_1 = x_2 - x_1$ and $k_2 = z_2 - z_1$

Let us denote

$$f(x_1,z_1)=f,\ \frac{\partial f}{\partial t_1}=f_x,\ \frac{\partial f}{\partial t_2}=f_z,\ \frac{\partial^2 f}{\partial t_1^2}=f_{xx},\ \frac{\partial^2 f}{\partial t_2^2}=f_{zz},\ \frac{\partial^2 f}{\partial t_1\partial t_2}=f_{xz}$$

Then the above expression can be written as

$$I_{ij}(x,z)=\left\{\begin{array}{l} \dfrac{k_1^{i+1}k_2^{j+1}}{(i+1)(j+1)}f+\dfrac{k_1^{i+2}k_2^{j+1}}{(i+2)(j+1)}f_x+\dfrac{k_1^{i+1}k_2^{j+2}}{(i+1)(j+2)}f_z+ \\ \dfrac{1}{2!}\left[\dfrac{k_1^{i+3}k_2^{j+1}}{(i+3)(j+1)}f_{xx}+\dfrac{k_1^{i+1}k_2^{j+3}}{(i+1)(j+3)}f_{zz}+\dfrac{2k_1^{i+2}k_2^{j+2}}{(i+2)(j+2)}f_{xz}\right]+O\left(k^{i+j+5}\right) \end{array}\right\} \tag{4.6.2}$$

where ‘k’ indicates k_1 or k_2

Now for i = j = 0,we have

$$I_{00}(x,z)=\left\{k_1k_2 f+\frac{k_1^2k_2}{(2)}f_x+\frac{k_1k_2^2}{(2)}f_z+\frac{1}{2!}\left[\frac{k_1^3k_2}{(3)}f_{xx}+\frac{k_1k_2^3}{(3)}f_{zz}+\frac{2k_1^2k_2^2}{(2)(2)}f_{xz}\right]+O\left(k^5\right)\right\} \tag{4.6.3}$$

Lemma 4.2:-Let $\mu_\eta(x,z)$ denotes the conditional expectation of the function $\eta(t_1,t_2)$,so that

$$\mu_\eta(x,z)=\frac{\int_{z_1}^{z_2}\int_{x_1}^{x_2}\eta(t_1,t_2)f(t_1,t_2)\partial t_1\partial t_2}{\int_{z_1}^{z_2}\int_{x_1}^{x_2}f(t_1,t_2)\partial t_1\partial t_2}$$

Then, the series expansion of $\mu_\eta(x,z)$ at point (t_1 , t_2)is given by

$$\mu_\eta(x,z)=\eta\left[\begin{array}{l}1+\frac{\eta'}{2\eta}(k_1+k_2)+\left(\frac{\eta'(f_x+f_z)+2f\eta''}{12f\eta}\right)(k_1+k_2)^2\\+\left(\frac{f(f_{xx}+f_{zz}+f_{xz})\eta'+f(f_x+f_z)\eta''+f^2\eta'''-\eta'(f_x+f_z)^2}{12f^2\eta}\right)(k_1+k_2)^3\\+O\left((k_1+k_2)^4\right)\end{array}\right]$$

(4.6.4)

Proof: We have given

$$\mu_\eta(x,z)\int_{z_1}^{z_2}\int_{x_1}^{x_2}f(t_1,t_2)\partial t_1\partial t_2=\int_{z_1}^{z_2}\int_{x_1}^{x_2}\eta(t_1,t_2)f(t_1,t_2)\partial t_1\partial t_2$$

Therefore $\qquad \mu_\eta(x,z)I_{00}(x,z)=\int_{z_1}^{z_2}\int_{x_1}^{x_2}\eta(t_1,t_2)f(t_1,t_2)\partial t_1\partial t_2$

Using the Taylors series expansion defined for two variables of $\eta(t_1,t_2)$ about the point $(t_1,t_2)=(x,z)$ is given as

$$\eta(t_1,t_2)=\left[\begin{array}{l}\eta+(t_1-x)\eta'+(t_2-z)\eta'+\frac{(t_1-x)^2}{2!}\eta''+\frac{(t_2-z)^2}{2!}\eta''+\frac{2(t_1-x)(t_2-z)}{2!}\eta''+\\ \frac{1}{3!}\left[(t_1-x)^3\eta'''+(t_2-z)^3\eta'''+(t_1-x)^2(t_2-z)\eta'''+(t_1-x)(t_2-z)^2\eta'''+\ldots\right]+\ldots\end{array}\right]$$

Thus, we have

$$\mu_\eta(x,z)I_{00}(x,z)=\int_{z_1}^{z_2}\int_{x_1}^{x_2}(s_1)f(t_1,t_2)\partial t_1\partial t_2$$

$$=I_{00}(x,z)\eta+I_{10}(x,z)\eta'+I_{01}(x,z)\eta'+I_{20}(x,z)\eta''+I_{02}(x,z)\eta''+I_{11}(x,z)\eta''+o\left(k^5\right)$$

where

$$s_1=\left[\eta+(t_1-x)\eta'+(t_2-z)\eta'+\frac{(t_1-x)^2}{2!}\eta''+\frac{(t_2-z)^2}{2!}\eta''+\frac{2(t_1-x)(t_2-z)}{2!}\eta''+\ldots\right]$$

while neglecting the higher order terms, it can be written as

$$\mu_\eta(x,z)=\eta+\left[\frac{I_{01}(x,z)\eta' + I_{20}(x,z)\eta'' + I_{02}(x,z)\eta'' + I_{11}(x,z)\eta''}{I_{00}(x,z)}\right] \tag{4.6.5}$$

Now in order to elaborate (4.6.5), we need to substitute values of $I_{00}(x,z), I_{10}(x,z)$, $I_{01}(x,z)$, $I_{20}(x,z), I_{02}(x,z)$ and $I_{11}(x,z)$ for which we need to substitute the values of 'i' and 'j' in equation (4.6.2) ,we get

$$I_{10}(x,z)=\frac{k_1^2k_2}{2}f+\frac{k_1^2k_2}{3}f_x+\frac{k_1^2k_2^2}{4}f_z+\frac{1}{2!}\left[\frac{k_1^4k_2}{4}f_{xx}+\frac{k_1^3k_2^2}{3}f_{xz}+\frac{k_1^2k_2^3}{6}f_{zz}\right]+O(k^6)$$

$$I_{01}(x,z)=\frac{k_1k_2^2}{2}f+\frac{k_1^2k_2^2}{4}f_x+\frac{k_1k_2^3}{3}f_z+\frac{1}{2!}\left[\frac{k_1^3k_2^2}{6}f_{xx}+\frac{k_1^2k_2^3}{3}f_{xz}+\frac{k_1k_2^4}{4}f_{zz}\right]+O(k^6)$$

$$I_{20}(x,z)=\frac{k_1^3k_2}{3}f+\frac{k_1^4k_2}{4}f_x+\frac{k_1^3k_2^2}{6}f_z+O(k^6)$$

$$I_{02}(x,z)=\frac{k_1k_2^3}{3}f+\frac{k_1^2k_2^3}{6}f_x+\frac{k_1k_2^4}{4}f_z+O(k^6)$$

$$I_{11}(x,z)=\frac{k_1^2k_2^2}{4}f+\frac{k_1^3k_2^2}{6}f_x+\frac{k_1^2k_2^3}{6}f_z+O(k^6)$$

$$I_{00}(x,z)=k_1k_2f+\frac{k_1^2k_2}{2}f_x+\frac{k_1k_2^2}{2}f_z+\frac{1}{2!}\left(\frac{k_1^3k_2}{3}f_{xx}+\frac{2k_1^2k_2^2}{4}f_{xz}+\frac{k_1k_2^3}{3}f_{zz}\right)$$

where $k_1=x_2-x_1$ and $k_2=z_2-z_1$

substituting the results obtained as above in equation (4.6.5) ,we get

$$\mu_\eta(x,z)=\eta\left[\begin{array}{l} 1+\frac{\eta'}{2\eta}(k_1+k_2)+\left(\frac{\eta'(f_x+f_z)+2f\eta''}{12f\eta}\right)(k_1+k_2)^2 \\ +\left(\frac{f(f_{xx}+f_{zz}+f_{xz})\eta'+f(f_x+f_z)\eta''+f^2\eta'''-\eta'(f_x+f_z)^2}{24f^2\eta}\right)(k_1+k_2)^3 \\ +O\left((k_1+k_2)^4\right) \end{array}\right]$$

Proceeding in the same fashion but using Taylor's expansion about the point 'z' , the expansion for $\mu_\eta(x,z)$ is obtained as

$$\mu_\eta(x,z)=\eta\left[\begin{array}{l}1-\frac{\eta'}{2\eta}(k_1+k_2)+\left(\frac{\eta'(f_x+f_z)+2f\eta''}{12f\eta}\right)(k_1+k_2)^2\\ -\left(\frac{f(f_{xx}+f_{zz}+f_{xz})\eta'+f(f_x+f_z)\eta''+f^2\eta'''-\eta'(f_x+f_z)^2}{24f^2\eta}\right)(k_1+k_2)^3\\ +O\left((k_1+k_2)^4\right)\end{array}\right]$$

(4.6.6)

where the η, f functions and their derivatives are evaluated at point z.

Also

$$\mu_\eta(x,z)I_{00}(x,z)=\left[\eta I_{00}(x,z)+\sum_h\sum_k\frac{\eta^{(ij)}I_{ij}(x,z)}{(ij)!}\right] \tag{4.6.7}$$

Lemma 4.3: If $\sigma_\eta^2(x,z)$ denotes the conditional variance of the function $\eta(t_1,t_2)$ in the interval (x, z) ,so that

$$\sigma_\eta^2(x,z)=\mu_\eta^2(x,z)-\left(\mu_\eta(x,z)\right)^2$$

Then,

$$\begin{aligned}\sigma_\eta^2(x,z)&=\frac{(k_1+k_2)^2}{12}\left(\eta'\right)^2\left[1+\frac{\eta''}{\eta'}(k_1+k_2)+O(k_1+k_2)^2\right]\\ &=\frac{(k)^2}{12}\left(\eta'\right)^2\left[1+\frac{\eta''}{\eta'}(k)^1+O(k)^2\right]\end{aligned} \tag{4.6.8}$$

where $(k)^1$ *and* $(k)^2$ denotes all k_i's with power '1' and '2' respectively.

Proof: The expression for $\mu_\eta^2(x,z)$ can be obtained from (4.6.4) and on replacing the function $\eta(t_1,t_2)$ by $\eta^2(t_1,t_2)$,we have

$$\mu_{\eta^2}(x,z)=\eta^2\left[\begin{array}{l}1+\frac{\eta'}{\eta}(k_1+k_2)+\left(\frac{\eta\eta'(f_x+f_z)+2f\left(\eta'^2+\eta\eta''\right)}{6f\eta^2}\right)(k_1+k_2)^2+(f_1)(k_1+k_2)^3\\ +O(k_1+k_2)^4\end{array}\right]$$

where

$$f_1 = \left(\frac{f\left(f_{xx}+f_{zz}+f_{xz}\right)\eta\eta' + 3f^2\eta'\eta'^2 + f\left(f_x+f_z\right)\eta''^2 + f\left(f_x+f_z\right)\eta'^2 + ff'\eta'^2 - \left(f_x+f_z\right)^2\eta\eta'}{12f^2\eta^2} \right)$$

and 'f' and 'η' are functions and their derivations are evaluated at the point $(t_1,t_2)=(x,z)$ Similarly from (4.6.4) ,we also have

$$\mu_\eta^2(x,z)=\eta^2\left[1+\frac{\eta'}{2\eta}(k_1+k_2)+\left(\frac{\eta'\left(f_x+f_z\right)+2f\eta''}{6f\eta}\right)(k_1+k_2)^2+(f_1)(k_1+k_2)^3+O(k_1+k_2)^4\right]^2$$

After simplifying it, we get

$$\mu_\eta^2(x,z)=\eta^2\left[1+\frac{\eta'}{2\eta}(k_1+k_2)+\left(\frac{4f\eta\eta''+3f\eta'^2}{12f\eta^2}\right)(k_1+k_2)^2+(f_2)(k_1+k_2)^3+O(k_1+k_2)^4\right]$$

where

$$f_2 = \left(\frac{f\left(f_{xx}+f_{zz}+f_{xz}\right)\eta\eta' + f\left(f_x+f_z\right)\eta\eta'' + f^2\eta\eta''' - \left(f_x+f_z\right)^2\eta\eta' + f\left(f_x+f_z\right)\eta'^2 + 2f^2\eta'\eta''}{12f^2\eta} \right)$$

and the functions f and η and their derivatives are evaluated at the point $(t_1,t_2)=(x,z)$.

Therefore, by substituting above equations, we get

$$\sigma_\eta^2(x,z)=\mu_\eta^2(x,z)-\left(\mu_\eta(x,z)\right)^2$$

$$\sigma_\eta^2(x,z)=\frac{(k_1+k_2)^2}{12}\eta'^2+\frac{\eta'\eta''}{12}(k_1+k_2)^3+O(k_1+k_2)^2$$

$$=\frac{(k_1+k_2)^2}{12}\left(\eta'\right)^2\left[1+\frac{\eta''}{\eta'}(k_1+k_2)^1+O(k_1+k_2)^2\right]$$

which can be written as

$$\sigma_\eta^2(x,z)=\frac{(k)^2}{12}\left(\eta'\right)^2\left[1+\frac{\eta''}{\eta'}(k)^1+O(k)^2\right]$$

where (k) and $(k)^2$ are all k_i's with power 1 and 2 respectively.

Hence the lemma.

Note: If we take the function of $\eta(t_1,t_2)=(t_1,t_2)$,we obtain

$$\sigma_{t_1,t_2}^2(x,z)=\frac{(k)^2}{12}\left[1+O(k)^2\right] \qquad (4.6.9)$$

$$\Rightarrow k=\sigma_\eta(x,z)\sqrt{12}\left(1+O(k)^2\right) \qquad (4.6.10)$$

Lemma 4.4:

$$\mu_\eta(x,z)=\frac{\sqrt{\eta(x)\eta(z)}\eta'\left(k_1^4k_2^2-k_1^2k_2^4\right)}{4k_1^2k_2^2f-2k_1^3k_2f_x}+O\left(k^7\right) \qquad (4.6.11)$$

Proof: To prove the lemma, we should use the relation obtained in (4.6.4) and (4.6.6) .From these, on multiplying and taking square roots on both sides, we obtain

$$\mu_\eta(x,z)$$

$$=\sqrt{\eta(x)\eta(z)}\left[\frac{\frac{\eta' f}{2}\left(k_1^2k_2+k_1k_2^2\right)}{k_1k_2f+k_1^2k_2f_x+\frac{k_1k_2^2f_z}{2}}+O\left(k^4\right)\right]^{\frac{1}{2}}\left[\frac{\frac{\eta' f}{2}\left(k_1^2k_2-k_1k_2^2\right)}{k_1k_2f-k_1^2k_2f_x-\frac{k_1k_2^2}{2}f_z}+O\left(k^4\right)\right]^{\frac{1}{2}}$$

On simplification, we get

$$\mu_\eta(x,z)=\frac{\sqrt{\eta(x)\eta(z)}\eta'\left(k_1^4k_2^2-k_1^2k_2^4\right)}{4k_1^2k_2^2f-2k_1^3k_2f_x}+O\left(k^7\right)$$

Hence the lemma.

Lemma 4.5:

$$\int_{z_1}^{z_2}\int_{x_1}^{x_2} f(t_1,t_2)\partial t_1\partial t_2=\frac{1}{xz}\int_{z_1}^{z_2}\int_{x_1}^{x_2}(t_1t_2)^2 f(t_1,t_2)\partial t_1\partial t_2\left[1+O(k^3)\right] \qquad (4.6.12)$$

Proof: In order to prove the lemma consider a function

$$\lambda(z)=\frac{1}{xz}\int_{z_1}^{z_2}\int_{x_1}^{x_2} f(t_1,t_2)\partial t_1\partial t_2-\int_{z_1}^{z_2}\int_{x_1}^{x_2} f(t_1,t_2)\partial t_1\partial t_2$$

we can see here

$$\lambda(x)=\left(\frac{\partial(\lambda(z))}{\partial z}\right)_{x=z}=\left(\frac{\partial^2(\lambda(z))}{\partial z^2}\right)_{x=z}=0$$

As there are the initial coefficients of k_1 or k_2 and k_1^2 or k_2^2 in the Taylor series expansion of $\lambda(z)$ about ‘n’, we find

$$\lambda(z)=O(k^3)$$

where k either k_1 or k_2

$$\int_{z_1}^{z_2}\int_{x_1}^{x_2} f(t_1,t_2)\partial t_1\partial t_2=\frac{1}{xz}\int_{z_1}^{z_2}\int_{x_1}^{x_2}(t_1t_2)^2 f(t_1,t_2)\partial t_1\partial t_2+O(k^3)$$

$$\therefore\int_{z_1}^{z_2}\int_{x_1}^{x_2} f(t_1,t_2)\partial t_1\partial t_2=\frac{1}{xz}\int_{z_1}^{z_2}\int_{x_1}^{x_2}(t_1t_2)^2 f(t_1,t_2)\partial t_1\partial t_2\left[1+O\left(k^3\right)\right]$$

Hence the lemma.

Lemma 4.6:

$$\left(k_1k_2\right)^{\lambda-1}\int_{z_1}^{z_2}\int_{x_1}^{x_2} f\left(t_1,t_2\right)\partial t_1\partial t_2=\left[\int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt[\lambda]{f\left(t_1,t_2\right)}\partial t_1\partial t_2\right]^{\lambda}\left[1+O(k^2)\right] \quad (4.6.13)$$

Proof: To prove the above lemma, we shall expand the term

$\left[\int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt[\lambda]{f\left(t_1,t_2\right)}\partial t_1\partial t_2\right]^{\lambda}$ in power of k_1 and k_2

Using Taylor's theorem and expanding $\sqrt[\lambda]{f(t)}$ about the point $\left(t_1,t_2\right)=\left(x,z\right)$, we obtain

$$\left[\int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt[\lambda]{f\left(t_1,t_2\right)}\partial t_1\partial t_2\right]^{\lambda}=\left[\int_{z_1}^{z_2}\int_{x_1}^{x_2}\left(\sqrt[\lambda]{f\left(t_1,t_2\right)}+\frac{\left(t_1-x_1\right)}{\lambda f^{1-\frac{1}{2}}}f_x+\frac{\left(t_2-z_1\right)}{\lambda f^{1-\frac{1}{2}}}f_z+O(k^2)\right)\partial t_1\partial t_2\right]^{\lambda}$$

$$=\left[k_1k_2\sqrt[\lambda]{f\left(t_1,t_2\right)}+\frac{k_1^2k_2^2}{4\lambda f^{1-\frac{1}{\lambda}}}f_x+\frac{k_1^2k_2^2}{4\lambda f^{1-\frac{1}{\lambda}}}f_z+O\left(k^5\right)\right]^{\lambda}$$

$$=\left(k_1k_2\right)^{\lambda} f\left(t_1,t_2\right)\left[1+\frac{k_1k_2}{4\lambda f^{1-\frac{1}{\lambda}}}\left(f_x-f_z\right)+O\left(k^3\right)\right]^{\lambda}$$

$$=\left(k_1k_2\right)^{\lambda-1}\int_{z_1}^{z_2}\int_{x_1}^{x_2} f\left(t_1,t_2\right)\partial t_1\partial t_2\left[1+O\left(k^2\right)\right]$$

This may be written as

$$\left(k_1k_2\right)^{\lambda-1}\int_{z_1}^{z_2}\int_{x_1}^{x_2} f\left(t_1,t_2\right)\partial t_1\partial t_2=\left[\int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt[\lambda]{f\left(t_1,t_2\right)}\partial t_1\partial t_2\right]^{\lambda}\left[1+O(k^2)\right]$$

Hence the lemma.

Lemma 4.7:

$$\begin{aligned}I_{00}\left(x,z\right)&\sqrt{\sigma_c^2\left(x,z\right)+\mu_\eta\left(x,z\right)}\\&=\int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt{\eta\left(t_1,t_2\right)}\,f\left(t_1,t_2\right)\partial t_1\partial t_1+n_1\left[1+O\left(k_1+k_2\right)^2\right]\end{aligned} \quad (4.6.14)$$

Where $n_1=\dfrac{\left(k_1+k_2\right)^2}{96}\displaystyle\int_{z_1}^{z_2}\int_{x_1}^{x_2}\left(\frac{4\eta c'^2+\eta'^2}{\sqrt[3]{\eta}}\right)_{\left(t_1,t_2\right)} f\left(t_1,t_2\right)\partial t_1\partial t_1$

Proof: In order to prove the above result the lema's already proved will be used. We shall prove (4.6.14) by taking first term on right hand side towards left and prove the required result. During its proof the functions η, c *and* f along with their derivatives are to be evaluated at point $(t_1,t_2)=(x,z)$.In other words we shall use for different parameters the series expansion about one variable $'t_1 = x'$ and the same result can be proved if one uses the expansion of $'t_2 = z'$.

From (4.6.4) and (4.6.8) ,we have

$$\sigma_c^2(x,z)+\mu_\eta(x,z)$$

$$=\eta\left[1+\frac{\eta'}{2\eta}(k_1+k_2)+\left(\frac{(f_x+f_z)\eta'+2f\eta''+fc'^2}{12f\eta}\right)(k_1+k_2)^2+n_2(k_1+k_2)^3+O(k_1+k_2)^4\right]$$

where $n_2=\left(\dfrac{f(f_{xx}+f_{zz}+f_{xz})\eta'+f(f_x+f_z)\eta''+f^2\eta'''-\eta'(f_x+f_z)^2+2f^2c'c''}{24f^2\eta}\right)$

Therefore, we have

$$\sqrt{\sigma_c^2(x,z)+\mu_\eta(x,z)}=\sqrt{\eta}\left[1+\frac{\eta'}{2\eta}(k_1+k_2)+g_1(k_1+k_2)^2+g_2(k_1+k_2)^3+O(k_1+k_2)^4\right] \tag{4.6.15}$$

where $g_1=\dfrac{4\eta\eta'(f_x+f_z)+8f\eta\eta''+4f\eta c'^2-3f\eta'^2}{96f\eta^2}$

and $g_2=\left\{\begin{array}{c}\dfrac{f(f_{xx}+f_{zz}+f_{xz})\eta'+f(f_x+f_z)\eta''+f^2\eta'''-\eta'(f_x+f_z)^2+2f^2c'c''}{48f^2\eta}\\ -\dfrac{\eta'\left(\eta'(f_x+f_z)+2f\eta''+fc'^2\right)}{96f\eta^2}+\dfrac{\eta'^3}{128\eta^3}\end{array}\right\}$

We have obtained the expansion of $I_{00}(x,z)$. Thus, multiplying it by (4.6.15) , we get

$$I_{00}(x,z)\sqrt{\sigma_c^2(x,z)+\mu_\eta(x,z)}$$

$$=(k_1+k_2)f\sqrt{\eta}\left[1+\left(\frac{f\eta'+2(f_x+f_z)\eta}{4f\eta}\right)(k_1+k_2)+m_1(k_1+k_2)^2+m_2(k_1+k_2)^3+O(k_1+k_2)^4\right]$$

where $m_1=\dfrac{16\eta\eta'(f_x+f_z)+8f\eta\eta''+4f\eta c'^2-3f\eta'^2+16\eta(f_{xx}+f_{zz}+f_{xz})}{96f\eta^2}$

$$m_2 = \frac{24\eta^2\left(f_{xx}+f_{zz}+f_{xz}\right)\eta' + 24\eta^2\left(f_x+f_z\right)\eta''' + 16f\eta^2 c'c'' - 10\eta\eta'^2\left(f_x+f_z\right)}{384f\eta^3} + \frac{3f\eta'^3 - 8f\eta\eta'\eta'' - 4f\eta\eta'c'^2 + 8\eta^2\left(f_x+f_z\right)c'^2}{384f\eta^3}$$

Replacing $\eta(t_1,t_2)$ obtained in (4.6.7) by $\sqrt{\eta(t_1,t_2)}$,we get the left hand side as

$$I_{00}(x,z)\mu_\eta(x,z) = \int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt{\eta(t_1,t_2)}\,f(t_1,t_2)\,\partial t_1\partial t_2$$

Thus, we have

$$\int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt{\eta(t_1,t_2)}\,f(t_1,t_2)\,\partial t_1\partial t_2 = \left[\sqrt{\eta}I_{00}(x,z) + \frac{\sum_i\sum_j\left(\sqrt{\eta}\right)^{ij}}{(ij)!}I_{ij}(x,z)\right] + O\left(k_1+k_2\right)^5$$ On

multiplication, we get

$$\int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt{\eta(t_1,t_2)}\,f(t_1,t_2)\,\partial t_1\partial t_2$$
$$= \left(k_1+k_2\right)f\sqrt{\eta}\left[1 + p_1\left(k_1+k_2\right) + p_2\left(k_1+k_2\right)^2 + p_3\left(k_1+k_2\right)^3 + O\left(k_1+k_2\right)^4\right] \tag{4.6.17}$$

where $p_1 = \dfrac{2\left(f_x+f_z\right)\eta + f\eta'}{4f\eta}$

$$p_2 = \frac{4\eta^2\left(f_{xx}+f_{zz}\right) + 6\eta\left(f_x+f_z\right)\eta' + 2f\eta\eta'' - 2\eta\eta'\left(f_x+f_z\right) - f\eta'^2}{24f\eta^2}$$

$$p_3 = \frac{12\eta^2\left(f_{xx}+f_{zz}+f_{xz}\right)\eta' + 12\eta^2\left(f_x+f_z\right)\eta'' - 6\eta\left(f_x+f_z\right)\eta'^2 - 6f\eta\eta'\eta'' + 4f\eta^2\eta''' + 3f\eta'^3}{192f\eta^3}$$

Now by subtraction, we have

$$I_{00}(x,z)\sqrt{\sigma_c^2(x,z) + \mu_\eta(x,z)} - \int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt{\eta(t_1,t_2)}\,f(t_1,t_2)\,\partial t_1\partial t_2$$
$$= \left(k_1+k_2\right)f\sqrt{\eta}\left[q_1\left(k_1+k_2\right)^2 + q_2\left(k_1+k_2\right)^3 + O\left(k_1+k_2\right)^4\right] \tag{4.6.18}$$

where $q_1 = \dfrac{4f\eta c'^2 + f\eta'^2}{96f\eta^2}$

$$q_2 = \frac{16 f\eta^2 c' c'' - 4 f\eta\eta' c'^2 + 8\eta^2 (f_x + f_z) c'^2 + 2\eta (f_x + f_x)\eta'^2 + 4 f\eta\eta'\eta'' - 3 f\eta'^3}{384 f\eta^3}$$

$$= \frac{1}{192 f\sqrt{\eta}} \frac{\partial}{\partial x \partial z}\left(\frac{4 f\eta c'^2 + f\eta'^2}{\eta^{\frac{3}{2}}} \right)$$

Thus (4.6.18) can be written as

$$I_{00}(x,z)\sqrt{\sigma_c^2(x,z) + \mu_\eta(x,z)} - \int_{z_1}^{z_2}\int_{x_1}^{x_2}\sqrt{\eta(t_1,t_2)}\, f(t_1,t_2)\,\partial t_1 \partial t_2$$

$$= \frac{(k_1+k_2)^2}{96}\int_{z_1}^{z_2}\int_{x_1}^{x_2}\left(\frac{4 f\eta c'^2 + f\eta'^2}{\eta^{\frac{3}{2}}} \right)_{(t_1,t_2)} f(t_1,t_2)\,\partial t_1 \partial t_2 \left[1 + O(k_1+k_2)^2\right]$$

Hence the lemma.

4.7 Minimal equations and their approximate solutions

In this section, we shall find the series expansion of the system of equations given in (4.5.2.7) about the point (x_h, z_k), the common boundary points of $(h,k)^{th}$ and $(h+1,k+1)^{th}$ strata in order to obtain their approximate solutions. To find the expansion of (4.5.2.7) we shall use the relations obtained in different Lemmas by replacing (x,z) by $((x_{h-1}, x_h),(z_{k-1}, z_k))$. Let us consider the development of right

hand side. The corresponding expansion for the L.H.S of the equation can be obtained from the expansion of the right side by merely changing the signs of the coefficients of odd powers of k_i *and* k_j, where $k_i = x_{h+1} - x_h$ and $k_j = z_{k+1} - z_k$, although the same result will be obtained if we develop the expansion of this side independently.

We have (4.6.4) after replacing x and z by x_h *and* x_{h+1}, respectively, we get

$$\mu_{ikc} = c\left[1 + \frac{c'}{2c}k_i + \left(\frac{c' f_x + 2 fc''}{12 fc}\right)k_i^2 + \left(\frac{ff_{xx}c' + ff_x c'' + f^2 c''' - c' f_x^2}{24 f^2 c}\right)k_i^3 + O(k_i)^4\right]$$

Similarly, replacing x and z by z_k *and* z_{k+1} respectively, we get

$$\mu_{hjc} = c\left[1 + \frac{c'}{2c}k_j + \left(\frac{c' f_z + 2 fc''}{12 fc}\right)k_j^2 + \left(\frac{ff_{zz}c' + ff_z c'' + f^2 c''' - c' f_z^2}{24 f^2 c}\right)k_j^3 + O(k_j)^4\right]$$

where the functions c, f and their derivatives like f_x, f_z, f_{xx}, f_{zz} etc are evaluated at x_h *and* z_k .Therefore we get

$$\left[\mu_{ikc}-c(x_h,z)\right]^2=c\left[\frac{c'}{2c}k_i+\left(\frac{c'f_z+2fc''}{12fc}\right)k_i^2+\left(\frac{ff_{zz}c'+ff_zc''+f^2c'''-c'f_z^2}{24f^2c}\right)k_i^3+O(k_i)^4\right]^2$$

$$=\frac{k_i^2}{4}\left[c'^2+\left(\frac{c'^2f_x+2fc'c''}{3f}\right)k_i+(f_5)k_i^2+O\left(k_i^3\right)\right]$$

$$=\left(\frac{c'k_i}{2}\right)^2\left[1+2\left\{\left(\frac{c'f_x+2fc''}{6fc'}\right)k_i+\left(\frac{ff_{xx}+ff_xc''+f^2c'''-c'f_x^2}{12f^2c'}\right)k_i^2\right\}+\left(\frac{\left(c'f_x+2fc''\right)^2}{36f^2c'^2}\right)k_i^2+O\left(k_i^3\right)\right] \quad (4.7.1)$$

where $f_5=\dfrac{6f_{xx}c'^2+10ff_xc'c''+6f^2c'c'''-5f_x^2c'^2+4f^2c''^2}{36f^2}$

Also (4.6.8) can be written as

$$\sigma_{ikc}^2=\frac{k_i^2}{4}\left[\frac{c'}{3}+\frac{c'c''}{3}k_i+O\left(k_i^2\right)\right] \quad (4.7.2)$$

Adding (4.7.1) and (4.7.2) ,we get

$$\left[\mu_{hjc}-c(x_h,z)\right]^2+\sigma_{ikc}^2=\frac{k_i^2}{12}\left[4c'^2+\left(\frac{c'f_x+3fc'c''}{f}\right)k_i+O\left(k_i^2\right)\right]$$

Also from (4.6.4) ,we can write as

$$\eta(x_h,z)+\mu_{ik\eta}=\eta\left[2+\frac{\eta'}{2\eta}k_i+\left(\frac{\eta'f_x+2f\eta''}{12f\eta}\right)k_i^2+\left(\frac{ff_{xx}\eta'+ff_x\eta''+f^2\eta'''-f_x^2\eta'}{24f^2\eta}\right)k_i^3\right]+O\left(k_i^4\right)$$

where, on the right hand side of this equation the functions f,η and their derivatives are evaluated at point 'x_h'.However, if they are evaluated at 'z_k'it would take the form as

$$\eta(x,z_k)+\mu_{hj\eta}=\eta\left[2+\frac{\eta'}{2\eta}k_j+\left(\frac{\eta'f_z+2f\eta''}{12f\eta}\right)k_j^2+\left(\frac{ff_{zz}\eta'+ff_z\eta''+f^2\eta'''-f_z^2\eta'}{24f^2\eta}\right)k_j^3\right]+O\left(k_j^4\right)$$

Thus, we get

$$\left[\mu_{hjc}-c\left(x_h,z\right)\right]^2+\sigma_{ikc}^2+\left(\eta\left(x_h,z\right)+\mu_{ik\eta}\right)$$

$$=2\eta\begin{bmatrix}1+\dfrac{\eta'}{4\eta}k_i+\left(\dfrac{4c'^2f+\eta' f_x+2f\eta''}{24f\eta}\right)k_i^2\\+\left(\dfrac{2fc'^2f_x+6f^2c'c''\eta+ff_{xx}\eta'+ff_x\eta''+f^2\eta'''-f_x^2\eta'}{48f^2\eta^2}\right)k_i^3\end{bmatrix}+O\left(k_i^4\right) \qquad (4.7.3)$$

Also, we have

$$\sigma_{ikc}^2+\mu_{ik\eta}=\eta\begin{bmatrix}1+\dfrac{\eta'}{2\eta}\left(k_i\right)+\left(\dfrac{f_x\eta'+2f\eta''+fc'^2}{12f\eta}\right)\left(k_i\right)^2\\+\left(\dfrac{ff_{xx}\eta'+ff_x\eta''+f^2\eta'''-\eta' f_x^2+2f^2c'c''}{24f^2\eta}\right)\left(k_i\right)^3+O\left(k_i\right)^4\end{bmatrix}$$

so that

$$\left(\sigma_{ikc}^2+\mu_{ik\eta}\right)^{-\frac{1}{2}}=\frac{1}{\sqrt{\eta}}\begin{bmatrix}1-\dfrac{1}{2}\begin{Bmatrix}\dfrac{\eta'}{2\eta}\left(k_i\right)+\left(\dfrac{f_x\eta'+2f\eta''+fc'^2}{12f\eta}\right)\left(k_i\right)^2\\+\left(\dfrac{ff_{xx}\eta'+ff_x\eta''+f^2\eta'''-\eta' f_x^2+2f^2c'c''}{24f^2\eta}\right)\left(k_i\right)^3\end{Bmatrix}\\+\dfrac{3}{8}\left\{\dfrac{\eta'}{2\eta}k_i+\left(\dfrac{f_x\eta'+2f\eta''+fc'^2}{12f\eta}\right)\left(k_i\right)^2\right\}^2-\dfrac{5}{16}\left(\dfrac{\eta'}{2\eta}\right)^3\left(k_i\right)^3+O\left(k_i\right)^4\end{bmatrix}$$

$$=\frac{1}{\sqrt{\eta}}\left[1-\frac{\eta'}{4\eta}k_i+\frac{\left(9f\eta'^2-4f_x\eta\eta'-8\eta\eta''-4f\eta c'^2\right)}{96f\eta^2}k_i^2-d_1k_i^3+O\left(k_i\right)^4\right] \qquad (4.7.4)$$

where $d_1=\dfrac{\begin{pmatrix}8f\eta^2f_{xx}\eta'+8f\eta^2f_x\eta''+8f^2\eta^2\eta'''-8\eta^2f_x^2\eta'+16f^2\eta^2c'c''\\-12f\eta f_x\eta'^2-24f^2\eta\eta'\eta''-12f^2\eta\eta'c'^2+15f^2\eta'^3\end{pmatrix}}{384f^2\eta^3}$

Now, using the expression in (4.7.3) and (4.7.4) ,and on multiplying them, we get

$$\frac{\left[\mu_{ikc}-c\left(x_h,z\right)\right]^2+\sigma_{ikc}^2+\left(\eta\left(x_h,z\right)+\mu_{ik\eta}\right)}{\left(\sigma_{ikc}^2+\mu_{ik\eta}\right)^{\frac{1}{2}}}=2\sqrt{\eta}\left[1+\left(M_1\right)k_i^2+\left(M_2\right)k_i^3+O\left(k_i^4\right)\right]$$

where $$M_1=\frac{4\eta c'^2+\eta'^2}{32\eta^2}$$

$$M_2=\frac{8f_x\eta^2c'^2+16f\eta^2c'c''+2\eta f_x\eta'^2+4f\eta\eta'\eta''-4f\eta\eta'c'^2-3f\eta'^3}{192f\eta^3}$$

that can be further expressed as

$$M_2=\frac{1}{96f\sqrt{\eta}}\frac{\partial}{\partial x_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)$$

However, the R.H.S of the equation obtained in (4.5.2.5) can be expressed as

$$\frac{\left[\mu_{ikc}-c\left(x_h,z\right)\right]^2+\sigma_{ikc}^2+\left(\eta\left(x_h,z\right)+\mu_{ik\eta}\right)}{\left(\sigma_{ikc}^2+\mu_{ik\eta}\right)^{\frac{1}{2}}}$$

$$=2\sqrt{\eta}\left[1+\left(\frac{4\eta c'^2+\eta'^2}{32\eta^2}\right)k_i^2+\frac{k_i^3}{96f\sqrt{\eta}}\frac{\partial}{\partial x_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_i^4\right)\right] \quad (4.7.5)$$

In the similar way the expansion of L.H.S of the same equation can be obtained. The expansion is given by

$$\frac{\left[\mu_{hkc}-c\left(x_h,z\right)\right]^2+\sigma_{hkc}^2+\left(\eta\left(x_h,z\right)+\mu_{hk\eta}\right)}{\left(\sigma_{hkc}^2+\mu_{hk\eta}\right)^{\frac{1}{2}}}$$

$$=2\sqrt{\eta}\left[1+\left(\frac{4\eta c'^2+\eta'^2}{32\eta^2}\right)k_h^2-\frac{k_h^3}{96f\sqrt{\eta}}\frac{\partial}{\partial x_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_h^4\right)\right] \quad (4.7.6)$$

where the function f,η,c and their derivatives are evaluated at x_h.

Similarly, we would get the results when the functions f,η,c and their derivatives are evaluated at z_k. Thus equations (4.7.5) and (4.7.6) can take the form as

$$\frac{\left[\mu_{hjc}-c\left(x,z_k\right)\right]^2+\sigma_{hjc}^2+\left(\eta\left(x,z_k\right)+\mu_{hj\eta}\right)}{\left(\sigma_{hjc}^2+\mu_{hj\eta}\right)^{\frac{1}{2}}}$$

$$=2\sqrt{\eta}\left[1+\left(\frac{4\eta c'^2+\eta'^2}{32\eta^2}\right)k_j^2+\frac{k_j^3}{96f\sqrt{\eta}}\frac{\partial}{\partial z_k}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_j^4\right)\right] \tag{4.7.7}$$

and

$$\frac{\left[\mu_{hkc}-c\left(x,z_k\right)\right]^2+\sigma_{hkc}^2+\left(\eta\left(x,z_k\right)+\mu_{hk\eta}\right)}{\left(\sigma_{hkc}^2+\mu_{hk\eta}\right)^{\frac{1}{2}}}$$

$$=2\sqrt{\eta}\left[1+\left(\frac{4\eta c'^2+\eta'^2}{32\eta^2}\right)k_k^2-\frac{k_k^3}{96f\sqrt{\eta}}\frac{\partial}{\partial z_k}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_k^4\right)\right] \tag{4.7.8}$$

The equation (4.5.2.5), after cancelling the common terms on both the sides and multiplying both sides by $f\left(x_h,z\right)$,can be put as

$$\begin{aligned}&k_h^2\left[\left(\frac{4\eta c'+f\eta'^2}{\eta^2}\right)-\frac{k_h}{3}\frac{\partial}{\partial x_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_h\right)^2\right]\\&\quad=k_i^2\left[\left(\frac{4\eta c'+f\eta'^2}{\eta^2}\right)+\frac{k_i}{3}\frac{\partial}{\partial x_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_i\right)^2\right]\end{aligned} \tag{4.7.9}$$

where i=h+1 and the functions f,η,c and their derivatives are to be taken at x_h.

Similarly when the functions f,η,c and their derivatives are evaluated at z_k then we have (4.5.2.5) by substituted values obtained in (4.7.7) and (4.7.8) as

$$\begin{aligned}&k_k^2\left[\left(\frac{4\eta c'+f\eta'^2}{\eta^2}\right)-\frac{k_k}{3}\frac{\partial}{\partial z_k}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_k\right)^2\right]\\&\quad=k_j^2\left[\left(\frac{4\eta c'+f\eta'^2}{\eta^2}\right)+\frac{k_j}{3}\frac{\partial}{\partial z_k}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O\left(k_j\right)^2\right]\end{aligned} \tag{4.7.10}$$

Combining (4.7.9) and (4.7.10) and simplifying or in other words the functions f, η, c and their derivatives are evaluated at x_h and z_k, we get the results as

$$\left(\frac{f\eta'^2 + 4f\eta c'^2}{\eta^{\frac{3}{2}}}\right)\left[\left(k_h k_k\right)^2 \int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1 \partial t_2 \left[1+O\left(k_h k_k\right)^2\right]\right]^{\frac{2}{3}}$$

$$= \left(\frac{f\eta'^2 + 4f\eta c'^2}{\eta^{\frac{3}{2}}}\right)\left[\left(k_i k_j\right)^2 \int_{x_h}^{x_{h+1}}\int_{z_k}^{z_{k+1}} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1 \partial t_2 \left[1+O\left(k_i k_j\right)^2\right]\right]^{\frac{2}{3}} \quad (4.7.11)$$

where

$$m_1(t_1,t_2) = \frac{\eta'^2(t_1,t_2) + 4\eta(t_1,t_2)\,c'^2(t_1,t_2)}{\left(\eta(t_1,t_2)\right)^{\frac{3}{2}}}, \quad \begin{array}{l} i = h+1, h = 1,2,\dots,L \\ j = k+1, k = 1,2,\dots,M \end{array} \quad (4.7.12)$$

and we have assumed that the function

$$f(x,z), m_1(x,z) \in \Omega \,\forall x \in [a,b], z \in [c,d]$$

In case, if we have large number of strata so that the strata widths k_h *and* k_k are small then the higher powers of the widths can be neglected and the system of minimal equations given in (4.5.2.7) can be written as

$$\left[\left(k_h k_k\right)^2 \int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1 \partial t_2\right]^{\frac{2}{3}}$$

$$= \left[\left(k_i k_j\right)^2 \int_{x_h}^{x_{h+1}}\int_{z_k}^{z_{k+1}} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1 \partial t_2\right]^{\frac{2}{3}} \quad (4.7.13)$$

In other words it can be written that

$$\left(k_h k_k\right)^2 \int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1 \partial t_2 = \text{constant} \quad (4.7.14)$$

where terms of $O\left(\underset{(a,b)}{Sup}\left(k_h\right)\right)^4$ *and* $O\left(\underset{(c,d)}{Sup}\left(k_k\right)\right)^2$ have been neglected on both the sides of the equation. Since $\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1 \partial t_2 = O\left(\underset{((a,b),(c,d))}{Sup}\left(k_h k_k\right)\right)$ in view of the fact that $\forall x \in [a,b], z \in [c,d], 0 < m_1(x,z)\, f(x,z)$. It can be seen from (4.7.13) that, if we have a function $P_1(x_{h-1}, x_h, z_{k-1}, z_k)$ is such that

$$\left(k_h k_k\right)\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1\left(t_1,t_2\right) f\left(t_1,t_2\right)\partial t_1\partial t_2 = P_1\left(x_{h-1},x_h,z_{k-1},z_k\right)\left[1+O\left(k_i k_j\right)^2\right]$$

and $P_1\left(x_{h-1},x_h,z_{k-1},z_k\right)$ is of order $O\left(\underset{\left(\left(a,b\right),\left(c,d\right)\right)}{Sup}\left(k_h k_j\right)\right)^3$,then the minimal equations (4.5.2.5) can be obtained ,to the same degree of approximation as involved in (4.7.13) take the form as

$Q_1\left(x_{h-1},x_h,z_{k-1},z_k\right)$ = Constant, h=1,2,..,L,k=1,2,...,M (4.7.15)

The solutions to the sets of equations (4.5.2.5) as an approximation to optimum $\left[x_h,z_k\right]$ can be obtained with the help of some iterative procedure where there approximate solutions can be taken as the starting points.

Theorem 4.1: If the regression of the estimation variable Y on the stratification variables X and Z, in the infinite super population, is given by

$$y = c\left(x,z\right)+e$$

where 'e' is the error term such that $E\left(e\,|\,x,z\right)=0$ & $V\left(e\,|\,x,z\right)=\eta\left(x,z\right)>0$ $\forall x\in\left(a,b\right), z\in\left(c,d\right)$ with non zero deviation of intervals, and further if the function $m_1\left(t_1,t_2\right) f\left(t_1,t_2\right)\in\Omega$,then the system of equations (4.5.2.7) giving strata boundaries $\left[x_h,z_k\right]$ which correspond to the minimum of $V\left(\bar{y}_{st}\right)_{opt}$ can be written as

$$\left[\left(k_h k_k\right)^2\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1\left(t_1,t_2\right) f\left(t_1,t_2\right)\partial t_1\partial t_2.\left[1+O\left(k_h k_k\right)^2\right]\right]^{\frac{2}{3}}$$

$$=\left[\left(k_i k_j\right)^2\int_{x_h}^{x_{h+1}}\int_{z_k}^{z_{k+1}} m_1\left(t_1,t_2\right) f\left(t_1,t_2\right)\partial t_1\partial t_2.\left[1+O\left(k_i k_j\right)^2\right]\right]^{\frac{2}{3}}$$

while neglecting the terms of order $O\left(\underset{\left(\left(a,b\right),\left(c,d\right)\right)}{Sup}\left(k_h k_j\right)\right)^4$,these equations can be replaced by the approximate system of equations

$$\left(k_h k_k\right)^2\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1\left(t_1,t_2\right) f\left(t_1,t_2\right)\partial t_1\partial t_2 = \text{Constant}$$

or equivalently

$$Q_1\left(x_{h-1},x_h,z_{k-1},z_k\right) = \text{constant}$$

Where $m_1(x,z)=\left(\dfrac{\eta'^2+4\eta c'^2}{\eta^{\frac{3}{2}}}\right)_{(x,z)}$, $\begin{array}{l} k_h = x_h - x_{h-1} \\ k_k = z_k - z_{k-1} \end{array}$

$$Q_1(x_{h-1},x_h,z_{k-1},z_k)\left[1+O(k_hk_k)^2\right]=(k_hk_k)^2\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\,f(t_1,t_2)\,\partial t_1\partial t_2$$

$and \quad i=h+1, h=1,2,...,L, \quad j=k+1, k=1,2,...,M$

From Lemma 4.7 with x and z replaced by (x_{h-1},x_h) *and* (z_{k-1},z_k) respectively so that $k=k_h, k=k_k$ *and* $I_{00}(x,z)=W_{hk}$, We obtain

$$W_{hk}\sqrt{\sigma^2_{hkc}+\mu_{hk\eta}}$$

$$=\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\sqrt{\eta(t_1,t_2)}\,f(t_1,t_2)\,\partial t_1\partial t_2+\frac{(k_hk_k)^2}{96}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\,f(t_1,t_2)\,\partial t_1\partial t_2\left[1+O(k_hk_k)^2\right]$$

while taking summation over all strata, we have

$$\sum_h\sum_k W_{hk}\sqrt{\sigma^2_{hkc}+\mu_{hk\eta}}$$

$$=\int_a^b\int_c^d\sqrt{\eta(t_1,t_2)}\,f(t_1,t_2)\,\partial t_1\partial t_2+\sum_h\sum_k\frac{(k_hk_k)^2}{96}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\,f(t_1,t_2)\,\partial t_1\partial t_2\left[1+O(k_hk_k)^2\right]$$

Since it is obvious that $\int_a^b\int_c^d\sqrt{\eta(t_1,t_2)}\,f(t_1,t_2)\,\partial t_1\partial t_2$ is a constant because of interval defined for the variables. So, minimization of $\sum_h\sum_k W_{hk}\sqrt{\sigma^2_{hkc}+\mu_{hk\eta}}$ is equivalent to minimization of $d_1=\sum_h\sum_k\frac{(k_hk_k)^2}{96}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\,f(t_1,t_2)\,\partial t_1\partial t_2\left[1+O(k_hk_k)^2\right]$ as the first term is constant in the above equation.

Thus, we find that if the function $m_1(x,z)\,f(x,z)$ belongs to the class Ω then the minimum value of $\sum_h\sum_k W_{hk}\sqrt{\sigma^2_{hkc}+\mu_{hk\eta}}$ and therefore $V\left(\bar{y}_{st}\right)_{opt}$ exists and the set of strata boundaries $[x_h,z_k]$, corresponding to the minimum, are the solutions of the system of equation (4.5.2.7) or equivalently of (4.7.11). These conditions in that capacity are extremely hard to understand and accordingly it is required to discover some approach to conquer this trouble. It is done by replacing this system of equations by other system of equations which are comparatively easier to solve but are only asymptotically equivalent to the exact minimal equations. The error factor is introduced because we neglect the terms of higher powers of strata widths which is of course justifiable if the number of strata is large. These option arrangement of conditions were given in (4.7.14) and (4.7.15) We have obtained these systems of equations after neglecting the terms of order $O\left(Sup\left(k_h k_k\right)\right)^4 = O\left(m^4\right)$ where $m = \left(\underset{((a,b),(c,d))}{Sup}\left(k_h k_k\right)\right)$ on both sides of the equation (4.7.11).If the number of

strata is large and therefore terms of order $O\left(m^4\right)$ are quite small, the error involved in the approximate systems of equations is expected to be quite small and the set of stratification points $\left[x_h, z_k\right]$ obtained from them shall be quite near to the optimum values.

Now we proceed to develop the approximate system of equations given in (4.5.33) and (4.5.34).Here, in finding various forms of the function $Q_1\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$,we shall keep in mind that the function $Q_1\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$ is such that $\left(k_h k_k\right)^2 \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} m_1\left(t_1, t_2\right) f\left(t_1, t_2\right) \partial t_1 \partial t_2 = Q_1\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$

4.8 Approximate system of equations

I. If in (4.7.9) and (4.7.10) we retain only the first term on both sides of the equation and neglect the others, the two sided are equalized if

x_h = constant = $\dfrac{b-a}{L}$, h=1,2,…,L and z_k =constant $= \dfrac{d-c}{M}$, $\quad k = 1, 2, ..., M$ (4.8.1)

Therefore ,

$$x_h = a + \left(\frac{(b-a)}{L}\right)h \quad with \quad x_0 = a \quad and \quad x_L = b$$

and $z_k = c + \left(\dfrac{(d-c)}{M}\right)k \quad with \quad z_0 = c \quad and \quad z_M = d$

This set of solutions can't be expected to yield very good results as we have neglected terms of order $O\left(m^3\right)$ on both sides of the exact minimal equations. This solution holds for all $m_1(x,z) f(x,z)$ provided of course they belong to Ω and all given density functions with finite range. Due to its universality of approximation it can be recommended in cases where not much is known about $m_1(x,z)$ *and* $f(x,z)$.

However, apart from this it gives the strata boundaries at once without any difficulty that may arise even in solving the approximate systems of equations. This approximate method fails if the range of X and Z is infinity, but in such cases one can resort to truncation of the density function $f(x,z), f(x)$ *and* $f(z)$ to any suitable probability level before using this approximation.

II. From Lemma 4.4, we have

$$\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1 \partial t_2$$
$$= \sqrt{m_1(x_{h-1})m_1(x_h)m_1(z_{k-1})m_1(z_k)}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} f(t_1,t_2)\,\partial t_1\partial t_2\left[1+O(k_h k_k)^2\right]$$

and therefore the next system of equation to obtain approximately optimum points of stratification is

$$(k_h k_k)^2\sqrt{m_1(x_{h-1})m_1(x_h)m_1(z_{k-1})m_1(z_k)}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} f(t_1,t_2)\,\partial t_1\partial t_2 = c_1 \quad (4.8.2)$$

h=1,2,...,L and k=1,2,...,M

III. We obtain next approximate system of equations from (4.7.14).From this relation, approximately optimum points of stratification are such that

$$(k_h k_k)^2\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\, f(t_1,t_2)\,\partial t_1\partial t_2 = C_2, h=1,2,...,L \text{ and } k=1,2,...,M \quad (4.8.3)$$

The solutions obtained from this approximation are expected to be quite close to the optimum points as only terms of $O\left(m^4\right)$ have been neglected. All the approximate systems that will now follow also give the points of stratification to the same degree of accuracy.

From Lemma 4.6 and equation (4.8.2) the following class of approximate equations are obtained. The approximation to optimum points $[x_h, z_k]$ are obtained from

$$\left[(k_h k_k)^{3\lambda-1}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\left(m_1(t_1,t_2)\, f(t_1,t_2)\right)^{\lambda}\partial t_1\partial t_2\right]^{\frac{1}{\lambda}} = \text{Constant}$$

h=1,2,...,L and k=1,2,...,M

For $\lambda=1$,we obtain the equation (4.8.3).However for $\lambda=\frac{1}{2}$,we have

$$\left[\sqrt{(k_h k_k)}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\sqrt{m_1(t_1,t_2)\, f(t_1,t_2)}\,\partial t_1\partial t_2\right]^2 = C_3 \quad (4.8.4)$$

For $\lambda=\frac{1}{3}$, we have system of equations as

$$\left[\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\sqrt[3]{m_1(t_1,t_2)\,f(t_1,t_2)}\,\partial t_1\partial t_2\right]^3 = C_4 \tag{4.8.5}$$

In all these system of equations, the $C_i^{'}s$ (i=1, 2, 3) are the constants to be determined.

4.9 Determination of constants

In all the approximations (4.8.2) (4.8.3),(4.8.4) and (4.8.5)the constants $C_i^{'}s$ (i=1, 2, 3) have to be determined so as to enable us to find set of points $[x_h, z_k]$.Since it is difficult to find exact values of these constants, we shall, therefore obtain the approximate values of $C_i^{'}s$.

It was noticed in relation (4.8.1) ,that k_h *and* k_k are constants gives asymptotically correct approximation to the minimal equations when all terms except first are neglected so that for large L and M, we have

$$k_h = \frac{b-a}{L} \quad and \quad \sigma_{hx}^2 = \frac{1}{12}\frac{(b-a)^2}{L^2} \;, z_k = \frac{d-c}{M} \quad and \quad \sigma_{kz}^2 = \frac{1}{12}\frac{(d-c)^2}{M^2}$$

Using these approximations values of the constants $C_1,\ C_2$ *and* C_3 denoted by $C_1^{'},\ C_2^{'}$ *and* $C_3^{'}$ respectively can be obtained as follows.

I. $C_1^{'} = \dfrac{(b-a)^2}{L^2}\dfrac{(d-c)^2}{M^2}\displaystyle\int_a^b\int_c^d m_1(t_1,t_2)\,f(t_1,t_2)\,\partial t_1\partial t_2$

Taking summation over all strata, we get

$$\frac{(b-a)^2}{L^2}\frac{(d-c)^2}{M^2}\sum_{h=1}^{L}\sum_{k=1}^{M}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} m_1(t_1,t_2)\,f(t_1,t_2)\,\partial t_1\partial t_2 = LMC_1^{'}$$

II. From (4.8.4), we get

$$\frac{(b-a)}{L}\frac{(d-c)}{M}\left[\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\sqrt{m_1(t_1,t_2)\,f(t_1,t_2)}\,\partial t_1\partial t_2\right]^2 = C_2^{'}$$

$$\sum_{h=1}^{L}\sum_{k=1}^{M}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\sqrt{m_1(t_1,t_2)\,f(t_1,t_2)}\,\partial t_1\partial t_2$$

or

$$= \int_a^b\int_c^d\sqrt{m_1(t_1,t_2)\,f(t_1,t_2)}\,\partial t_1\partial t_2 = \left(\frac{(LM)^3 C_2^{'}}{(b-a)(d-c)}\right)$$

which on summing over all the strata gives

$$\frac{(b-a)(d-c)}{(LM)^3}\left[\int_a^b\int_c^d\sqrt{m_1(t_1,t_2)\,f(t_1,t_2)}\partial t_1\partial t_2\right]^2=C_2^{'} \tag{4.9.1}$$

III. From (4.8.5) ,we get

$$\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\sqrt{m_1(t_1,t_2)\,f(t_1,t_2)}\partial t_1\partial t_2=C_3^{1/3}$$

on summing over all the strata, we obtain

$$C_3=\frac{1}{(LM)^3}\left[\int_a^b\int_c^d\sqrt{m_1(t_1,t_2)\,f(t_1,t_2)}\partial t_1\partial t_2\right]^3 \tag{4.9.2}$$

This is interesting to note that the value of constant C_3 is exact, since no approximation based on the asymptotic equivalence of k_h *and* k_k has been used. Hence, the approximate system of equations (4.5.8) is more useful from practical point of view. Now we shall give a rule to find out the AOSB in case two auxiliary variables.

4.10 Cum $\sqrt[3]{D_1(x,z)}$ Rule

If the function

$D_1(x,z)=m_1(x,z)\,f(x,z)$

where

$$m_1(x,z)=\frac{\eta'^2(x,z)+4\eta(x,z)c'^2(x,z)}{\left[\eta(x,z)\right]^{\frac{3}{2}}}$$

is bounded and its first derivative exists for all x in $[a,b]$ and z in $[c,d]$,then for a given value of L and M taking equal intervals on the cumulative cube root of $D_1(x,z)$ will give AOSB $[x_h,z_k]$.

Remarks

1. Let $c(x,z)=\alpha+\beta x+\gamma z$ then $c'(x,z)=\beta$ *and* $c'(x,z)=\gamma$, by differentiating w.r.t. x *and* z, keeping z and x constant , respectively, and ultimately

$$m_1(x,z)=\frac{\eta'^2(x,z)+4\eta(x,z)c'^2(x,z)}{\left[\eta(x,z)\right]^{\frac{3}{2}}}$$ =constant .Therefore, for such a case the

proposed rule reduces to the $Cum\sqrt[3]{f(x,z)}$.

2. Let $c(x,z)=\alpha+\beta x+\gamma z,$ then if either $c(x,z)=\alpha+\beta x$ or $c(x,z)=\alpha+\gamma z$, then

$$m_1(x)=\frac{\eta'^2(x)+4\eta(x)c'^2(x)}{[\eta(x)]^{\frac{3}{2}}} \quad or \quad m_1(z)=\frac{\eta'^2(z)+4\eta(z)c'^2(z)}{[\eta(z)]^{\frac{3}{2}}}$$

reduces to the method proposed by Singh and Sukhatme (1969) for single auxiliary variable.

3. For any distribution and given number of strata, the AOSB $[x_h, z_k]$ will remain unchanged with respect to the form of conditional variance $\eta(x,z)$,however, the efficiency of stratification as compared to no stratification will be changed with the choice of various forms of conditional variance.

4.11 Empirical study

We shall now demonstrate empirically the effectiveness of the proposed method of findings the set of AOSB. For the sake of simplicity, the linear regression line of Y on X and Z have been taken of the form $y=\alpha+\beta x+\gamma z+e$. For the conditional variance function $\eta(x,z)$ we have taken two forms viz. $\eta(x,z)=\alpha$ and $\eta(x,z)=\lambda xz$, where α *and* λ are constants.

The origin is deliberately excluded from the range of the auxiliary variables X and Z , otherwise $\eta(x,z)=\lambda xz$ which will give $m_1(x,z)=\infty$ at x = 0, z = 0 and the function $m_1(x,z)\,f(x,z)$ in that case does not belong to the class Ω of functions. We could have also avoided this difficulty by taking some other suitable forms to the functions. For the empirical studies under optimum allocation let us assume values of $\alpha=0.0214, \lambda=0.00437$ which are quite small so that the effect of taking $\eta(x,z)=\alpha$ and $\eta(x,z)=\lambda xz$ is negligibly small.

In order to obtain AOSB,let us assume that the correlation coefficient, ρ, between X and Z is equal to 0.65 For this purpose the various forms of density functions of the stratification variables X and Z have been considered as follows:

4.11.1: Let us suppose that the auxiliary variable X follows standard normal distribution with pdf as

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \qquad ,x\geq 0$$

and the variable Z also follows standard nor mal distribution with pdf as

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \qquad ,z\geq 0$$

In order to obtain the OSB when both the variables are standard normally distributed let us assume the value of regression coefficients $\beta=0.65$ and $\gamma=0.57$.For obtaining total 16 strata, 4 along the X variable and 4 along the Z variable, using the proposed Cum $\sqrt[3]{D_1(x,z)}$ rule the distributions of X and Z are truncated at x=6 and z=4, respectively. By solving it in Mathematica Software we get the stratification points as below:

Table 4.11.1: OSB when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\alpha$

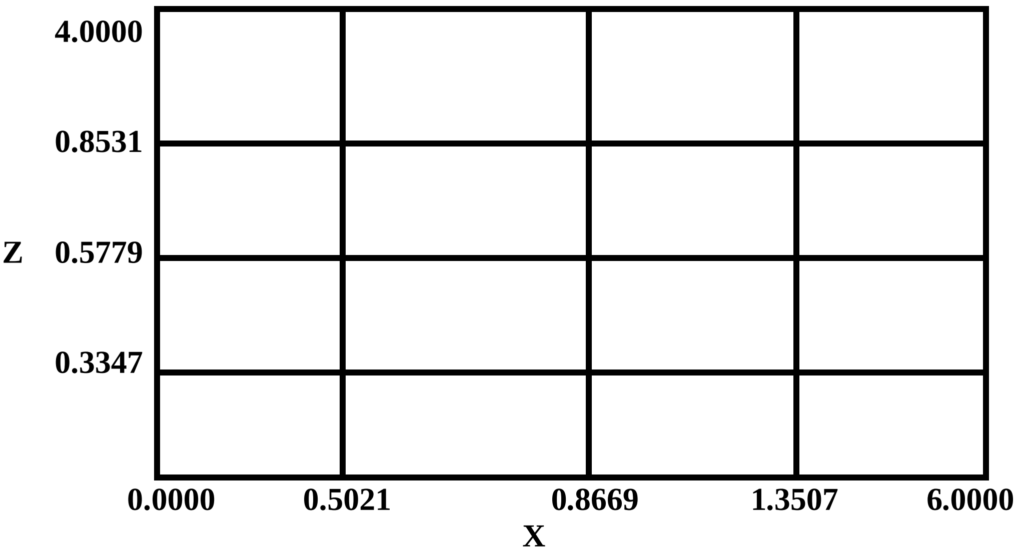


Table 4.11.2: OSB and Variance when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\alpha$

OSB (x_h, z_k)	**Variance** Cum $\sqrt[3]{D_1(x,z)}$ Rule	**Variance (Singh and Sukhatme, 1969)**	**% R.E.**
(0.5021,0.3347)	0.16254785	0.396418	243.89
(0.8669,0.3347)			
(1.3507,0.3347)			
(6.0000,0.3347)			
(0.5021,0.5779)			
(0.8669,0.5779)			
(1.3507,0.5779)			
(6.0000,0.5779)			
(0.5021,0.8531)			
(0.8669,0.8531)			
(1.3507,0.8531)			
(6.0000,0.8531)			
(0.5021,4.0000)			
(0.8669,4.0000)			
(1.3507,4.0000)			
(6.0000,4.0000)			

Table 4.11.3: OSB when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\lambda xz$

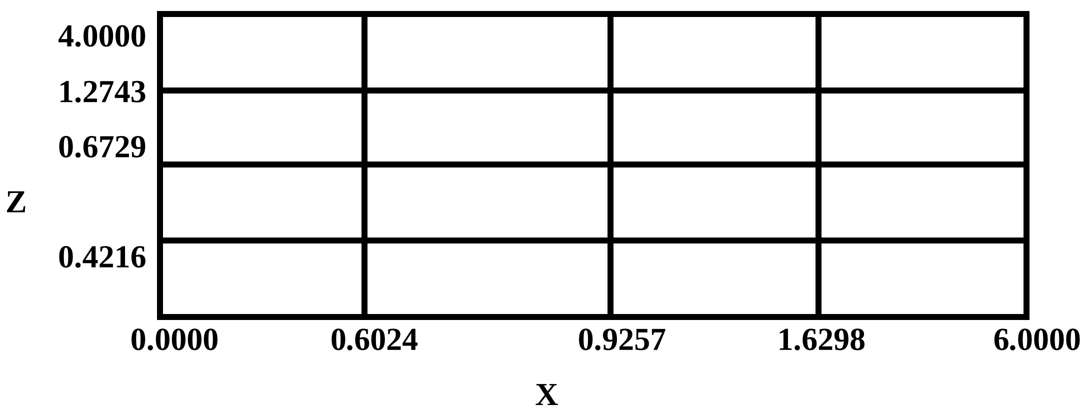


Table 4.11.4: OSB and Variance when the auxiliary variables X and Z follow standard normal distribution and are dependent with $\eta(x,z)=\lambda xz$

OSB (x_h, z_k)	**Variance** (Cum $\sqrt[3]{D_1(x,z)}$ Rule)	**Variance (Singh and Sukhatme. 1969)**	**% R.E.**
(0.6024,0.4216) (0.9257,0.4216) (1.6298,0.4216) (6.0000,0.4216) (0.6024,0.6729) (0.9257,0.6729) (1.6298,0.6729) (6.0000,0.6729) (0.6024,1.2743) (0.9257,1.2743) (1.6298,1.2743) (6.0000,1.2743) (0.6024,4.0000) (0.9257,4.0000) (1.6298,4.0000) (6.0000,4.0000	0.15942873	0.428619	268.85

4.11.2 : Let us consider the distribution of X as uniform having probability density function (pdf) as

$$f(x)=\frac{1}{b-a}, a \le x \le b$$

and Z follows exponential distribution with pdf as

$$f(z)=e^{-z+1}, z \ge 0$$

In order to obtain OSB for the above pdf let us suppose that the variable X is defined in [1,2] and Z in [1,6] and assume that values of β *and* γ be 0.56 and 0.72 , respectively. Using the above pdf's for constructing stratification points using

Cum $\sqrt[3]{D_1(x,z)}$ Rule for total 6 strata i,e 2 strata along X variable and 3 along Z variable, using Mathematica software for solving the function , we get values as presented in the following table.

Table 4.11.5: OSB and Variance when the auxiliary variables are dependent and follow uniform and exponential distributions, respectively, having $\eta(x,z)=\alpha$

OSB (x_h, z_k)	**Variance** Cum $\sqrt[3]{D_1(x,z)}$ Rule	**Variance Singh and Sukhatme (1969)**	**% R.E.**
(1.46099,1.77003) (2.00000,1.77003) (1.46099,4.07294) (2.00000,4.07294) (1.46099,6.00000) (2.00000,6.00000)	0.096473	0.1553	160.98

Table 4.11.6: OSB and Variance when the auxiliary variables are dependent and follow uniform and exponential distributions, respectively, having $\eta(x,z)=\lambda xz$

OSB (x_h, z_k)	**Variance** (Cum $\sqrt[3]{D_1(x,z)}$ Rule)	**Variance (Singh and Sukhatme, 1969)**	**% R.E**
(1.5000,2.7856) (2.0000,2.7856) (1.5000,3.17044) (2.0000,3.17044) (1.5000,6.0000) (2.0000,6.0000)	0.0627351	0.173	275.76

4.12 : For independent auxiliary variables (ρ = 0)

Since in many real life problems we come across the problem of stratification having two independent auxiliary variables. So in order to tackle the problem a method is to be developed that could solve the problem. Following the same procedure as followed for dependent variables, we can write (4.7.11) as

$$\left(\frac{f(t_1)f(t_2)\eta'^2(t_1)\eta'^2(t_2)+4f(t_1)f(t_2)\eta(t_1)\eta(t_2)c'^2(t_1)c'^2(t_2)}{\left[\eta(t_1)\eta(t_2)\right]^{\frac{3}{2}}}\right)$$

$$\left[(k_hk_k)^2\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}m_1(t_1)m_1(t_2)f(t_1)f(t_2)\partial t_1\partial t_2\left[1+O(k_hk_k)^2\right]\right]^{\frac{2}{3}}$$

$$=\left(\frac{f(t_1)f(t_2)\eta'^2(t_1)\eta'^2(t_2)+4f(t_1)f(t_2)\eta(t_1)\eta(t_2)c'^2(t_1)c'^2(t_2)}{\left[\eta(t_1)\eta(t_2)\right]^{\frac{3}{2}}}\right)$$

$$\left[\left(k_ik_j\right)^2\int_{x_h}^{x_{h+1}}\int_{z_k}^{z_{k+1}}m_1(t_1)m_1(t_2)f(t_1)f(t_2)\partial t_1\partial t_2\left[1+O\left(k_ik_j\right)^2\right]\right]^{\frac{2}{3}}$$

where

$$m_1(t_1)=\frac{\eta'^2(t_1)+4\eta(t_1)c'^2(t_1)}{\left(\eta(t_1)\right)^{\frac{3}{2}}} \quad \text{and} \quad m_1(t_2)=\frac{\eta'^2(t_2)+4\eta(t_2)c'^2(t_2)}{\left(\eta(t_2)\right)^{\frac{3}{2}}}$$

i=h+1,h=1,2,…,L ; j=k+1,k=1,2,…,M and we assume that the functions $f(x)m_1(x)\in\Omega\forall x\in[a,b]$ and $f(z)m_1(z)\in\Omega\forall z\in[c,d]$ and $\eta(t_1)$ *and* $\eta(t_2)$ represent conditional variance of x and z as $V(e/x)$ and $V(e/z)$, respectively.

Proceeding in the same way as the case of dependent auxiliary variables, the theorem 2.1 would take another form, the approximate system of equations and the determination of constants can be obtained. The only difference is that in latter case we have joint density function while as in current case we have marginal density functions. Thus, keeping in view the rule is given as follows.

4.13 Cum $\sqrt[3]{D_2(x,z)}$ Rule

If the function $D_2(x,z)=D_2(x)D_2(z)$

$$D_2(x)=m_2(x)f(x) \quad and \quad D_2(z)=m_2(z)f(z)$$

where $m_2(x)=\dfrac{\eta'^2(x)+4\eta(x)c'^2(x)}{(\eta(x))^{\frac{3}{2}}}$ and $m_2(z)=\dfrac{\eta'^2(z)+4\eta(z)c'^2(z)}{(\eta(z))^{\frac{3}{2}}}$

is bounded and its first two derivatives exists and defined in $x\in[a,b]$ and $z\in[c,d]$,then for a the given value of total strata along X variable and along Y variable taking equal intervals in the cumulative cube root of $D_1(x)$ and $D_2(z)$, that is

$$\int_{x_{h-1}}^{x_h}\sqrt[3]{D_2(x)}dx=\frac{1}{L}\int_a^b\sqrt[3]{D_2(x)}dx \quad and \int_{x_{h-1}}^{x_h}\sqrt[3]{D_2(z)}dz=\frac{1}{M}\int_c^d\sqrt[3]{D_2(z)}dz$$

will provide the AOSB as $[x_h,z_k]$.

Remarks:

$m_1(x)=\dfrac{\eta'^2(x)+4\eta(x)c'^2(x)}{(\eta(x))^{\frac{3}{2}}}$ and $m_1(z)=\dfrac{\eta'^2(z)+4\eta(z)c'^2(z)}{(\eta(z))^{\frac{3}{2}}}$ are constants if we take $c(x,z)=\alpha+\beta x \quad or \quad c(x,z)=\alpha+\gamma z$ and differentiating partially it w.r.t. x and z ,then it would give constant value. Then for the case proposed technique reduces to Cum $\sqrt[3]{f(x)f(z)}$.

4.14 Empirical study

In order to have empirical study for the above proposed rule, we have to choose the individual distribution for x and z. For the sake of simplicity, the linear regression line Y on X and Z have been taken as, of the form $y=\alpha+\beta x+\gamma z+e$ .For the conditional variance function $\eta(x,z)$ we have taken two forms viz.

1. $\eta(x)=\alpha_1,\eta(z)=\alpha_2 \quad where \quad \eta(x,z)=\eta(x)\eta(z)$

2. $\eta(x)=\lambda_1 x,\eta(z)=\lambda_2 z \quad where \quad \eta(x,z)=\eta(x)\eta(z)$

where $\alpha_1,\alpha_2,\lambda_1,\lambda_2$ are constants

The origin is deliberately excluded from the range of the auxiliary variables X and Z, otherwise $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$ we have $m_1(x)=\infty$ at x = 0 and $m_1(x)=\infty$ at z = 0 and the functions $m_1(x)f(x)$ *and* $m_2(z)f(z)$ in that case do not belong to the class Ω of functions. We could have also avoided this difficulty by taking some other suitable forms to the functions. For the empirical studies under optimum allocation let us assume values of $\alpha_1 = 0.0432, \alpha_2 = 0.0334, \lambda_1 = 0.00432$, $\lambda_2 = 0.00331$ which are quite small so that the effect of taking $\eta(x) = \alpha_1, \eta(z) = \alpha_2$ and $\eta(x) = \lambda_1 x, \eta(z) = \lambda_2 z$ is negligibly small.

4.14.1: X follows uniform distribution with pdf as

$$f(x) = \frac{1}{b-a}, a \le x \le b$$

and Z follows exponential distribution with pdf as

$$f(z) = e^{-z+1}, z \ge 1$$

In order to obtain OSB for the above case let us suppose that the variable X is defined in [1,2] and Z in [1,6] and that values of β *and* γ be 0.56 and 0.72, respectively. Using the above pdf's for constructing strata points using Cum $\sqrt[3]{D_2(x,z}$ rule for total 6 strata, using Mathematica software for solving the function , we get the values which has been presented in the following table.

Table 4.14.1: OSB when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, and with $\eta(x) = \alpha_1$ *and* $\eta(z) = \alpha_2$

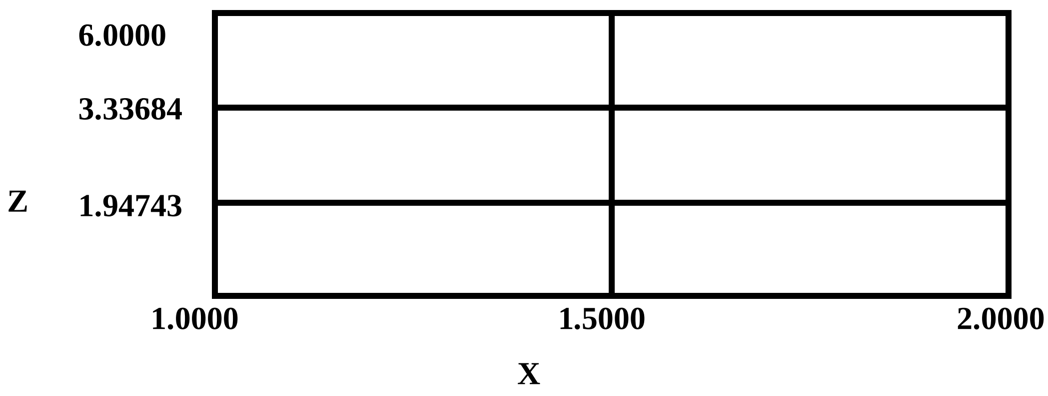


Table 4.14.2: OSB and Variance when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, and with $\eta(x)=\alpha_1 \quad and \quad \eta(z)=\alpha_2$

OSB (x_h, z_k)	**Variance** Cum $\sqrt[3]{D_2(x,z)}$ Rule	**Variance (Singh and Sukhatme 1969)**	**% R.E.**
(1.5000,1.9474) (2.0000,1.9474) (1.5000,3.33684) (2.0000,3.33684) (1.5000,6.0000) (2.0000,6.0000)	0.0815364	0.1553	190.468

Table 4.14.3: OSB when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, with $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$

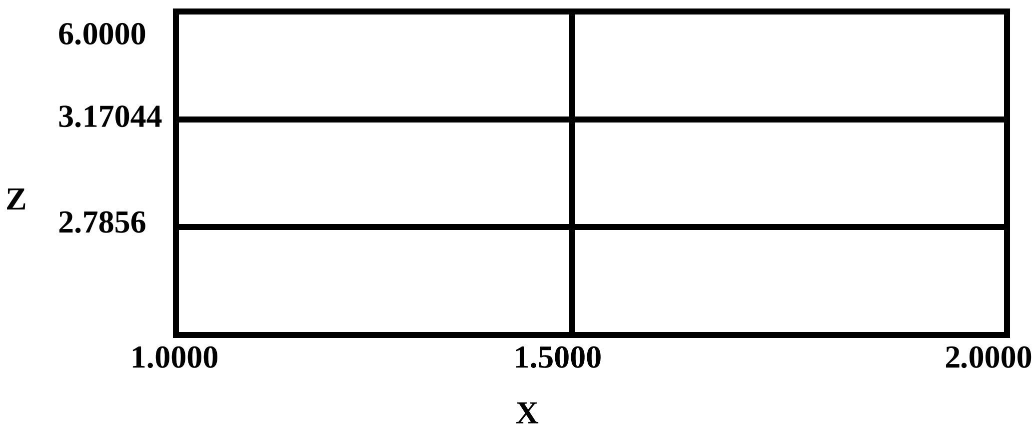


Table 4.14.4: OSB and Variance when the auxiliary variables X and Z are independent following uniform and exponential distributions, respectively, and with $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$

OSB (x_h, z_k)	**Variance** (Cum $\sqrt[3]{D_2(x,z)}$ Rule)	**Variance (Singh and Sukhatme, 1969)**	**% R.E**
(1.5000,2.7856) (2.0000,2.7856) (1.5000,3.17044) (2.0000,3.17044) (1.5000,6.0000) (2.0000,6.0000)	0.05326	0.173	324.82

4.14.2: X has right triangular distribution with pdf as

$$f(x)=2(2-x),1\le x\le 2$$

and Z follows uniform distribution with pdf as

$$f(z)=\frac{1}{b-a},1\le z\le 2$$

The OSB can be obtained when the auxiliary variable X is having right triangular distribution defined in [1,2] and Z is having uniform distribution defined in [1,2].In order to obtain stratification points for total 6 number of strata, 3 along the X variable and 2 along the Z variable by assuming the values of $\alpha=0.57$ and $\beta=0.42$ .After using the method of Cum $\sqrt[3]{D_2(x,z)}$ for obtaining stratification points, using Mathematica Software for solving the function, the results obtained are presented below:

Table 4.14.5: OSB and Variance when the auxiliary variables are independent following right triangular and uniform distributions, respectively, with $\eta(x)=\alpha_1 \quad and \quad \eta(z)=\alpha_2$

OSB (x_h, z_k)	**Variance (**Cum $\sqrt[3]{D_2(x,z)}$ Rule)
(1.2598,1.5000) (1.6257,1.5000) (2.0000,1.5000) (1.2598,2.0000) (1.6257,2.0000) (2.0000,2.0000)	0.03985

Table 4.14.6: OSB and Variance when the auxiliary variables are independent following right triangular and uniform distributions, respectively, with $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$

OSB (x_h, z_k)	**Variance** (Cum $\sqrt[3]{D_2(x,z)}$ Rule)
(1.2530,1.5000)	0.074621
(1.5523,1.5000)	
(2.0000,1.5000)	
(1.2530,1.5000)	
(1.5523,1.5000)	
(2.0000,1.5000)	

4.15 Case of Proportional allocation

In order to discuss the solution of minimal equations it is necessary to prove the following Lemma

Lemma 4.8: With $I_{00}(x,z)$ and $\sigma_\eta^2(x,z)$ defined as in Lemma 4.1 and Lemma 4.3 respectively, we have

$$I_{00}(x,z)\sigma_\eta^2(x,z)=\frac{k^2}{12}\int_{z_1}^{z_2}\int_{x_1}^{x_2}\eta'^2(t_1,t_2)f(t_1,t_2)\partial t_1\partial t_2 \tag{4.15.1}$$

where k denotes any k_h or k_k.

Proof: We have from the Lemma 4.1 and Lemma 4.3 the value of $I_{00}(x,z) \quad and \quad \sigma_\eta^2(x,z)$ respectively. Thus we have after multiplying them as

$$I_{00}(x,z)\sigma_\eta^2(x,z)=\frac{(k)^3 f\eta'^2}{12}\left[1+\left(\frac{(f_x+f_z)\eta'+2f\eta''}{2f\eta'}\right)k+O(k)^2\right]$$

$$=\frac{(k)^2}{12}\left[f\eta'^2(k)+\left((f_x+f_z)\eta'^2+2f\eta'\eta''\right)\frac{(k)^2}{2}+O(k)^3\right]$$

$$\therefore\quad I_{00}(x,z)\sigma_\eta^2(x,z)=\frac{(k)^2}{12}\left[\int_{z_1}^{z_2}\int_{x_1}^{x_2}\eta'^2(t_1,t_2)f(t_1,t_2)\partial t_1\partial t_2+O(k)^3\right]$$

Where k represents either $k_h \quad or \quad k_k$ and the functions f,η and their derivatives are evaluated at both $t_1=x \quad and \quad t_2=z$.

4.16 Minimal equations and their approximate solutions

After discussing the case of optimal allocation we come to consider the case of proportional allocation. In this section we shall find the series expansions of the system of equations given in (4.5.1.4) about the points x_h and z_k the common boundary of $(h,k)^{th}$ *and* $(h+1,k+1)^{th}$ strata and obtain the approximate systems of equation which give approximately optimum points of stratification as their solutions. In doing so we shall make use of some of the Lemma already proved in Article (4.6).

The minimal equations for this method are given by

$$c(x_h,z_k)-\mu_{hkc}=\mu_{ijc}-c(x_h,z_k)$$

$$i=h+1, h=1,2,...,L \quad and \; j=k+1, k=1,2,...,M$$

Let us first consider the development of right hand side. The corresponding expansion for the left hand side may be obtained by changing the signs of coefficients of even powers of (k_i,k_j) the width of $(i,j)^{th}$ stratum where $k_i=x_{h+1}-x_h$, $k_j=z_{k+1}-z_k$ From (4.6.4), after replacing (x_1,x_2) *by* (x_h,x_{h+1}) and (z_1,z_2) *by* (z_k,z_{k+1}), we have

$$\mu_{ikc}=c\left[1+\frac{c'}{2c}k_i+\left(\frac{c'f_x+2fc''}{12fc}\right)k_i^2+\left(\frac{ff_{xx}c'+ff_xc''+f^2c'''}{24f^2c}\right)k_i^3+O\left(k_i^4\right)\right]$$

where $k_i=x_{h+1}-x_h$ and the functions c, f and their derivatives are evaluated at x_h.

However when the same functions are differentiated w.r.t. z_k, we have

$$\mu_{hjc}=c\left[1+\frac{c'}{2c}k_j+\left(\frac{c'f_z+2fc''}{12fc}\right)k_j^2+\left(\frac{ff_{zz}c'+ff_zc''+f^2c'''}{24f^2c}\right)k_j^3+O\left(k_j^4\right)\right]$$

where $k_j=z_{k+1}-z_k$

When the functions c, f and their derivatives are evaluated at both x_h *and* z_k, we get

$$\mu_{ijc}=c\left[1+\frac{c'}{2c}\left(k_ik_j\right)+\left(\frac{c'\left(f_x+f_z\right)+2fc''}{12fc}\right)\left(k_ik_j\right)^2+\left(f_4\right)\left(k_ik_j\right)^3+O\left(k_ik_j\right)^4\right]$$

where $f_4=\dfrac{f\left(f_{xx}+f_{zz}+f_{xz}\right)c'+f\left(f_x+f_z\right)c''+f^2c'''}{24f^2c}$

Thus we have

$$\mu_{ijc} - c(x_h, z_k) = \frac{c^{'}}{2}(k_i k_j) + \left(\frac{c^{'}(f_x + f_z) + 2fc^{''}}{12f}\right)(k_i k_j)^2 + (f_4)(k_i k_j)^3 + O(k_i k_j)^4$$

$$= \frac{(k_i k_j)}{2}\left[c^{'} + \left(\frac{c^{'}(f_x + f_z) + 2fc^{''}}{12f}\right)(k_i k_j) + O(k_i k_j)^2\right]$$

Similarly, we get

$$\mu_{hkc} - c(x_h, z_k) = \frac{(k_h k_k)}{2}\left[c^{'} + \left(\frac{c^{'}(f_x + f_z) + 2fc^{''}}{12f}\right)(k_h k_k) + O(k_h k_k)^2\right]$$

where also in this case the functions f, η and their derivatives are evaluated at x_h *and* z_k. Therefore equation (4.5.1.4) can be put as

$$\frac{(k_h k_k)}{2}\left[c^{'} + \left(\frac{c^{'}(f_x + f_z) + 2fc^{''}}{12f}\right)(k_h k_k) + O(k_h k_k)^2\right]$$

$$= \frac{(k_i k_j)}{2}\left[c^{'} + \left(\frac{c^{'}(f_x + f_z) + 2fc^{''}}{12f}\right)(k_i k_j) + O(k_i k_j)^2\right] \qquad (4.16.1)$$

Now let us consider an expansion of the function

$$B_{hk} = \int_{z_{k-1}}^{z_k}\int_{x_{h-1}}^{x_h} c^{'2}(t_1, t_2) f(t_1, t_2)\, \partial t_1 \partial t_2$$

about the point $[x_h, z_k]$. Expanding the integral about x_h *and* z_k with the help of Taylors expansion for two variables, we get

$$B_{hk} = c^{'2} f k_h k_k \left[1 - \frac{(f_x + f_z)c^{'} + 2fc^{''}}{fc^{'}}\frac{(k_h k_k)}{2} + O(k_h k_k)^2\right] \qquad (4.16.2)$$

where in (4.16.2) also the function of f, η and their derivatives are evaluated at x_h *and* z_k. Thus we find that

$$\frac{(k_h k_k)^2}{8}\frac{B_{hk} c^{'}}{f} = \frac{(k_h k_k)^3 c^{'3}}{8}\left[1 - \left(\frac{(f_x + f_z)c^{'} + 2fc^{''}}{2fc^{'}}\right)(k_h k_k) + O(k_h k_k)^2\right]$$

or

$$\left[\frac{(k_h k_k)^2}{8}\frac{B_{hk}c'}{f}\right]^{\frac{1}{3}} = \frac{(k_h k_k)c'}{2}\left[1-\left(\frac{(f_x+f_z)c'+2fc''}{2fc'}\right)(k_h k_k)+O(k_h k_k)^2\right]^{\frac{1}{3}}$$

$$= \frac{(k_h k_k)c'}{2}\left[1-\left(\frac{(f_x+f_z)c'+2fc''}{6fc'}\right)(k_h k_k)+O(k_h k_k)^2\right] \quad (4.16.3)$$

Similarly, we obtain

$$\left[\frac{(k_i k_j)^2}{8}\frac{B_{ij}c'}{f}\right]^{\frac{1}{3}} = \frac{(k_h k_k)c'}{2}\left[1-\left(\frac{(f_x+f_z)c'+2fc''}{6fc'}\right)(k_i k_j)+O(k_i k_j)^2\right]$$

Therefore the minimal equations (4.16.1) can be put as

$$\left[\frac{(k_h k_k)^2}{8}\frac{B_{hk}c'}{f}\right]^{\frac{1}{3}}\left[1+O(k_h k_k)^2\right] = \left[\frac{(k_i k_j)^2}{8}\frac{B_{ij}c'}{f}\right]^{\frac{1}{3}}\left[1+O(k_i k_j)^2\right] \quad (4.16.4)$$

Hence if the terms of order $O\left(\underset{((a,b),(c,d))}{Sup}(k_h k_j)\right)^3$ can be neglected, we can replace the minimal equations approximately by

$$\left[\frac{(k_h k_k)^2}{8}\frac{B_{hk}c'}{f}\right]^{\frac{1}{3}} = \left[\frac{(k_i k_j)^2}{8}\frac{B_{ij}c'}{f}\right]^{\frac{1}{3}}$$

or $\quad (k_h k_k)^2 B_{hk}$ = Constant (4.16.5)

In case it is possible to find a function $Q_1'(x_{h-1}, x_h, z_{k-1}, z_k)$ such that

$$(k_h k_k)^2 B_{hk} = (k_h k_k)^2 \int_{z_{k-1}}^{z_k}\int_{x_{h-1}}^{x_h} c'^2(t_1,t_2) f(t_1,t_2)\,\partial t_1 \partial t_2$$
$$= Q_1'(x_{h-1}, x_h, z_{k-1}, z_k)\left[1+O(k_h k_k)^2\right] \quad (4.16.6)$$

Thus the system of equations (4.16.5) to the same degree of accuracy can be put as

$Q_1'(x_{h-1}, x_h, z_{k-1}, z_k)$ = Constant (4.16.7)

the above results can be put in the form of theorem as follows.

Theorem 4.2: If the regression of the estimation variable Y in the stratification variables X and Z in the infinite super population is given by

$$Y = C(X,Z) + \varepsilon$$

where 'ϵ' is the error component such that $E(e\,|\,x,z)$ =0 and $V(e\,|\,x,z) = \eta(x,z) > 0$, $\forall x \in (a,b)$ *and* $z \in (c,d)$ with $(b-a) < \infty, (d-c) < \infty$, and further if the function $c'^{2}(x,z) f(x,z)$ belong to Ω, then the system of equations (4.5.1.4) giving strata boundaries $[x_h, z_k]$ which correspond to the minimum of $V(\bar{y}_{st})_{prop}$ can be put as

$$\left\{(k_h k_k)^2 \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} c'^{2}(t_1,t_2) f(t_1,t_2) \partial t_1 \partial t_2 \left[1 + O(k_h k_k)^2\right]\right\}^{\frac{1}{3}}$$

$$= \left\{(k_i k_j)^2 \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} c'^{2}(t_1,t_2) f(t_1,t_2) \partial t_1 \partial t_2 \left[1 + O(k_i k_j)^2\right]\right\}^{\frac{1}{3}}$$

if the terms of order $O\left(\underset{((a,b),(c,d))}{Sup}(k_h k_j)\right)^3$ can be neglected ,these equations can be replaced by the approximate system of equations

$$(k_h k_k)^2 \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} c'^{2}(t_1,t_2) f(t_1,t_2) \partial t_1 \partial t_2 \left[1 + O(k_h k_k)^2\right] = \text{Constant}$$

Or equivalently by

$$Q_1'(x_{h-1}, x_h, z_{k-1}, z_k) = \text{Constant}$$

where $k_h = x_h - x_{h-1}$ *and* $k_k = z_k - z_{k-1}$

Therefore,

$$\therefore \begin{aligned} &Q_1'(x_{h-1}, x_h, z_{k-1}, z_k)\left[1 + O(k_h k_k)^2\right] \\ &= (k_h k_k)^2 \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} c'^{2}(t_1,t_2) f(t_1,t_2) \partial t_1 \partial t_2 \left[1 + O(k_h k_k)^2\right] \end{aligned}$$

where $\begin{aligned} i &= h+1, h = 1,2,...,L \\ j &= k+1, k = 1,2,...,M \end{aligned}$

The same result can also be obtained by minimizing the function

$\sum_h \sum_k \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} c'^2 (t_1, t_2) f(t_1, t_2) \partial t_1 \partial t_2 \left[1 + O(k_h k_k)^2\right]$ as in the light of Lemma 4.8,

$12 \sum_h \sum_k W_{hk} \sigma_{hkc}^2$ equals to this function.

Thus we find that if the function $c'^2 (x,z) f(t_1, t_2)$ belongs to the class Ω the minimum value of $\sum_h \sum_k W_{hk} \sigma_{hkc}^2$ and therefore of $V\left(\bar{y}_{st}\right)_{prop}$ exists and set of strata boundaries $[x_h, z_k]$,corresponding to this minimum are the solutions of the systems of equations (4.5.1.4) or equivalently of (4.16.4).As in case of optimum allocation method, these equations are very difficult to solve exactly and it becomes essential to find some approximation to optimum points of stratification. It can be done by replacing the exact minimal equations by other systems of equations which are comparatively easy to solve but are only asymptotically equivalent to the exact equations. The error is introduced because we neglect terms of higher powers of the strata widths which can be justified when the number of strata is large. The approximate systems of equations are obtained by neglecting terms of order $O\left(m^3\right)$ where $m = \underset{((a,b),(c,d))}{Sup} (k_h k_k)$,on both sides of (4.16.4) .If the number of strata is large, the terms of order $O\left(m^3\right)$ are small and therefore the error involved in the approximate systems of equations is small. Although this error is comparatively large than the one involved in case of optimum allocation. Here we shall develop the approximate systems of equations given in (4.16.5) and (4.16.7).

4.17 Approximate systems of equations

I. If in the expansion of the minimal equations (4.5.1.4) we neglect all terms except the first on both sides of the equation, the solution is obtained by taking

x_h = constant $= \dfrac{b-a}{L}$, h=1,2,…,L and z_k =constant $= \dfrac{d-c}{M}$, $\quad k = 1, 2, ..., M$ (4.17.1)

Therefore

$$x_h = a + \left(\frac{(b-a)}{L}\right) h \quad with \quad x_0 = a \quad and \quad x_L = b$$

and $z_k = c + \left(\dfrac{(d-c)}{M}\right) k \quad with \quad z_0 = c \quad and \quad z_M = d$

As in case of optimum allocation method, this set of approximations can not be expected to yield very good results although it is easiest to obtain. However the method is not applicable in case of infinite range.

II. An approximation to the optimum points of stratification is obtained by solving the systems of equations

$$\begin{gathered}(k_h k_k)^2 \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} c'^2 (t_1,t_2)\, f(t_1,t_2)\, \partial t_1 \partial t_2 = C_1 \\ h = 1,2,\ldots,L \text{ and } k = 1,2,\ldots,M\end{gathered} \tag{4.17.2}$$

as shown in (4.16.5).The solutions of this system of equations and also of those that will now follow ,are expected to be closer to the optimum points of stratification as compared to the solutions obtained from (4.17.1).

III. From Lemma 4.6 and equation (4.17.2),we get a general class of approximate systems of equations as

$$\left[(k_h k_k)^{3\lambda-1} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \left(c'^2 (t_1,t_2)\, f(t_1,t_2)\right)^{\lambda} \partial t_1 \partial t_2\right]^{\frac{1}{\lambda}} = \text{Constant}$$

h=1,2,...,L and k=1,2,...,M

However, for $\lambda = \frac{1}{2}$,we have

$$\left[(k_h k_k) \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} c'^2 (t_1,t_2) \sqrt{f(t_1,t_2)}\, \partial t_1 \partial t_2\right]^2 = C_2$$

and for $\lambda = \frac{1}{3}$, we have system of equations as

$$\left[\int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \sqrt[3]{c'^2 (t_1,t_2)\, f(t_1,t_2)}\, \partial t_1 \partial t_2\right]^3 = C_3$$

as the system of equations giving approximations to optimum points of stratification $[x_h, z_k]$ As remarked in case of optimum allocation method that in some particular cases some of the approximate systems given in above equations may be meaningless. Therefore depending upon the situation one should make the approximate choice of the systems of equations for obtaining the approximations to optimum points $[x_h, z_k]$.

4.18 Cum $\sqrt[3]{D_3(x,z)}$ Rule

If the function $D_3(x,z) = c'^2(x,z)\, f(x,z)$ is bounded and its first two derivatives exists for all x in $[a,b]$ and z in $[c,d]$ then for a given value of L and M taking equal intervals on the cumulative cube root of $D_3(x,z)$ will give AOSB $[x_h, z_k]$.

Remarks:

I. If we take either $c(x,z) = \alpha + \beta x$ or $c(x,z) = \alpha + \gamma z$ in $D_3(x,z)$ it reduces to the method proposed by Singh and Sukhatme (1969).

II. If the function $c'^2(x,z)$ is constant ,therefore proposed method reduces to Cum $\sqrt[3]{f(x,z)}$ rule.

And for any distribution and given number of strata the set of AOSB will remain unchanged with respect to the form of conditional variance. However, the efficiency of the stratification will differ from stratified simple random sampling estimators as well as other estimator with the choice of various forms of conditional variance.

4.19 Empirical study

For the purpose of empirical study, the effectiveness of the methods of finding approximation to the optimum points of stratification, we have considered the system of minimal equations obtained for the case of proportional allocation. In this illustration we shall consider equal interval approximation and the system of approximations given in (4.17) article. The former approximation is specially considered due to its simplicity. From all the later approximations we have only chosen one suitable method. Since the order of approximation involved in all these methods is same, this one approximation will give the idea about the effectiveness of all other approximation given (4.17). For the sake of simplicity, the linear regression line Y on X and Z have been taken as, of the form $y = \alpha + \beta x + \gamma z + e$ .Here it is considered that the two auxiliary variables used for stratification are dependent. From all the latter approximation we have only chosen one suitable method. Since the order of approximation involved in all these methods is same, this one approximation will give the idea about the effectiveness of all other approximation. For obtaining the stratification points under proportional allocation let us assume $c(x,z) = \alpha + \beta x + \gamma z.$ Further, let us assume that the correlation coefficient between X and Z is denoted by ρ and is equal to 0.65.Let us consider the following examples:

4.19.1 : Let us suppose that the auxiliary variable X follows standard normal distribution with pdf as

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad , x \geq 0$$

and the variable Z also follows standard normal distribution with pdf as

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad , z \geq 0$$

In order to obtain the OSB when both the variables are standard normally distributed let us assume the value of regression coefficients $\beta = 0.65$ and $\gamma = 0.57$ .For obtaining total 16 strata,4 along the X variable and 4 along the Z variable using the proposed rule Cum $\sqrt[3]{D_3(x,z)}$, by solving it in Mathematica software assuming that the distribution of X and Z are truncated at x = 6 and z = 4, respectively, we get the stratification points as below:

Table 4.19.1: OSB when the auxiliary variables X and Z have both standard normal distribution for the case of dependent variables

	4.0000				
	0.9865				
Z	**0.5284**				
	0.2673				
	0.0000	**0.3347**	**0.5779**	**1.9004**	**6.0000**
			X		

Table 4.19.2: OSB and Variance when the auxiliary variables are both standard normally distributed for dependent variables under proportional allocation

OSB (x_h, z_k)	**Variance** Cum $\sqrt[3]{D_3(x,z)}$ Rule	**Variance (Singh 1975a)**	**% R.E.**
(0.3347,0.2673) (0.5779,0.2673) (1.9004,0.2673) (6.0000,0.2673) (0.3347,0.5284) (0.5779,0.5284) (1.9004,0.5284) (6.0000,0.5284) ((0.3347,0.9865) (0.5779,0.9865) (1.9004,0.9865) (6.0000,0.9865) (0.3347,4.0000) (0.5779,4.0000) (1.9004,4.0000) (6.0000,4.0000)	0.06798628	0.182346	268.21

4.19.2 : Let us consider the distribution of X Uniform having probability density function (pdf) as

$$f(x) = 2(2-x), 1 \le x \le 2$$

and Z follows Exponential distribution with pdf as

$$f(z) = e^{-z+1} \quad , \quad 1 \le x \le 6$$

In order to obtain OSB for the above pdf let us suppose that the variable X is defined in [1, 2] and Z in [1, 6] and assume that values of β *and* γ be 0.576 and 0.257 respectively. Using the above pdf's for constructing stratification points using Cum $\sqrt[3]{D_3(x,z)}$ Rule for total 6 strata i,e 2 strata along X variable and 3 along Z variable, using Mathematica software for solving the function , we get the solution as shown in table given below.

Table 4.19.3: OSB and Variance when the auxiliary variables X and Z are independent following right triangular and exponential distribution, respectively

OSB $\left(x_h, z_k\right)$	Variance (Cum $\sqrt[3]{D_3(x,z)}$ Rule)
(1.48963,1.91467) (2.00000,1.91467) (1.48963,3.61677) (2.00000,3.61677) (1.48963,6.0000) (2.00000,6.0000)	0.097625

4.19.3 : Lets us consider the distribution of X as right triangular with pdf as

$$f\left(x\right)=2\left(2-x\right),1\le x\le 2$$

and the variable Z is having a exponential distribution with pdf as

$$f\left(z\right)=e^{-z+1} \quad , \quad 1\le x\le 6$$

In order to find the OSB when one of the auxiliary variable is following Right triangular distribution and other Exponential distribution we assume the value of $\beta = 0.576$ and $\gamma = 0.257$.The stratification points obtained for total 6 strata among that 2 along the X variable and 3 along Z variable for the Cum $\sqrt[3]{D_3\left(x,z\right)}$ rule using Mathematica Software for solving the function are presented in the following tales:

Table 4.19.4: OSB and Variance when the auxiliary variables X and Z are independent following right triangular and exponential distribution, respectively

OSB $\left(x_h, z_k\right)$	Variance (Cum $\sqrt[3]{D_3(x,z)}$ Rule)
(1.3492,1.9146) (2.0000,1.9146) (1.3492,3.6167) (2.0000,3.6167) (1.3492,6.0000) (2.0000,6.0000)	0.0602879

4.19.4 : Let us assume that one of the auxiliary variable, say, X follows log-normal distribution with pdf as

$$f(x)=\begin{cases}\dfrac{1}{\sigma x\sqrt{2\pi}}e^{-\dfrac{(\log x-\mu)^2}{2\sigma^2}} & ;x>0,\sigma>0\\ 0 \quad ;otherwise & \end{cases}$$

and the other auxiliary variable Z having pdf as:

$$f(z)=\begin{cases}\dfrac{1}{b-a} & ,a\le z\le b\\ o & ,otherwise\end{cases}$$

To obtain the OSB using the proposed method Cum $\sqrt[3]{D_3(x,z)}$ under proportional allocation when the auxiliary variables are independent and are having log-normal and uniform distribution let us assume log-normal distribution is to be standardized i,e μ=0 , σ = 1 and is defined in the interval $x\in[0,10] i,e. x_0=0, x_L=10$ and the other variable $z\in[0,1]$, $i,e. z_0=0,\ z_M=1$ and β=0.82 and γ=0.437.Further let us assume that the total strata to be made are 3×2(L×M)=6 i,e 3 along then X variable and 2 along the Z variable, the results obtained after solving the function using Mathematica software are presented in the following tables as:

Table 4.19.5: OSB for the case of dependent auxiliary variables X and Z following standard log-normal and uniform distributions

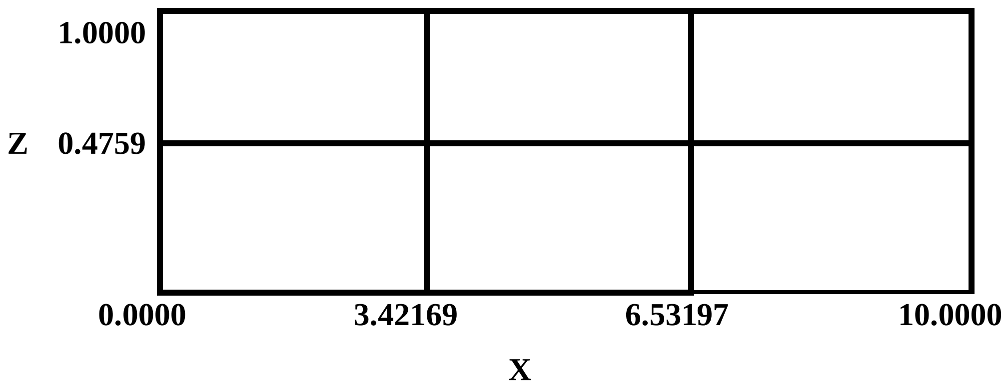


Table 4.19.6: OSB and Variance for the case of dependent auxiliary variables X and Z following standard log-normal and uniform distributions

OSB (x_h, z_k)	**Variance** (Cum $\sqrt[3]{D_3(x,z)}$ Rule)
(3.4216,0.4759) (6.5319,0.4759) (10.0000,0.4759) (3.4216,1.0000) (6.5319,1.0000) (10.0000,1.0000)	0.035281796

4.20 For independent auxiliary variables under proportional allocation (ρ = 0)

In order to propose a technique under proportional allocation when the two auxiliary variables are independent to each other we need to proceed in the same way as proceeded in case when they are dependent only with the difference that here in this case we have to take marginal densities rather than joint densities under consideration. We can write (4.16.3) as

$$\left[\frac{(k_h)^2}{8}\frac{B_h c^{'}(x)}{f(x)}\right]^{\frac{1}{3}}\left[\frac{(k_h)^2}{8}\frac{B_h c^{'}(z)}{f(z)}\right]^{\frac{1}{3}}$$

$$=\frac{(k_h)c^{'}(x)}{2}\left[1-\left(\frac{(f_x)c^{'}+2f(x)c^{''}(x)}{6f(x)c^{'}(x)}\right)(k_h)+O(k_h)^2\right]$$

$$\frac{(k_k)c^{'}(z)}{2}\left[1-\left(\frac{(f_z)c^{'}+2f(z)c^{''}(z)}{6f(z)c^{'}(z)}\right)(k_k)+O(k_k)^2\right]$$

In the similar way, we have

$$\left[\frac{(k_i)^2}{8}\frac{B_i c'(x)}{f(x)}\right]^{\frac{1}{3}}\left[\frac{(k_j)^2}{8}\frac{B_j c'(z)}{f(z)}\right]^{\frac{1}{3}}$$

$$=\frac{(k_i)c'(x)}{2}\left[1-\left(\frac{(f_x)c'+2f(x)c''(x)}{6f(x)c'(x)}\right)(k_i)+O(k_i)^2\right]$$

$$\frac{(k_j)c'(z)}{2}\left[1-\left(\frac{(f_z)c'+2f(z)c''(z)}{6f(z)c'(z)}\right)(k_j)+O(k_j)^2\right]$$

where $B_h=\int_{x_{h-1}}^{x_h} c'^2(t_1)f(t_1)dt_1 \quad and \quad B_k=\int_{z_{k-1}}^{z_k} c'^2(t_2)f(t_2)dt_2$

However, if the terms of order $O\left(\underset{a,b}{Sup}(k_h)\right)^3 \quad and \quad O\left(\underset{c,d}{Sup}(k_k)\right)^3$ can be neglected, we replace the minimal equations approximately by

$$\left[\frac{(k_h)^2}{8}\frac{B_h c'(x)}{f(x)}\right]^{\frac{1}{3}}\left[\frac{(k_h)^2}{8}\frac{B_h c'(z)}{f(z)}\right]^{\frac{1}{3}}=\left[\frac{(k_i)^2}{8}\frac{B_i c'(x)}{f(x)}\right]^{\frac{1}{3}}\left[\frac{(k_j)^2}{8}\frac{B_j c'(z)}{f(z)}\right]^{\frac{1}{3}}$$

or in other words we have that $k_h^2 B_h \quad and \quad k_k^2 B_k$ are constants. In case it is possible to find a function $Q_1'(x_{h-1},x_h) \quad and \quad Q_1'(z_{k-1},z_k)$ such that

$$(k_h)^2 B_h=(k_h)^2\int_{x_{h-1}}^{x_h} c'^2(t_1)f(t_1)dt_1$$
$$=Q_1'(x_{h-1},x_h)\left[1+O(k_h)^2\right]$$

and

$$(k_k)^2 B_k=(k_k)^2\int_{z_{k-1}}^{z_k} c'^2(t_2)f(t_2)dt_2$$
$$=Q_1'(z_{k-1},z_k)\left[1+O(k_k)^2\right]$$

The theorem 4.2 can be proceeded in case of independent variables too. Similarly approximate system of equations can be proposed in the same way as proposed in case when auxiliary variables are dependent like (4.17.2) and can be written as

$$(k_h k_k)^2\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k} c'^2(t_1)c'^2(t_2)\,f(t_1)\,f(t_2)\partial t_1\partial t_2=C_1$$
$$h=1,2,\ldots,L \text{ and } k=1,2,\ldots,M$$

The solution of this system of equations and also those that will now follow, are expected to be closer to the optimum points of stratification as compared to the strata obtained composed from $k_h = \frac{b-a}{L}$ = Constant and $k_k = \frac{d-c}{M}$ = Constant. Therefore, depending upon the situation one should make the approximate choice of system of equations for obtaining the approximation of optimum points of stratification.

4.21 Cum $\sqrt[3]{D_4(x,z)}$ Rule

For a given values of L and M taking equal intervals on the cumulative cube root of $D_4(x,z)$ will give AOSB if the function $D_4(x,z) = D_4(x)D_4(z) = c'^2(x) f(x) c'^2(z) f(z)$ is bounded and its first derivative exists in all $x \in [a,b]$ and $z \in [c,d]$.

Remarks:

1. If the functions $c'^2(x)$ *and* $c'^2(z)$ are constants, then the proposed method deduced to cum $\sqrt[3]{f(x) f(z)}$ rule.
2. If we take $c(x) = c(z)$ and $f(x) = f(z)$, then the proposed method reduced to the Yadava and Singh (1984) method of Cum $\sqrt[3]{B_2(x)}$

   where $B_2 = \frac{f(x)x^2c'^2(x) + x\phi'(x) - \phi(x)}{x^3}$

4.22 Empirical Study

We shall demonstrate empirically the effectiveness of proposed method of finding the set of approximately optimum strata boundaries (AOSB).For this purpose, we have considered the system of minimal equations obtained for the case of proportional allocation when the two auxiliary variables used for stratification are independent. From all the latter approximation we have only chosen one suitable method. Since the order of approximation involved in all these methods is same, this one approximation will give the idea about the effectiveness of all other approximation. For obtaining the stratification points under proportional allocation let us assume $c(x,z) = \alpha + \beta x + \gamma z$ .Let us consider the following examples:

4.22.1 : X has Right triangular distribution with pdf as

$$f(x) = 2(2-x), 1 \le x \le 2$$

and Z follows Exponential distribution with pdf as

$$f(z) = e^{-z+1}, 1 \le z \le 6$$

In order to obtain stratification points when the auxiliary variable X follows right triangular distribution defined in [1,2] and auxiliary variable Z follows exponential distribution defined in [1,6] we assume the values of $\beta = 0.567$ and $\gamma = 0.257$.While execution for obtaining OSB using Cum $\sqrt[3]{D_4(x,z)}$ Rule by solving the function using Mathematica Software for 6 strata ,2 along the X variable and 3 along the Z variable. The results obtained are presented in the following tables as:

Table 4.22.1: OSB when the auxiliary variables X and Z are independent having right triangular and exponential distribution, respectively

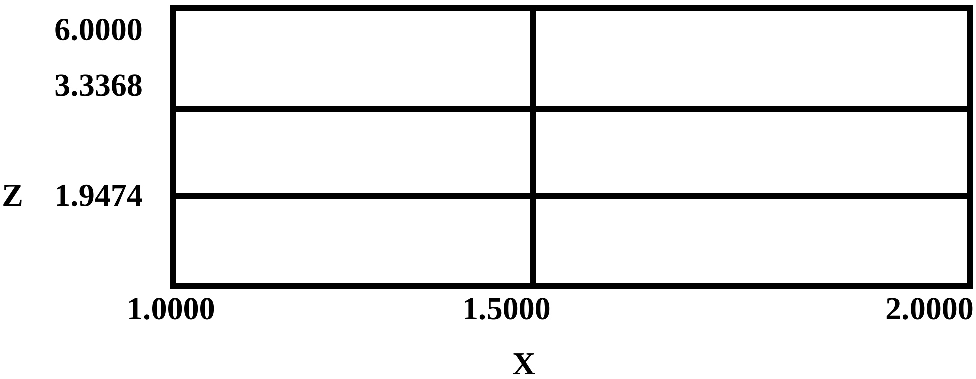


Table 4.22.2: OSB and Variance when the auxiliary variables X and Z are independent having right triangular and exponential distribution, respectively

OSB (x_h, z_k)	**Variance** (Cum $\sqrt[3]{D_4(x,z)}$ Rule)	**Variance (Yadava and Singh, 1984)**	**% R.E.**
(1.5000,1.9474) (1.0000,1.9474) (1.5000,3.3368) (1.0000,3.3368) (1.5000,6.0000) (1.0000,6.0000)	0.089542	0.152122	169.89

4.22.2 : Lets us consider the distribution of X as right triangular with pdf as

$$f(x) = 2(2-x), 1 \le x \le 4$$

and the variable Z is having a uniform distribution with pdf as

$$f(z) = \frac{1}{b-a}, 1 \le z \le 2$$

In order to find the OSB when one of the auxiliary variable is following Right triangular distribution and other uniform distribution we assume the value of $\beta = 0.56$ and $\gamma = 0.762$.The stratification points obtained for total 6 strata among that 3 along the X variable and 2 along Z variable for the Cum $\sqrt[3]{D_4(x,z)}$ Rule using Mathematica Software for solving the function are presented in the following tales as:

Table 4.22.3: OSB and Variance when the auxiliary variables X and Z are independent having right triangular and exponential distribution, respectively

OSB (x_h, z_k)	**Variance** (Cum $\sqrt[3]{D_4(x,z)}$ Rule)	**Variance (Khan *et al.* 2008)**	**% R.E.**
(1.7880,1.5000) (2.6870,1.5000) (4.0000,1.5000) (1.7880,2.0000) (2.6870,2.0000) (4.0000,2.0000)	0.0354952	0.08293	233.64

4.22.3 : Let us assume that one of the auxiliary variable, say, X follows log-normal distribution with pdf as

$$f(x) = \begin{cases} \dfrac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} & ; x > 0, \sigma > 0 \\ 0 & ; otherwise \end{cases}$$

and the other auxiliary variable Z having pdf as:

$$f(z) = \begin{cases} \dfrac{1}{b-a} & , a \le z \le b \\ o & , otherwise \end{cases}$$

To obtain the OSB using the proposed method Cum $\sqrt[3]{D_4\left(x,z\right)}$ under proportional allocation when the auxiliary variables are independent and are having log-normal and uniform distributions. Let us assume log-normal distribution is to be standardized i,e μ=0 , σ = 1 and is defined in the interval $x \in \left[0,10\right] i,e. x_0 = 0, x_L = 10$ and the other variable $z \in \left[0,1\right],\ i,e. z_0 = 0,\ \ z_M = 1$and β=0.82 and γ=0.437.Further let us assume that the total strata to be made are 3×2(L×M)=6 i,e 3 along the X variable and 2 along the Z variable, the results obtained after solving the function using Mathematica software are presented in the following table.

Table 4.22.4: OSB and Variance when the auxiliary variables X and Z are independent and follow standardized log-normal and uniform distributions, respectively

OSB $\left(x_h, z_k\right)$	**Variance** (Cum $\sqrt[3]{D_4(x,z)}$ Rule)	**Variance (Khan et al. 2015)**	**% R.E.**
(2.4719,0.5000) (5.5357,0.5000) (10.0000,0.5000) (2.4719,1.0000) (5.5357,1.0000) (10.0000,1.0000)	0.01025863	0.052942795	516.05

CHAPTER 5

MATHEMATICAL PROGRAMMING TECHNIQUE

5.1. Formulation of problem

Let the target population consisting of 'N' units be stratified into L×M strata based on two auxiliary variables 'X' and 'Z' when the estimation of mean of the study variable 'Y' is of interest. In order to have the estimate ,we divide the whole population into the desired number of strata, say L×M, such that each stratum is homogenous within itself and heterogeneous between strata with respect to the character under study such that the number of units in the (h, k)th stratum is N_{hk} ,so that

$$\sum_{h=1}^{L}\sum_{k=1}^{M} N_{hk} = N$$

From each stratum, a sample of size n_{hk} (h=1,2,...,L; k=1,2,...,M) is to be drawn such that a sample consisting of 'n' units be selected from a population under study of size 'N' such that $\sum_{h}\sum_{k} n_{hk} = n$.

Let y_{hki} ,(i=1,2,3,...,N_{hk}) denotes the population unit in the (h , k)th stratum ,then the population total can be expressed as

$$Y = \sum_{h}\sum_{k}\sum_{i} y_{hki}$$

Under stratified random sampling the unbiased estimator of the population mean $\overline{Y}_N$, is

$$\overline{y}_{st} = \sum_{h}\sum_{k} W_{hk}\,\overline{y}_{hk}$$

where, W_{hk} denotes the weight of the (h, k)th stratum given as $\frac{N_{hk}}{N}$ and $\overline{y}_{hk} = \frac{1}{n_{hk}}\sum_{i} y_{hki}$.

The sampling variance of the unbiased estimator '$\overline{y}_{st}$ ' is

$$V\left(\overline{y}_{st}\right) = \sum_{h}\sum_{k}\left(\frac{1}{n} - \frac{1}{N}\right) W_{hk}^2 \sigma_{hky}^2$$

where ,W_{hk} and σ_{hky}^2 are the weight and stratum variance for the $(h,k)^{th}$ stratum (h = 1,2,...,L ; k = 1,2,...,M), respectively. However, if the finite population correction (f.p.c) is ignored, $V\left(\overline{y}_{st}\right)$ can be expressed as

$$V\left(\bar{y}_{st}\right) = \sum_{h}\sum_{k} \frac{W_{hk}^{2}\sigma_{hky}^{2}}{n} \tag{5.1.1}$$

When the study variable 'Y' itself is not used for stratification variable, we propose a model based on two auxiliary variables. Let the regression model of study variable on auxiliary variables is of the form as:

$$Y = \lambda(x,z) + e \tag{5.1.2}$$

where, $\lambda(x,z)$ be a linear or non-linear function of 'X' and 'Z' and 'e' denotes the error term such that

$E(e \mid x,z) = 0$ and $V(e \mid x,z) = \phi(x,z)$ for all (x, z)

Under model (5.1.2) the stratum mean 'μ_{hky}' and the stratum variance 'σ_{hky}^2' can be written as

$$\mu_{hky} = \mu_{hk\lambda} \tag{5.1.3}$$

and

$$\sigma_{hky}^{2} = \sigma_{hk\lambda}^{2} + \mu_{hk\phi} \tag{5.1.4}$$

where $\mu_{hk\lambda}$ and $\mu_{hk\phi}$ are the expected values of $\lambda(x,z)$ and $\phi(x,z)$,respectively and $\sigma_{hk\lambda}^2$ denotes the variance of $\lambda(x,z)$ in the (h, k)th stratum.

If 'λ' and 'ε' are uncorrelated, then in model (5.1.2) 'σ_{hky}^2' can be expressed as

$$\sigma_{hky}^{2} = \sigma_{hk\lambda}^{2} + \sigma_{hk\varepsilon}^{2} \tag{5.1.5}$$

where $\sigma_{hk\varepsilon}^2$ is the variance of error term in (h, k)th stratum.

Let the joint density function of (X,Y,Z) in the super population is f(x, y, z) and joint marginal density function of X and Z is f(x, z) .Let f(x) and f(z) be the frequency function of the auxiliary variables X and Z ,respectively, defined in the interval [a, b] and [c, d].

If the population mean of the study variable 'Y' is estimated under the variance given in equation (5.1.1) ,then the problem of determining the strata boundaries is to cut up the ranges $d_x = b - a$ and $t_z = d - c$, at (L-1) and (M-1) intermediate points as $a = x_0 \le x_1 \le ... \le x_{L-1} \le x_L = b$ and $c = z_0 \le z_1 \le ... \le z_{M-1} \le z_M = d$, respectively, such that the equation (5.1.1) is minimum.

For a fixed size 'n' ,minimizing the expression of the right hand side of (5.1.1) is equivalent to minimizing

$$\sum_h \sum_k W_{hk}^2 \sigma_{hky}^2$$

as the value of ‘n’ is known in advance. Thus, while using (5.1.4), we have

$$\sum_h \sum_k W_{hk}^2 \left(\sigma_{hk\lambda}^2 + \mu_{hk\phi}\right) \tag{5.1.6}$$

If $f(x,z), \lambda(x,z)$ and $\phi(x,z)$ are known and also integrable then, $W_{hk}, \sigma_{hk\lambda}^2$ and $\mu_{hk\phi}$ can be obtained as a function of boundary points $(x_{h-1}, x_h, z_{k-1}, z_k)$ by using the following expression

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f(x,z)\,\partial x \partial z \tag{5.1.7}$$

$$\sigma_{hk\lambda}^2 = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda^2(x,z)\, f(x,z)\,\partial x \partial z - \mu_{hk\lambda}^2 \tag{5.1.8}$$

and
$$\mu_{hk\phi} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \phi(x,z) f(x,z)\,\partial x \partial z \tag{5.1.9}$$

Where,
$$\mu_{hk\lambda} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda(x,z) f(x,z)\,\partial x \partial z \tag{5.1.10}$$

and $(x_h, x_{h-1}, z_k, z_{k-1})$ are the boundary points of the (h, k)th stratum.

Thus, the objective function (5.1.6) could be expressed as the function of boundary points $(x_{h-1}, x_h, z_{k-1}, z_k)$ only.

Let
$$\phi_{hk}(x_h, x_{h-1}, z_k, z_{k-1}) = W_{hk}^2 \left(\sigma_{hk\lambda}^2 + \mu_{hk\phi}\right) \tag{5.1.11}$$

and the ranges as:

$$d_x = b - a = x_L - x_0 \tag{5.1.12}$$

$$t_z = d - c = z_M - z_0 \tag{5.1.13}$$

Then, in the bivariate stratification a problem of determining the strata boundaries (x_h, z_k) is to break up the ranges of (5.1.12) and (5.1.13) at intermediate points in order to estimate $x_1 \le x_2 \le \ldots \le x_{L-2} \le x_{L-1}$ and $z_1 \le z_2 \le \ldots \le z_{M-2} \le z_{M-1}$.Then, the reasonable criterion for determining optimum strata boundaries(OSB) $(x_h, z_k)$is to minimize

Minimize $\sum_h \sum_k \phi_{hk}(x_h, x_{h-1}, z_k, z_{k-1})$

Subject to (5.1.14)

$$a = x_0 \le x_1 \le \ldots \le x_{L-1} \le x_L = b$$

$$c = z_0 \le z_1 \le \ldots \le z_{M-1} \le z_M = d$$

and

$$\sum_h \sum_k n_{hk} = n$$

When, the marginal frequency functions are known then σ^2_{hky} can be expressed as a function of boundary points (x_h, z_k).For the rectangular stratification, let $V_h = x_h - x_{h-1}$ and $U_k = z_k - z_{k-1}$ denotes the total length or width of the (h, k)th stratum. Then, using (5.1.12) and (5.1.13) ,the ranges can be expressed as

$$\sum_h V_h = \sum_h (x_h - x_{h-1}) = b - a = d_x \tag{5.1.15}$$

$$\sum_k U_k = \sum_k (z_k - z_{k-1}) = d - c = t_z \tag{5.1.16}$$

The objective function in (5.1.14) suggests that, for determination of two way stratification, a two-dimensional dynamic programming approach should be used. Employing the general concept of dynamic programming with the state and decision variables by the pairs (h, k).Then problem of two way optimum stratification can be expressed as to

$$\begin{aligned}
&\textit{Minimize} \quad \sum_h \sum_k \phi_{hk}(x_h, x_{h-1}, z_k, z_{k-1}) \\
&\textit{Subject to} \\
&\quad (x_h, z_k) = (x_{h-1} + V_h, z_{k-1} + U_k) \\
&\qquad (x_h, z_k) \in [a, d] \times [c, d] \\
&\qquad (V_h, U_k) \in B_h(x_{h-1}) \times B_k(z_{k-1}) \\
&\qquad\quad = [0, b - x_{h-1}] \times [0, d - z_{k-1}] \\
&\qquad (x_0, z_0) = [a, c] \\
&\qquad h = 1, 2, ..., L \quad \textit{and} \quad k = 1, 2, ..., M
\end{aligned} \tag{5.1.17}$$

Though the formulation of (5.1.17) seems to be difficult, it can in fact be seen with respect to the decision space $B_h(x_{h-1}) \times B_k(z_{k-1})$.This decision space depends upon the states of both past and future stages since the variable x_{h-1} is present in the decision space of the M stages (h , 1) to (h , M) and the variable z_{k-1} in the decision spaces of the L stages (1,k) to (L , K).Hence, Bellman's principle of optimality that states, "An optimal policy has the property that, whatever the initial state and decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decisions" is not applicable. It should be noted that the problem is partially due to the way the stages have been defined. In fact instead of $L \times M$ stages which can be viewed through the decision variables V_h and U_k.However, due to nature of objective function given in general by (5.1.14),its transformation to reflect the (L x M) stages does not seem to be mathematically tractable for most allocations.

We propose a simple approach which permits a solution to the problem (5.1.17) using the unidmensional dynamic programming iteratively. Before the first iteration, some trail values say x_0 and z_0, such that $a = x_0 \le x_1 \le ... \le x_{L-1} \le x_L = b$ and $c = z_0 \le z_1 \le ... \le z_{M-1} \le z_M = d$ are choosen for the initial points of the stratification. Then for the i^{th} iteration (i=1,2,...) the points of stratification z^{i-1} are first considered as fixed. Note that the points of stratification x^{i-1}could also be chosen instead of z^{i-1}.Fixing the values of z^{i-1} has in fact the effect of reducing the problem exactly to the one of two-way optimum stratification with one categorical stratification variable. This can be seen by comparing the formulation (5.1.17) to the one which is defined on univarite auxiliary variable used as stratification variable with the values of the points of stratification Z taken as constant in (5.1.17).

Let $\phi^*_{x_h}\left(x_{h-1}, z^{i-1}\right)$ be the optimal value for the objective function (5.1.14) for the strata (h, k) to (L, k) for all k=1,2,...,M given that the lower bound for the strata (h, k) for k = 1,2,...,M is x_{h-1}.The functional equation of Bellman with respect to the first part of the i^{th} iteration is then given by

$$\phi^*_{x_h}\left(x_{h-1}, z^{i-1}\right) = \underset{V_h \in B_h\left(x_{h-1}\right)}{Minimize} \left\{ \sum_{k=1}^{M} \phi\left(x_{h-1}, x_h, z_{k-1}^{i-1}, z_k^{i-1}\right) + \phi^*_{x_{h+1}}\left(x_h, z^{i-1}\right) \middle| x_h = x_{h-1} + V \right\}$$

where $B_h\left(x_{h-1}\right)$ is defined in (5.1.17).

Using this last equation, new points of stratification x^i with respect to the variable 'X' can be obtained to response the proceeding value x^{i-1} .Hence, the OSB for the first part of the i^{th} iteration are given by $\left(x^i, z^{i-1}\right)$ .For the second part of the i^{th} iteration ,the points of stratification x^i are in turn considered as fixed. Restating the problem of determining OSB as the problem of determining optimum points $\left(V_h, U_k\right)$,adding equation (5.1.15) and (5.1.16) as a constraint ,the problem (5.1.14) can be treated as an equation problem of determining Optimum Strata Width (OSW), $V_1,V_2,...,V_L and U_1,U_2,...,U_M$ and is expressed as the following Mathematical Programming Problem (MPP):

$$\text{Minimize } \sum_h \sum_k \phi_{hk}\left(x_h, x_{h-1}, z_k, z_{k-1}\right)$$

Subject to (5.1.18)

$$\sum_h V_h = d_x$$

$$\sum_k U_k = t_z \qquad \text{h=1,2,...,L and k =1,2,...,M}$$

and

$$V_h \geq 0 \quad and \quad U_k \geq 0$$

Initially, $\left(x_0, z_0\right)$ the initial values of the auxiliary variables X and Z, respectively, are known. Therefore, the first term $\phi_{11}\left(x_1, x_0, z_1, z_0\right)$ in the objective function (5.1.18) is the function of $\left(V_1, U_1\right)$ alone, once the $\left(V_1, U_1\right)$ is known. The second term $\phi_{22}\left(x_2, x_1, z_2, z_1\right)$ will be the function of $\left(V_2, U_2\right)$ alone, and so on. Due to special nature of function, the MPP (5.1.18) may be treated as the function of $\left(V_h, U_k\right)$ and can be expressed as

$$\text{Minimize } \sum_h \sum_k \phi_{hk}\left(V_h, U_k\right)$$

Subject to (5.1.1)

$$\sum_h V_h = d_x$$

$$\sum_k U_k = t_z \qquad \text{,h=1,2,...,L and k =1,2,...,M}$$

and $$V_h \geq 0 \quad and \quad U_k \geq 0$$

5.2 The solution procedure

The problem (5.1.19) is a problem of multistage decision in which the objective function and the constraints are separable functions of $\left(V_h, U_k\right)$, which allows us to use a dynamic programming technique. Dynamic programming determines optimal solution of a multi-variable problem by decomposing into stages, each stage compromising a single variable sub problem. A dynamic programming model is generally a recursive equation. These recursive equation links to different stages of the problem.

Consider the following sub problem of equation (5.1.19) for first $\left(L_1 \times M_1\right)$ strata, where $\left(L_1 \times M_1\right) \le \left(L \times M\right)$, i, e $L_1 < L, M_1 < M$

$$\text{Minimize } \sum_{h=1}^{L_1} \sum_{k=1}^{M_1} \phi_{hk}\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$$

Subject to (5.2.1)

$$\sum_{h=1}^{L-1} V_h = d_{L_1}$$

$$\sum_{k=1}^{M-1} U_k = t_{M_1} \qquad ,h=1,2,...,L_1 \text{ and } k=1,2,...,M_1$$

and $V_h \ge 0 \quad and \quad U_k \ge 0$

where $d_{L_1} < d_x, t_{M_1} < t_z$

Note: If $d_{L_1} = d_x$ and $t_{M_1} = t_z$ then $\left(L_1 \times M_1\right) = \left(L \times M\right)$

The transformation functions are given by

$$d_{L_1} = V_1 + V_2 + ... + V_{L_1}$$
$$d_{L_1 - 1} = V_1 + V_2 + ... + V_{L_1 - 1} = d_{L_1} - V_{L_1}$$
$$d_{L_1 - 2} = V_1 + V_2 + ... + V_{L_1 - 2} = d_{L_1 - 1} - V_{L_1 - 1}$$

.

.

.

$$d_2 = V_1 + V_2 = d_3 - V_3$$
$$d_1 = V_1 = d_2 - V_2$$

Similarly, we have

$$t_{M_1} = U_1 + U_2 + ... + U_{M_1}$$
$$t_{M_1 - 1} = U_1 + U_2 + ... + U_{M_1 - 1} = t_{M_1} - U_{M_1}$$
$$t_{M_1 - 2} = U_1 + U_2 + ... + U_{M_1 - 2} = t_{M_1 - 1} - U_{M_1 - 1}$$

.

.

.

$$t_2 = U_1 + U_2 = t_3 - U_3$$
$$t_1 = U_1 = t_2 - U_2$$

Let $\phi_{L_1 \times M_1}\left(V_{L_1} \times U_{M_1}\right)$ denotes the minimum value of the objective function of the equation (5.2.1), that is,

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = Min\left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1} \phi_{hk}\left(V_h, U_k\right) \middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1-1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1-1}\right]$$

$$and \qquad V_h \geq 0, U_k \geq 0; h = 1,2,3,\ldots,L_1 \qquad ; \qquad k = 1,2,3,\ldots,M_1$$

with the above definition of $\phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$,the MPP (5.1.19) is equivalent to finding $\phi_{L \times M}\left(d_x, t_z\right)$ recursively by defining $\phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$ for $L_1 = 1,2,\ldots,L$ and $M_1 = 1,2,\ldots,M \;\; ; 0 \leq d_{L_1} \leq V, 0 \leq t_{M_1} \leq U$.

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right)$$
$$= Min \left[\begin{array}{l} \phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right) \\ + \left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1} \phi_{hk}\left(V_h, U_k\right) \middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1}\right] \end{array}\right]$$
$$and \qquad V_h \geq 0, U_k \geq 0; h = 1,2,3,\ldots,L_1 \qquad and \qquad k = 1,2,3,\ldots,M_1 \tag{5.2.2}$$

For fixed value of $\left(V_{L_1}, U_{M_1}\right)$, $0 \leq d_{L_1} \leq V \quad , \quad 0 \leq t_{M_1} \leq U$.

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = \phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$$
$$+ Min \left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1} \phi_{hk}\left(V_h, U_k\right) \middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1}\right]$$

and $\quad V_h \geq 0 \quad , h = 1,2,\ldots,L_1, U_k \geq 0 \quad , k = 1,2,\ldots,M_1, 1 \leq L_1 \leq L \quad , 1 \leq M_1 \leq M$

Using the same procedure to write the forward recursive equation of the dynamic programming technique and could obtain OSB.

Let us assume the regression model defined in equation (5.1.2) be linear as:

$$Y = \alpha + \beta x + \gamma z + e$$

$$\text{then} \qquad \sigma_{hky}^2 = \beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2 \tag{5.2.3}$$

The weight and variance of the (h, k)th stratum having auxiliary variables as 'X' and 'Z'.

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f\left(x, z\right) \partial x \partial z \tag{5.2.4}$$

$$\sigma_{hkx}^2 = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x^2 f(x) \partial x \partial z - \mu_{hkx}^2 \tag{5.2.5}$$

$$\sigma_{hkz}^2 = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z^2 f(z) \partial z \partial x - \mu_{hkz}^2 \tag{5.2.6}$$

where $\mu_{hkx} = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x f(x) \partial x \partial z$, $\mu_{hkz} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z f(z) \partial z \partial x$

5.3 Empirical Study (for general variance)

To illustrate the computation details of the proposed design, we consider the population of size 2000.

5.3.1: Let the auxiliary variable 'X' follows exponential distribution with probability density function (pdf) as below:

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} : x > 0 \\ 0 \quad : \quad otherwise \end{cases} \tag{5.3.1}$$

and the second auxiliary variable Z follows a right-triangular distribution with pdf as:

$$f(z, a, b) = \begin{cases} \dfrac{2(b-z)}{(b-a)^2}; a \le z \le b \\ 0 \qquad ; otherwise \end{cases} \tag{5.3.2}$$

Case 1: For Independent auxiliary variables

If the two stratification variables are independent then substitute the values of f(x) and f(z) taken from (5.3.1) and (5.3.2) equations, respectively in (5.2.4) , (5.2.5) and (5.2.6), we have

$$W_{hk} = e^{-\lambda x_h} \left(e^{\lambda V_h} - 1 \right) \frac{2bU_k - 2z_{k-1}U_k - U_k^2}{(b-a)^2} \tag{5.3.3}$$

$$\sigma_{hkx}^2 = \frac{\left(\lambda^2 x_{h-1} + 2\lambda x_{h-1} + 2\right) e^{\lambda V_h} - \lambda \left(V_h^2 + 2x_{h-1}V_h + x_{h-1}^2 \right) - 2\left(x_{h-1} + V_h\right)}{\lambda^2 \left(e^{\lambda x_h} - 1 \right)} - \left[\frac{\left(\lambda x_{h-1} + 1\right) - \lambda \left(x_{h-1} + V_h\right) - 1}{\lambda \left(e^{\lambda x_h} - 1 \right)} \right]^2 \tag{5.3.4}$$

$$\sigma_{hkz}^2 = \frac{\left(2bU_k - U_k^2\right) z_{k-1}^2 - 2z_{k-1}^3 U_k}{2bU_k - 2z_{k-1}U_k - U_{k-1}^2} - \left[\frac{\left(2bU_k - U_k\right)^2 - 2z_{k-1}^2 U_k}{2bU_k - 2z_{k-1}U_k - U_k^2} \right]^2 \tag{5.3.5}$$

Using the variance formula given in (5.1.1) and substitute in equation (5.2.3), we get MPP as

Minimize $\sum_{h}\sum_{k} W_{hk}^{2}\left(\beta^{2}\sigma_{hkx}^{2}+\gamma^{2}\sigma_{hkz}^{2}\right)$

Subject to

$$\begin{aligned}&\sum_{h} V_h = d_x\\&\sum_{k} U_k = t_z\\&\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{matrix} h = 1,2,...,L \\ k = 1,2,...,M \end{matrix}\end{aligned} \tag{5.3.6}$$

By substituting values obtained in equations (5.3.4) to (5.3.5) in MPP (5.3.6), we can put it as:

Minimize

$$\sum_{h}\sum_{k}\left(e^{-\lambda x_h}\left(e^{\lambda V_h}-1\right)\frac{2bU_k-2z_{k-1}U_k-U_k^2}{(b-a)^2}\right)^2\left\{\beta^2\frac{\left(\lambda^2 x_{h-1}+2\lambda x_{h-1}+2\right)e^{\lambda V_h}-[a_1]}{\lambda^2\left(e^{\lambda x_h}-1\right)}\right.$$

$$\left.-\left[\frac{(\lambda x_{h-1}+1)-\lambda(x_{h-1}+V_h)-1}{\lambda\left(e^{\lambda x_h}-1\right)}\right]^2+\gamma^2\frac{\left(2bU_k-U_k^2\right)z_{k-1}^2-2z_{k-1}^3U_k}{2bU_k-2z_{k-1}U_k-U_{k-1}^2}-[a_2]^2\right\}$$

Subject to

$$\begin{aligned}&\sum_{h} V_h = d_x\\&\sum_{k} U_k = t_z\end{aligned} \tag{5.3.7}$$

and $\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{matrix} h = 1,2,...,L \\ k = 1,2,...,M \end{matrix} \qquad a_1 = \lambda\left(V_h^2+2x_{h-1}V_h+x_{h-1}^2\right)+2\left(x_{h-1}+V_h\right),$

$$a_2 = \frac{\left(2bU_k-U_k\right)^2-2z_{k-1}^2U_k}{2bU_k-2z_{k-1}U_k-U_k^2}$$

Now, let the population divided into $L\times M = 12$ strata i,e L=4 and M=3, and suppose estimates of $x_0 = 0, d_x = 10, a = z_0 = 0, z_k = 2, t_z = 2.$ Also by simulation of the linear regression model using the pdf's (5.3.1) and (5.3.2) we get the estimates of coefficients $\beta = 0.5$ *and* $\gamma = 1.5$.

By executing the programme coded in LINGO of NLPP given in (5.3.7) for determing the OSB.The results were obtained as shown in following Tables.

Table 5.3.1: Stratification points when the auxiliary variables X and Z are independent and follow exponential and right triangular distributions

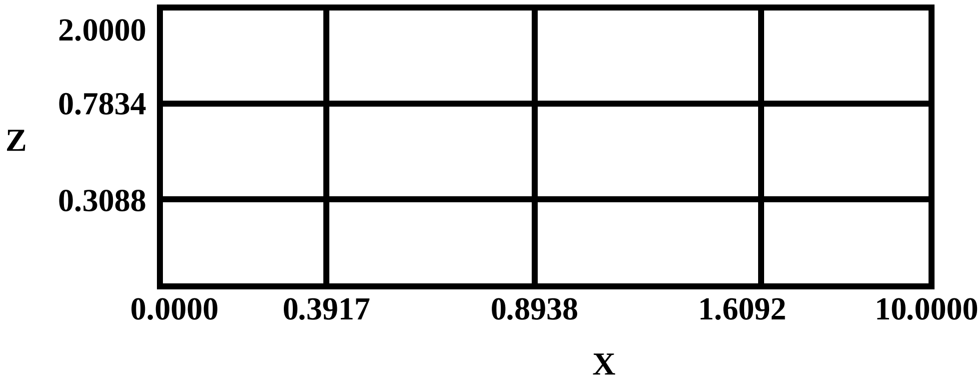
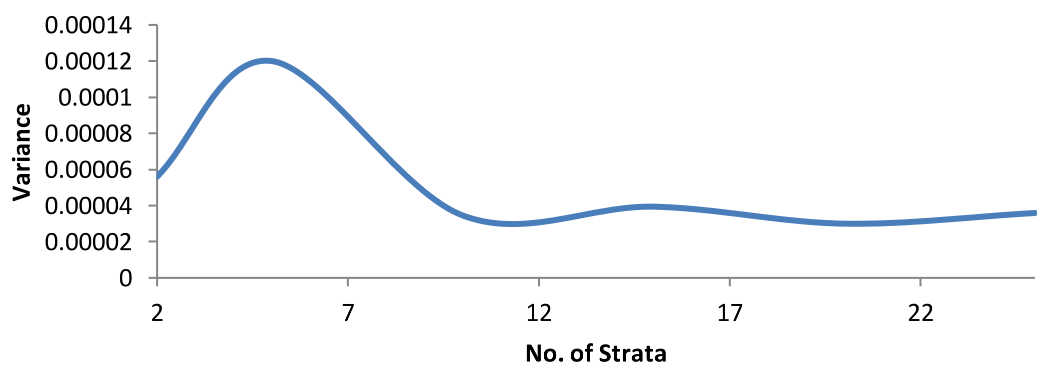


Table 5.3.2: OSB, Stratum Weight and Variance of the exponential and right triangularly distributed independent auxiliary variables

OSB (x_h, z_k)	Stratum Weight (W_{hk})	Variance
(0.3917, 0.3088)	0.0924	0.00003691
(0.8938, 0.3088)	0.0760	
(1.6092, 0.3088)	0.0596	
(10.0000, 0.3088)	0.0570	
(0.3917, 0.7834)	0.1118	
(0.8938, 0.7834)	0.0920	
(1.6092, 0.7834)	0.0721	
(10.0000, 0.7834	0.0696	
(0.3917, 2.0000)	0.1199	
(0.8938, 2.0000)	0.0987	
(1.6092, 2.0000)	0.0774	
(10.0000, 2.0000)	0.0740	

Fig. 1:Curve of number of strata and corresponding variance



It can be large even for small values of L and M. Here in Fig.1, we considered 25 number of strata but it can be observed that (which displays the graph between number of strata and variance) there is no substantial gain in efficiency for more than 20 strata. The Fig.1 shows that the variance remains constant for some strata when the number of strata reached 20 and then it shows an increasing trend. Thus the sufficient strata boundaries for each of the auxiliary variables would be 4 or 5.

Now, in order to make the comparison of the proposed methods and the existing methods, a simulation study is carried out to discuss the efficiency of the proposed method .The existing methods that are to be considered for comparison are:

1) Dalenius and Hodges (1959) cum $\sqrt{f}$ method

2) Gunning and Horgan (2004) geometric method

3) Lavallee-Hidiroglou (1988) method using Kozak's (2004) method

4) Khan *et al*. (2008) mathematical programming approach

5) Proposed method

In order to have the above comparison, the population of size N = 5000 with the stratification variables 'X' and 'Z', that follow an exponential and Right-triangular distributions, respectively, was randomly generated using the R-software.

For auxiliary variable X :

$$x_0 = 0.00006, x_L = 11.27211, d_x = 11.27205$$

Similarly, for the auxiliary variable Z:

$$z_0 = 0.00043, z_M = 2.96877, t_z = 2096834$$

If L=M=4, then the OSB's are determined by using R-package 1-3 and LINGO for 4 and 5.The results obtained, after executing the programme in both the softwares, are given in Table:5.3.3

Table 5.3.3: The variance of variables and their total variances

Method of Stratification	$v\left(\bar{x}_{st}\right)$ **(in E -07)**	$v\left(\bar{z}_{st}\right)$ **(in E -07)**	**Total variance (in E-07)**
Dalenius and Hodges (1959)cum$\sqrt{f}$ method	583.1884	275.1031	585.2915
Gunning and Horgan (2004) geometric method	3068.8590	1783.4450	4852.304
Lavallee-Hidiroglou (1988) method using Kozak's (2004) method	559.6009	439.4406	999.0415
Khan *et al.* (2008)	531.1695	340.9203	872.0898
Proposed method	499.7879	0.2451	500.0330

Thus, it reveals that the proposed method of stratification shows maximum reduction in variance of both the estimates as compared to other methods. Thereby, the proposed bi-variate stratification method is a better option for obtaining the OSB.

Case II: when the auxiliary variables are depended

If the auxiliary variables are dependent, then the MPP (5.3.6) would take the form as

Minimize

$$\sum_h \sum_k \left(e^{-\lambda x_h} \left(e^{\lambda V_h} - 1 \right) \frac{2bU_k - 2z_{k-1}U_k - U_k^2}{(b-a)^2} \right)^2 \left(\begin{array}{l} \beta^2 \dfrac{\left(\lambda^2 x_{h-1} + 2\lambda x_{h-1} + 2\right) e^{\lambda V_h} - [a_1]}{\lambda^2 \left(e^{\lambda x_h} - 1 \right)} \\ - \left[\dfrac{(\lambda x_{h-1} + 1) - \lambda\left(x_{h-1} + V_h\right) - 1}{\lambda \left(e^{\lambda x_h} - 1 \right)} \right]^2 \\ + \gamma^2 \dfrac{\left(2bU_k - U_k^2\right) z_{k-1}^2 - 2z_{k-1}^3 U_k}{2bU_k - 2z_{k-1}U_k - U_{k-1}^2} \\ -[a_2]^2 + 2\alpha\beta\rho(A)(B) \end{array} \right)$$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \\ &\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{aligned} h &= 1,2,\ldots,L \\ k &= 1,2,\ldots,M \end{aligned} \end{aligned} \tag{5.3.7}$$

where ρ is the correlation between X and Z variables and A and B denotes the standard deviation of X and Z in $(h, k)^{th}$ stratum and are given as

$$A = \sigma_{hkx} = sqrt\left[\frac{\left(\lambda^2 x_{h-1} + 2\lambda x_{h-1} + 2\right)e^{\lambda V_h} - [a_1]}{\lambda^2\left(e^{\lambda x_h} - 1\right)} - \left[\frac{(\lambda x_{h-1} + 1) - \lambda\left(x_{h-1} + V_h\right) - 1}{\lambda\left(e^{\lambda x_h} - 1\right)}\right]^2\right]$$

and

$$B = \sigma_{hkz} = Sqrt\left[\frac{\left(2bU_k - U_k^2\right)z_{k-1}^2 - 2z_{k-1}^3 U_k}{2bU_k - 2z_{k-1}U_k - U_{k-1}^2} - [a_2]^2\right]$$

Now as above, let the population be divided into $L \times M = 6$ strata ,i,e L=3 and M=2 and suppose that estimates of $x_0 = 0, d_x = 10, a = z_0 = 0, z_k = 2, t_z = 2.$ Also by simulation of the linear regression model using the pdf's (5.3.1) and (5.3.2), we get the estimates of coefficients $\beta = 0.5$ *and* $\gamma = 1.5$.and correlation coefficient $\rho = 0.342$.

By executing the programme coded in LINGO of NLPP given in (5.3.7) for determing the OSB would give the required result as given in below Tables

Table 5.3.4: OSB when the auxiliary variables X and Z are correlated and are exponentially and tight- triangularly distributed

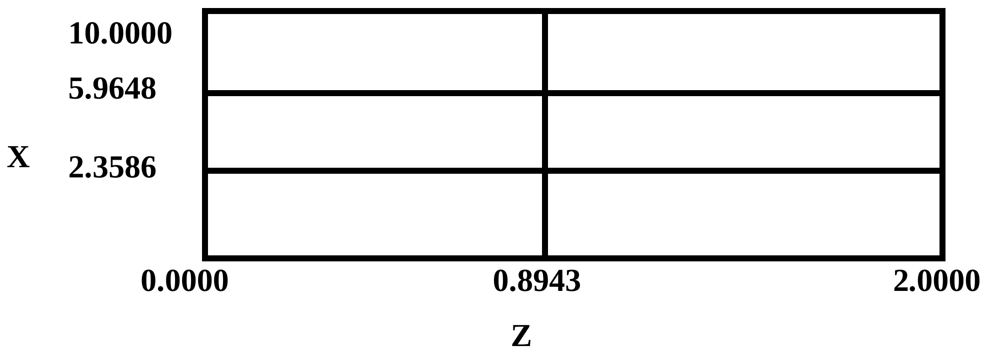


Table 5.3.5: Displays OSB and Variance for correlated auxiliary variables having exponentially and right- triangularly distribution

OSB (x_h, z_k)	**Variance**
(2.3586,0.8943)	
(5.9648,0.8943)	
(10.0000,0.8943)	0.0045369
(2.3586,2.0000)	
(5.9648,2.0000)	
(10.0000,2.0000)	

Table 5.3.4 shows the stratification points of the stratum while as Table 5.3.5 shows the OSB (x_h, z_k) with the variance obtained when having 6 number of strata, 3 along X variable and 2 along Z variable.

5.3.2: Let the auxiliary variable X follows uniform distribution with pdf as

$$f(x)=\begin{cases}\dfrac{1}{b-a} & ,a\le x\le b\\ o & ,otherwise\end{cases} \tag{5.3.8}$$

and the ‘Z follows the exponential distribution with pdf as

$$f(z)=\begin{cases}e^{-z+1} & ;\quad 1\le z\le 4\\ 0 & ;otherwise\end{cases} \tag{5.3.9}$$

For the case of zero correlation between X and Z, substitute values of (5.3.8) and (5.3.9) in (5.2.4), (5.2.5) and (5.2.6), we get

$$W_{hk}=\frac{V_h}{b-a}e^{-z_k+1}\left[e^{U_k}-1\right] \tag{5.3.10}$$

$$\sigma^2_{hkx}=\frac{4U_k\left(V_h^2+3x_hx_{h-1}\right)\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^3-3U_k^2\left(V_h+2x_{h-1}\right)^2}{12\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^2} \tag{5.3.11}$$

$$\sigma^2_{hkz}=\frac{b-a}{\left(e^{U_k}-1\right)^2}\begin{Bmatrix}z_{k-1}^2\left(e^{U_k}-1\right)^2-\left(e^{U_k}-1\right)\left[U_k^2+2e^{U_k}\left(1+z_{k-1}\right)-2\left(1+z_k\right)\right]\\ -\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]^2\end{Bmatrix}$$

(5.3.12)

Substituting (5.3.10),(5.3.11) and (5.3.12) in equation (5.3.5),we get MPP as
Minimize

$$\sum_h \sum_k \left(\frac{V_h}{b-a} e^{-z_k+1}\left[e^{U_k}-1\right]\right)^2$$

$$\left\{ \beta^2 \left(\frac{4U_k\left(V_h^2+3x_h x_{h-1}\right)\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^3-3U_k^2\left(V_h+2x_{h-1}\right)^2}{12\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^2} \right) \right.$$

$$\left. +\gamma^2 \left(\frac{b-a}{\left(e^{U_k}-1\right)^2} \left\{ \begin{array}{l} z_{k-1}^2\left(e^{U_k}-1\right)^2-\left(e^{U_k}-1\right)\left[\begin{array}{l} U_k^2+2e^{U_k}\left(1+z_{k-1}\right) \\ -2\left(1+z_k\right) \end{array}\right] \\ -\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]^2 \end{array} \right\} \right) \right\}$$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \\ &\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h=1,2,...,L \\ k=1,2,...,M \end{array} \end{aligned} \tag{5.3.13}$$

By simulation in R-software with the truncated exponential distribution and uniform distribution we get the estimate of $\beta=0.5$ and γ = 0.65.Further execution of a programme in LINGO by saving the above MPP with b=2, a=1 while taking L=3 and M=2, we get

The above results can be written in better form as

Table 5.3.6: OSB and Variance when the auxiliary variables are uniformly and exponentially distributed

OSB (x_h, z_k)	**Variance (Proposed method)**	**Variance (Singh 1977)**	**% R.E.**
(1.4192,1.8213) (1.4192,4.2457) (1.4192,6.0000) (2.0000,1.8213) (2.0000,4.2457) (2.0000,6.0000)	0.0093158	0.02426	260.42

While comparing its variance with the already existing developed method by Singh (1977) it is to be concluded that the percent relative efficiency is 260.42.Hence the superiority of the proposed method is proved.

5.3.3: Let us consider that one of the auxiliary variables 'X' follows a distribution with pdf as

$$f(x)=\begin{cases} e^{-x+1} & ; \quad 1\le z\le 4 \\ 0 & ; otherwise \end{cases} \tag{5.3.14}$$

and the other variable 'Z' follows Pareto distribution with pdf as

$$f(z)\begin{cases} \dfrac{ab^{a}}{z^{a+1}} & ; \quad z\in(b,\infty) \\ 0 & ; \quad otherwise \end{cases} \tag{5.3.15}$$

where $a>0$ is the shape of the parameter and $b>0$ is the scale parameter.

In order to obtain OSB, when the auxiliary variables X and Z have distribution functions given in (5.3.14) and (5.3.15),we need to find the values of (5.2.4) , (5.2.5) and (5.2.6),we have

$$W_{hk}=b^{a}e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\frac{\left(U_k+z_{k-1}\right)^2-\left(z_{k-1}\right)^a}{\left(U_k+z_{k-1}\right)^2\left(z_{k-1}\right)^a}\right] \tag{5.3.16}$$

$$\sigma^2_{hkx}=U_kV_h\left(U_k+z_{k-1}\right)^a\left(z_{k-1}\right)^a \left\{\frac{\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-\frac{1}{2}\left[\left(V_h+x_{h-1}\right)\left(V_h\left(V_h+x_{h-1}\right)+x_{h-1}\left(1+x_{h-1}\right)\right)\right]}{b^{a}e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k+z_{k-1}\right)^a-\left(z_{k-1}\right)^a\right]}\right\} -\left(A_1\right)^2 \tag{5.3.17}$$

$$\sigma^2_{hkz}=\frac{a\left(U_k+z_{k-1}\right)^2\left(z_{k-1}\right)^2\left[\left(U_k+z_{k-1}\right)^{a-2}-\left(z_{k-1}\right)^{a-2}\right]}{(a-2)e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k+z_{k-1}\right)^a-\left(z_{k-1}\right)^a\right]} -\left(\frac{a\left(U_k+z_{k-1}\right)\left(z_{k-1}\right)\left[\left(U_k+z_{k-1}\right)^{a-1}-\left(z_{k-1}\right)^{a-1}\right]}{(a-1)e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k+z_{k-1}\right)^a-\left(z_{k-1}\right)^a\right]}\right)^2 \tag{5.3.18}$$

where $$A_1 = \frac{U_k (U_k + z_{k-1})^2 (z_{k-1})^a \left[x_{h-1} - (V_h + x_{h-1} + 1) e^{-V_h} + 1 \right]}{b^a \left(1 - e^{-V_h}\right) \left[(U_k + z_{k-1})^a - (z_{k-1})^a \right]}$$

Substituting (5.3.16),(5.3.17) and (5.3.18) in equation (5.3.5),we get MPP as

Minimize

$$\sum_h \sum_k \left[b^a e^{-x_{h-1}+1} \left(1 - e^{-V_h}\right) \left[\frac{(U_k + z_{k-1})^2 - (z_{k-1})^a}{(U_k + z_{k-1})^2 (z_{k-1})^a} \right] \right]^2$$

$$\left\{ \begin{array}{l} \beta^2 U_k V_h (U_k + z_{k-1})^a (z_{k-1})^a \\ \left\{ \dfrac{\frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[(V_h + x_{h-1})\left(V_h (V_h + x_{h-1}) + x_{h-1}(1 + x_{h-1})\right) \right]}{b^a e^{-x_{h-1}+1} \left(1 - e^{-V_h}\right) \left[(U_k + z_{k-1})^a - (z_{k-1})^a \right]} \right\} \end{array} \right.$$

$$-(A_1)^2 + \gamma^2 \frac{a (U_k + z_{k-1})^2 (z_{k-1})^2 \left[(U_k + z_{k-1})^{a-2} - (z_{k-1})^{a-2} \right]}{(a-2) e^{-x_{h-1}+1} \left(1 - e^{-V_h}\right) \left[(U_k + z_{k-1})^a - (z_{k-1})^a \right]}$$

$$\left. - \left(\frac{a (U_k + z_{k-1})(z_{k-1}) \left[(U_k + z_{k-1})^{a-1} - (z_{k-1})^{a-1} \right]}{(a-1) e^{-x_{h-1}+1} \left(1 - e^{-V_h}\right) \left[(U_k + z_{k-1})^a - (z_{k-1})^a \right]} \right)^2 \right\}$$

Subject to

$$\sum_h V_h = d_x$$
$$\sum_k U_k = t_z \tag{5.3.19}$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, ..., L \\ k = 1, 2, ..., M \end{array}$$

Assuming the variable Z that follows pareto distribution is defined in the interval [1.000 , 10.000] and also assume that a=1.342 and this implies that b =1.000421.The (5.3.19) MPP can be put as

Minimize

$$\sum_h \sum_k \left[(1.00056) e^{-x_{h-1}+1} \left(1-e^{-V_h}\right) \left[\frac{\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}}{\left(U_k + z_{k-1}\right)^{1.342} \left(z_{k-1}\right)^{1.342}} \right] \right]^2$$

$$\left\{ \begin{array}{l} \beta^2 U_k V_h \left(U_k + z_{k-1}\right)^{1.342} \left(z_{k-1}\right)^{1.342} \\ \left\{ \dfrac{\frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[\left(V_h + x_{h-1}\right)\left(V_h\left(V_h + x_{h-1}\right) + x_{h-1}\left(1 + x_{h-1}\right)\right)\right]}{(1.00056) e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]} \right\} \end{array} \right.$$

$$-\left(\frac{U_k \left(U_k + z_{k-1}\right)^{1.342} \left(z_{k-1}\right)^{1.342} \left[x_{h-1} - \left(V_h + x_{h-1} + 1\right) e^{-V_h} + 1 \right]}{(1.00056)\left(1-e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]} \right)^2$$

$$+\gamma^2 \frac{1.342\left(U_k + z_{k-1}\right)^2 \left(z_{k-1}\right)^2 \left[\left(U_k + z_{k-1}\right)^{-0.658} - \left(z_{k-1}\right)^{-0.658}\right]}{\left(-0.658\right) e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]}$$

$$\left. -\left(\frac{1.342\left(U_k + z_{k-1}\right)\left(z_{k-1}\right)\left[\left(U_k + z_{k-1}\right)^{0.342} - \left(z_{k-1}\right)^{0.342}\right]}{\left(0.342\right) e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]} \right)^2 \right\}$$

Subject to

$$\sum_h V_h = 3$$

$$\sum_k U_k = 9 \tag{5.3.20}$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, \ldots, L \\ k = 1, 2, \ldots, M \end{array}$$

By using the pdf's given in (5.3.15) and (5.3.16) the simulation has been made to get the values of $\beta = 0.0743, \lambda = 0.421$ in R-Software. Solving the (5.3.20) following the recursive equations by executing a computer programme in LINGO to obtain OSB for the above objective function having total 12 (3×4) strata. are presented in table 5.3.8.

Table 5.3.7: OSB and Variance for exponential and pareto distributed auxiliary variables

OSB $\left(x_h, z_k\right)$	Variance
(1.6214,2.0512) (2.9782,2.0512) (4.0000,2.0512) (1.6214,3.9126) (2.9782,3.9126) (4.0000,3.9126) (1.6214,6.1421) (2.9782,6.1421) (4.0000,6.1421) (1.6214,10.0000) (2.9782,10.0000) (4.0000,10.0000)	0.002412

This shows the OSB when we have the total of 12 strata, 3 along X variable and 4 along Z variable as also the variance obtained by the proposed method.

5.4 Equal allocation

One of the type of allocations in stratified sampling design which is used in a situation of considerable practical interest for administrative or field convenience. In this method the total sample size $\sum_h \sum_k n_{hk} = n$ is divided equally among all the strata desired i,e for (h,k)th stratum

$$n_{hk} = \frac{n}{L \times M}$$

where 'n_{hk}' is the sample size taken from the (h, k)th stratum ,'n' is the total sample size need to be selected and L×M is the desired number of strata that are to be made.

Under equal allocation the variance is given below

$$V\left(\bar{y}_{st}\right) = \left(\frac{L \times M}{n}\right) \sum_h \sum_k (1-f) W_{hk}^2 \sigma_{hky}^2 \tag{5.4.1}$$

However, if the finite population correction is ignored, then it is sufficient to minimize $(L \times M) \sum_h \sum_k W_{hk}^2 \sigma_{hky}^2$ because 'n' is constant. Thus in this case the MPP would take the form as

Minimize $(L\times M)\sum_h\sum_k W_{hk}^2\sigma_{hky}^2$

Subject to

$$\begin{aligned}&\sum_h V_h = d_x\\&\sum_k U_k = t_z\end{aligned} \tag{5.4.2}$$

$$\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{aligned}h&=1,2,...,L\\k&=1,2,...,M\end{aligned}$$

By substituting values obtained in equation (5.2.3) in (5.4.2), we get the problem as below:

Minimize

$$(L\times M)\sum_h\sum_k W_{hk}^2\left(\beta^2\sigma_{hkx}^2+\gamma^2\sigma_{hkz}^2\right)$$

Subject to

$$\begin{aligned}&\sum_h V_h = d_x\\&\sum_k U_k = t_z\end{aligned} \tag{5.4.3}$$

$$\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{aligned}h&=1,2,...,L\\k&=1,2,...,M\end{aligned}$$

5.4.1: Empirical study

For empirical study let us assume that one of the auxiliary variables, say X, follows uniform distribution having pdf as

$$f(x)=\begin{cases}1 & ,1\le x\le 2\\ o & ,otherwise\end{cases} \tag{5.4.4}$$

and the other auxiliary variable follows a distribution with pdf

$$f(z)=\begin{cases}2(2-z) & ;1\le z\le 2\\ 0 & ;otherwise\end{cases} \tag{5.4.5}$$

Case I: When X and Z are independent

In order to obtain the OSB for a study variable having linear regressed with two auxiliary variables with distribution function defined in equations (5.4.4) and (5.4.5) we need to find the values of (5.2.4),(5.2.5) and (5.2.6).By substituting above density functions in them we get

$$W_{hk}=V_h\left(U_k-U_k^2-2z_{k-1}\right) \tag{5.4.6}$$

$$\sigma_{hkx}^2=\frac{\left(4-U_k-2z_{k-1}\right)\left(V_h^2+3x_{h-1}+3V_hx_{h-1}\right)-3V_h^2+12x_{h-1}^2+12V_hx_{k-1}}{3\left(4-U_k-2z_{k-1}\right)^2} \tag{5.4.7}$$

$$\sigma_{hkz}^2=\frac{\left(4-U_k-2z_{k-1}\right)\left[\frac{4}{3}\left(U_k^2+3z_{k-1}^2+3U_kx_{h-1}\right)-\frac{1}{2}\left(U_k+2z_{k-1}\right)\left(U_k^2+2z_{k-1}^2+2U_kz_{k-1}\right)\right]}{\left(4-U_k-2z_{k-1}\right)^2}$$

$$\frac{-4\left[U_k+2z_{k-1}-\frac{1}{3}\left(U_k^2+3z_{k-1}^2+3U_kz_{k-1}\right)\right]^2}{\left(4-U_k-2z_{k-1}\right)^2} \tag{5.4.8}$$

while substituting (5.4.6) to (5.4.7) in MPP (5.4.3),we have

Minimize

$$\left(L\times M\right)\sum_h\sum_k V_h^2\left(U_k-U_k^2-2z_{k-1}\right)^2$$

$$\left\{\beta^2\frac{\left(4-U_k-2z_{k-1}\right)\left(V_h^2+3x_{h-1}+3V_hx_{h-1}\right)-3V_h^2+12x_{h-1}^2+12V_hx_{k-1}}{3\left(4-U_k-2z_{k-1}\right)^2}\right.$$

$$\left.+\gamma^2\frac{\left(4-U_k-2z_{k-1}\right)\left[u_2\right]-4\left[u_1\right]^2}{\left(4-U_k-2z_{k-1}\right)^2}\right\}$$

Subject to

$$\begin{aligned}&\sum_h V_h=d_x\\&\sum_k U_k=t_z\end{aligned} \tag{5.4.8}$$

$$\forall V_h\geq 0, U_k\geq 0 \quad , \quad \begin{array}{l}h=1,2,\ldots,L\\k=1,2,\ldots,M\end{array}$$

where $u_1=U_k+2z_{k-1}-\frac{1}{3}\left(U_k^2+3z_{k-1}^2+3U_kz_{k-1}\right)$.

$$u_2=\frac{4}{3}\left(U_k^2+3z_{k-1}^2+3U_kx_{h-1}\right)-\frac{1}{2}\left(U_k+2z_{k-1}\right)\left(U_k^2+2z_{k-1}^2+2U_kz_{k-1}\right)$$

However, it is given in the distribution functions defined above that $x_0 = 1, d_x = 2, t_z = 2, z_0 = 1$ .Also by simulation of the two distributions in R-software we get the value of $\beta = 0.24, \gamma = 0.479$ .For (3×2) = 6 strata the MPP that we have to minimize for obtaining OSB is

Minimize

$$6\sum_{h=1}^{3}\sum_{k=1}^{2} V_h^2 \left(U_k - U_k^2 - 2z_{k-1}\right)^2$$
$$\left\{(0.57)\frac{\left(4-U_k-2z_{k-1}\right)\left(V_h^2+3x_{h-1}+3V_h x_{h-1}\right)-3V_h^2+12x_{h-1}^2+12V_h x_{k-1}}{3\left(4-U_k-2z_{k-1}\right)^2}\right.$$
$$\left.+(0.229)\frac{\left(4-U_k-2z_{k-1}\right)[u_2]-4[u_1]^2}{\left(4-U_k-2z_{k-1}\right)^2}\right\}$$

Subject to

$$\sum_{h=1}^{3} V_h = 1$$
$$\sum_{k=1}^{2} U_k = 1 \qquad (5.4.9)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2,3 \\ k = 1,2 \end{array}$$

By executing a computer programme in LINGO to the above MPP we get the following results

Table 5.4.1: OSB when the auxiliary variables X and Z are uniformly and right triangularly distributed

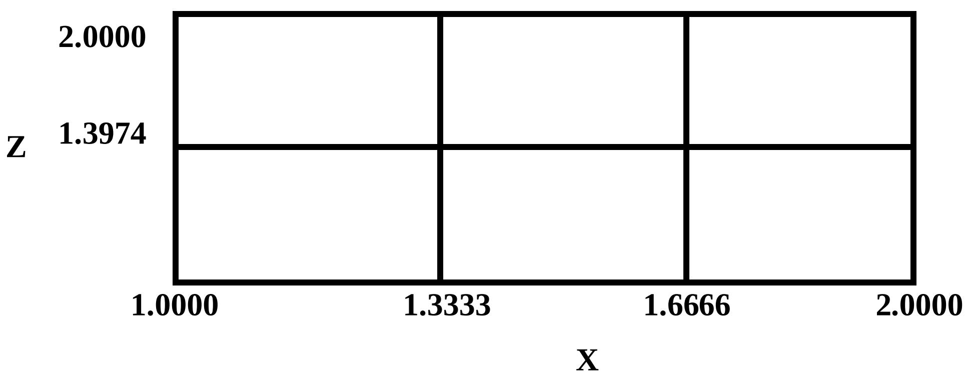


Table 5.4.2: OSB and Variance for uniform and right triangular distributed auxiliary variables

OSB (x_h, z_k)	Variance (Proposed method)	Variance (Singh 1977)	% R.E.
(1.3333,1.3974) (1.6666,1.3974) (2.0000,1.3974) (1.3333,2.0000) (1.6666,2.0000) (2.0000,2.0000)	0.016215	0.01486	276.66

The above table reveals that the variance obtained by the proposed method is much less than the method proposed by Singh (1977). Also the percent relative efficiency of the proposed method over Singh (1977) is 276.66, which shows the efficiency of the proposed method. Hence, we conclude that the proposed method of obtaining OSB when the two independent auxiliary variables are uniformly and Right triangularly distributed is preferable.

Case II: When X and Z are dependent

When the two auxiliary variables are not independent then the term of covariance would be included in the objective function. Having the same condition with same interval as taken in case of independent with ρ=0.62 obtained by simulation in R-software, the MPP that we need to minimize when the variables are dependent would be

Minimize

$$6\sum_{h=1}^{3}\sum_{k=1}^{2} V_h^2\left(U_k - U_k^2 - 2z_{k-1}\right)^2$$

$$\left[\begin{array}{l}(0.57)\dfrac{(4-U_k-2z_{k-1})\left(V_h^2+3x_{h-1}+3V_hx_{h-1}\right)-3V_h^2+12x_{h-1}^2+12V_hx_{k-1}}{3(4-U_k-2z_{k-1})^2} \\ +(0.229)\dfrac{(4-U_k-2z_{k-1})[u_2]-4[u_1]^2}{(4-U_k-2z_{k-1})^2} \\ +(0.1618)(A)(B) \\ \\ \end{array}\right]$$

Subject to

$$\sum_{h=1}^{3} V_h = 1$$
$$\sum_{k=1}^{2} U_k = 1 \qquad (5.4.9)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{matrix} h = 1,2,3 \\ k = 1,2 \end{matrix}$$

where A and B are the standard deviation of X and Z respectively and are given by

$$A = \left[\frac{\left(4 - U_k - 2z_{k-1}\right)\left(V_h^2 + 3x_{h-1} + 3V_h x_{h-1}\right) - 3V_h^2 + 12x_{h-1}^2 + 12V_h x_{k-1}}{3\left(4 - U_k - 2z_{k-1}\right)^2} \right]^{\frac{1}{2}}$$

and $B = \left[\dfrac{\left(4 - U_k - 2z_{k-1}\right)\left[u_2\right] - 4\left[u_1\right]^2}{\left(4 - U_k - 2z_{k-1}\right)^2} \right]^{\frac{1}{2}}$

By executing a computer programme in LINGO to the above MPP, we get the following results

Table 5.4.3: OSB and Variance for uniform and right triangular distributed auxiliary variables

OSB $\left(x_h, z_k\right)$	**Variance (Proposed method)**
(1.2371,1.5000) (1.7825,1.5000) (2.0000,1.5000) (1.23825,2.0000) (1.7825,2.0000) (2.0000,2.0000)	0.0045812

It is to be revealed that the strata boundaries changes if the variables are dependent. However the variance is less in case of dependent rather than independent.

5.5 Proportional Allocation

The allocation generally known as proportional allocation was originally proposed by Bowley (1926).This procedure of allocation is very common in practice because of its simplicity, when no other information other than N_{hk}, which denotes the total number of units in the $(h,k)^{th}$ stratum is available, the allocation of a given sample size 'n' to different strata is done in proportion to their sizes i,e in the $(h,k)^{th}$ stratum

$$n_{hk} = \frac{n}{N} N_{hk}$$

This means that the sampling fraction is the same in all strata. It gives a self weighing sample by which numerous estimates can be made with greater speed and a higher degree of precision.

Under proportional allocation the variance is given by

$$V\left(\bar{y}_{st}\right) = \frac{(1-f)}{n} \sum_h \sum_k W_{hk} \sigma_{hky}^2$$

where $f = \frac{n}{N}$ is sampling fraction. If the finite population correction is ignored, we get

$$V\left(\bar{y}_{st}\right) = \frac{1}{n} \sum_h \sum_k W_{hk} \sigma_{hky}^2$$

Minimizing this function is equivalent to minimizing

$$\sum_h \sum_k W_{hk} \sigma_{hky}^2 \tag{5.5.1}$$

Using the same procedure as discussed in the case of general and equal allocation, we need to replace the equation the objective function by $\sum_h \sum_k W_{hk} \sigma_{hky}^2$, Thus the MPP that we have to minimize is

Minimize $\sum_h \sum_k W_{hk} \sigma_{hky}^2$

Subject to

$$\sum_h V_h = d_x$$
$$\sum_k U_k = t_z \tag{5.5.2}$$

$$\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{array}{l} h = 1,2,...,L \\ k = 1,2,...,M \end{array}$$

The solution procedure would too be same as discussed earlier in previous allocation, which further can be elaborated as by replacing value from (5.1.1) in (5.4.2) MPP. Thus, we have

Minimize $\sum_h \sum_k W_{hk} \left(\beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2 \right)$

Subject to

$$\sum_h V_h = d_x$$
$$\sum_k U_k = t_z \tag{5.5.3}$$

$$\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{array}{l} h = 1,2,...,L \\ k = 1,2,...,M \end{array}$$

5.5.1 Empirical study

Let the variable X follows a distribution with pdf as

$$f(x) = \begin{cases} 2(2-x) & ; 1 \le x \le 2 \\ 0 & ; otherwise \end{cases} \tag{5.5.4}$$

and the other auxiliary variable Z follows truncated exponential distribution with pdf

$$f(z) = \begin{cases} e^{-z+1} & ; \quad 1 \le z \le 6 \\ 0 & ; otherwise \end{cases} \tag{5.5.5}$$

In order to obtain OSB under proportional allocation having the pdf's of the auxiliary variables defined in (5.5.4) and (5.5.5),we need to obtain the value of W_{hk} and σ_{hky}^2 for that we have to substitute (5.5.4) and (5.5.5) in equations (5.2.4) , (5.2.5) and (5.2.6),we get

$$W_{hk} = V_h e^{-z_k+1}\left(e^{U_k} - 1\right)\left(4 - V_h - 2x_{h-1}\right) \tag{5.5.6}$$

$$\sigma_{hkx}^2=\frac{a_1U_ke^{-z_k+1}}{\left(a_1e^{-z_k+1}\right)^2}\left\{\begin{array}{l}\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)\\-\frac{1}{2}\left[\begin{array}{l}\left(V_h+2x_{h-1}\right)\\\left[V_h\left(V_h+2x_{h-1}\right)+x_{h-1}\left(1+x_{h-1}\right)\right]\end{array}\right]\end{array}\right\}-4U_k^2 \qquad (5.5.7)$$

and $\sigma_{hkz}^2=\dfrac{a_1a_2-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{a_1^2}$ (5.5.8)

where $a_1=\left(e^{U_k}-1\right)\left(4-V_h-2x_{h-1}\right)$,

$$a_2=z_{k-1}^2e^{U_k}-U_k^2-z_{k-1}^2-U_kz_{k-1}+2\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]$$

Case I: for Independent case of auxiliary variables

Substituting the values obtained in (5.5.8) to (5.5.8) in equation (5.5.3), we have MPP as Minimize

$$\sum_h\sum_k\left(Sqrt\left(a_1V_he^{-z_k+1}\right)\right)$$

$$\left\{\beta^2\frac{a_1U_ke^{-z_k+1}}{6\left(a_1e^{-z_k+1}\right)^2}\left\{\left[8\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-3\left[\left(V_h+2x_{h-1}\right)\left[\begin{array}{l}V_h\left(V_h+2x_{h-1}\right)\\+x_{h-1}\left(1+x_{h-1}\right)\end{array}\right]\right]\right]-4U_k^2\right\}\right.$$

$$\left.+\gamma^2\frac{a_1a_2-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{a_1^2}\right\}$$

Subject to

$$\sum_h V_h=d_x$$
$$\sum_k U_k=t_z \qquad (5.5.9)$$

$$\forall V_h\geq 0,U_k\geq 0 \quad , \quad \begin{array}{l}h=1,2,...,L\\k=1,2,...,M\end{array}$$

By using the given pdf's a simulation has been done in R-software and the values of β=0.576 and γ = 0.257 have been obtained. Thus using values of β, γ , d_x =1 and t_z =5 as given above the defined interval for X and Z respectively for total 6 (2×3) strata. Thus (5.5.9) can be written as

Minimize

$$\sum_h\sum_k \left[Sqrt\left(a_1V_he^{-z_k+1}\right)\right]$$

$$\left\{(0.055)\frac{a_1U_ke^{-z_k+1}}{\left(a_1e^{-z_k+1}\right)^2}\left\{\left[\left[8\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-3\left[\left(V_h+2x_{h-1}\right)\begin{bmatrix}V_h\left(V_h+2x_{h-1}\right)\\+x_{h-1}\left(1+x_{h-1}\right)\end{bmatrix}\right]\right]\right]-4U_k^2\right\}\right.$$

$$\left.+(0.666)\frac{a_1a_2-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{a_1^2}\right\}$$

Subject to

$$\begin{aligned}&\sum_h V_h=1\\&\sum_k U_k=5\end{aligned}\qquad(5.5.10)$$

$$\forall V_h\geq 0, U_k\geq 0 \quad , \quad \begin{aligned}h&=1,2\\k&=1,2,3\end{aligned}$$

Executing a computer programme for the MPP (5.5.10) using LINGO software, we get the OSB as given in tables below:

Table 5.5.1: OSB when the auxiliary variables X and Z are independent having right triangular and exponential distribution respectively

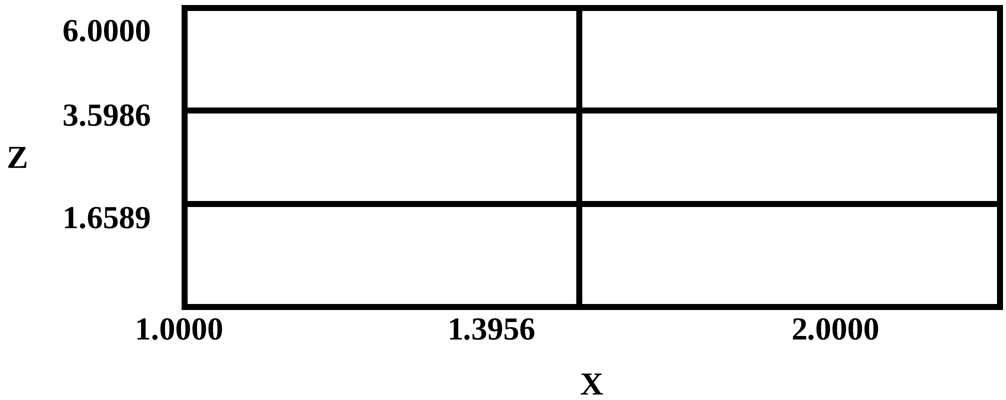


Table 5.5.2: OSB and Variance when the auxiliary variables X and Z are independent having right triangular and exponential distribution respectively

OSB (x_h, z_k)	Variance (Proposed method)	Variance (Thomson, 1973)	% R.E.
(1.3956,1.6568) (2.0000,1.6568) (1.3956,3.5986) (2.0000,3.5986) (1.3956,6.0000) (2.0000,6.0000)	0.000864	0.00412	476.85

Thus while making 2 strata along X variable and 3 along Z variable when the auxiliary variables X and Z are having Right triangular and Exponential distribution respectively independently. The results obtained in Table 5.5.1 and 5.5.2 reveal that the variance obtained by the proposed method is much less than Thomson (1973) for which the percent relative efficiency comes out to be 476.85.Thereby, it is revealed that use two auxiliary variables is better than using one auxiliary variable.

Case II: For dependent case of auxiliary variables

When the two auxiliary variables X and Z are dependent to each other then the MPP (5.5.10) would take the form as

Minimize

$$\sum_h \sum_k Sqrt\left(a_1 V_h e^{-z_k+1}\right)$$
$$\left\{ (0.055)\frac{a_1 U_k e^{-z_k+1}}{\left(a_1 e^{-z_k+1}\right)^2} \left\{ \left[8\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - 3\left[\left(V_h + 2x_{h-1}\right) \left[\begin{matrix} V_h\left(V_h + 2x_{h-1}\right) \\ + x_{h-1}\left(1 + x_{h-1}\right) \end{matrix} \right] \right] \right] - 4U_k^2 \right\} \right.$$
$$\left. + \frac{(0.666)}{a_1^2}\left\{ a_1 a_2 - e^{U_k}\left(1 + z_{k-1}\right) - U_k - z_{k-1} - 1 \right\} + 0.0736\rho b_1 b_2 \right\}$$

Subject to

$$\sum_h V_h = 1$$
$$\sum_k U_k = 5 \qquad (5.5.11)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2 \\ k = 1,2,3 \end{array}$$

where ρ , b_1 and b_2 are the correlation coefficient between X and Z and standard deviation of X and Z respectively which are given

$$a_2 = \left[\frac{a_1 U_k e^{-z_k+1}}{\left(a_1 e^{-z_k}+1\right)^2} \left\{ \begin{array}{l} \frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) \\ -\frac{1}{2}\left[\left(V_h + 2x_{h-1}\right)\left[V_h\left(V_h + 2x_{h-1}\right) + x_{h-1}\left(1 + x_{h-1}\right)\right]\right] \end{array} \right\} - 4U_k^2 \right]^{\frac{1}{2}}$$

$$a_3 = \left[\frac{a_1 a_2 - e^{U_k}\left(1 + z_{k-1}\right) - U_k - z_{k-1} - 1}{a_1^2} \right]^{\frac{1}{2}}$$

By simulating of data developed using pdf's, we get the value of ρ=0.42.By executing the programme in LINGO for obtaining OSB keeping all the values same as above, we get

Table 5.5.3: OSB and Variance when the auxiliary variables X and Z are dependent having right triangular and exponential distribution respectively

OSB $\left(x_h, z_k\right)$	**Variance**
(1.2568,1.5469)	0.004168
(2.0000,1.5469)	
(1.2568,3.0215)	
(2.0000,3.0215)	
(1.2568,6.0000)	
(2.0000,6.0000)	

The values in Tab.5.5.3 reveal that the strata boundaries and the variance are both different when the auxiliary variables are dependent. However; it shows that due to dependence the variance increases.

5.4.2: The log normal distribution is a positively skewed distribution. Surveyors may use the log normal distribution for a positive valued study variable that might increase without limit, such as the value of securities in financial problem or the values of properties in real estate or the failure rate of electronic parts in the engineering problems.

Let us assume that one of the auxiliary variable, say, X follows log-normal distribution with pdf as

$$f(x)=\begin{cases}\frac{1}{\sigma x\sqrt{2\pi}}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} & ; x>0, \sigma>0\\ 0 & ; otherwise\end{cases} \tag{5.4.12}$$

and the other auxiliary variable Z having pdf as:

$$f(z)=\begin{cases}\frac{1}{b-a} & , a\le z\le b\\ o & , otherwise\end{cases} \tag{5.5.13}$$

Then, in order to estimate OSB we need to find the value of W_{hk} and σ_{hky}^2.Substituting the pdf's (5.5.12) and (5.5.13) in equations (5.2.4) , (5.2.5) and (5.2.6),we shall get

$$W_{hk}=\frac{U_k}{2(b-a)}E_1 \tag{5.5.14}$$

$$\sigma_{hkx}^2=\frac{(b-a)\left[e^{2(\sigma^2+\mu)}(E_2)(E_1)\right]-U_k^2(b-a)^2\left[e^{\frac{1}{2}(\sigma^2-2\mu)}(E_3)\right]^2}{E_1^2} \tag{5.5.15}$$

and $$\sigma_{hkz}^2=\frac{2E_1\left(U_k^2+3z_{k-1}^2+3U_kz_{k-1}\right)-3V_h^2\left(U_k+2z_{k-1}\right)^2}{3E_1^2} \tag{5.5.16}$$

where $E_1=erf\left(\frac{\log(V_h+x_{h-1})-\mu}{\sqrt{2\sigma^2}}\right)-erf\left(\frac{\log(x_{h-1})-\mu}{\sqrt{2\sigma^2}}\right)$

$$E_2=erf\left(\frac{\log(V_h+x_{h-1})-\mu-2\sigma^2}{\sqrt{2\sigma^2}}\right)-erf\left(\frac{\log(x_{h-1})-\mu-2\sigma^2}{\sqrt{2\sigma^2}}\right)$$

and $E_3=erf\left(\frac{\log(V_h+x_{h-1})-\mu-\sigma^2}{\sqrt{2\sigma^2}}\right)-erf\left(\frac{\log(x_{h-1})-\mu-\sigma^2}{\sqrt{2\sigma^2}}\right)$

It is to be noted here that the function 'erf' which repeats many times in the above result is an error function which is used to counter the integration with log-normal density function. It is defined as

$$erf\left(\omega\right)=\frac{2}{\sqrt{\pi}}\int_{0}^{\omega}e^{-j^{2}}\partial j$$

and some of its properties that need to be noted are

$$\begin{aligned}
&erf\left(-\omega\right)=-erf\left(\omega\right)\\
&erf\left(0\right)=0\\
&erf\left(\infty\right)=1\\
&erf\left(-\infty\right)=1
\end{aligned}$$

Substituting values (5.5.14) to (5.5.16) in (5.5.3), we have MPP as

Minimize

$$\sum_{h}\sum_{k}\left[Sqrt\left(\frac{U_k}{2\left(b-a\right)}E_1\right)\right]\left(\begin{array}{l}\beta^2\dfrac{\left(b-a\right)\left[e^{2\left(\sigma^2+\mu\right)}\left(E_2\right)\left(E_1\right)\right]-U_k^2\left(b-a\right)^2\left[e^{\frac{1}{2}\left(\sigma^2-2\mu\right)}\left(E_3\right)\right]^2}{E_1^2}\\+\gamma^2\dfrac{2E_1\left(U_k^2+3z_{k-1}^2+3U_kz_{k-1}\right)-3V_h^2\left(U_k+2z_{k-1}\right)^2}{3E_1^2}\end{array}\right)$$

Subject to

$$\begin{aligned}
&\sum_{h}V_h=d_x\\
&\sum_{k}U_k=t_z
\end{aligned}\tag{5.5.17}$$

$$\forall V_h\geq 0,U_k\geq 0\quad,\quad \begin{array}{l}h=1,2,\ldots,L\\k=1,2,\ldots,M\end{array}$$

In this case let us assume that the log-normal distribution is to be standardize i,e μ=0, σ= 1. $z\in\left[0,1\right]i,e.z_0=0,z_M=1$ and the other variable $x\in\left[0,10\right],\ i,e.x_0=0,\ x_L=10$.Further let us assume that the total strata to be made are 3×2(L×M)=6 and by simulation in R-software the value of β=0.82 and γ=0.437.Then to obtain OSB we need to solve MPP

Minimize

$$\sum_{h=1}^{3}\sum_{k=1}^{2}\left[Sqrt\left(\frac{U_k}{2}E_1\right)\right]$$

$$\left(\begin{array}{l}(0.0722)\dfrac{\left[\left[(7.389)\left(E_2^{'}\right)\left(E_1^{'}\right)\right]-U_k^2\left[(1.648)\left(E_3^{'}\right)\right]^2\right]}{E_1^2}\\[2ex] +(0.1909)\dfrac{2E_1\left(U_k^2+3z_{k-1}^2+3U_k z_{k-1}\right)-3V_h^2\left(U_k+2z_{k-1}\right)^2}{3E_1^2}\end{array}\right)$$

Subject to

$$\begin{aligned}&\sum_{h=1}^{3}V_h=10\\ &\sum_{k=1}^{2}U_k=1\end{aligned}\qquad (5.5.18)$$

$$\forall V_h\geq 0, U_k\geq 0 \quad , \quad \begin{array}{l}h=1,2,3\\ k=1,2\end{array}$$

where $E_1^{'}=erf\left(\dfrac{\log\left(V_h+x_{h-1}\right)}{1.141}\right)-erf\left(\dfrac{\log\left(x_{h-1}\right)}{1.141}\right)$

$$E_2^{'}=erf\left(\frac{\log\left(V_h+x_{h-1}\right)-2}{1.141}\right)-erf\left(\frac{\log\left(x_{h-1}\right)-2}{1.141}\right)$$

and $E_3^{'}=erf\left(\dfrac{\log\left(V_h+x_{h-1}\right)-1}{1.414}\right)-erf\left(\dfrac{\log\left(x_{h-1}\right)-1}{1.414}\right)$

By executing the computer programme of the above MPP in LINGO, we get the OSB value presented in the following table

Table 5.5.4: OSB when the auxiliary variables X and Z follow log-normal and uniform distribution respectively

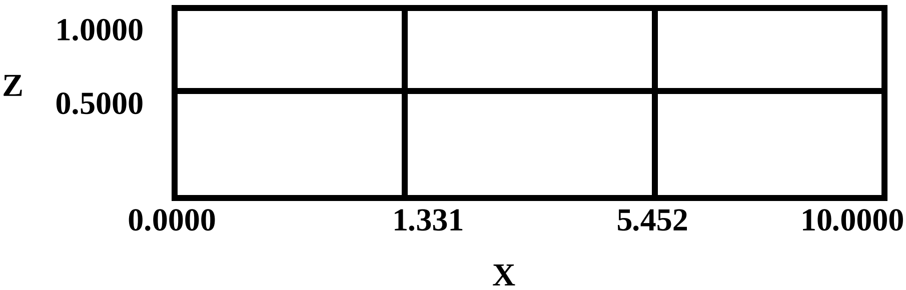


Table 5.5.5: OSB, and Variance when the auxiliary variables X and Z follow log-normal and uniform distribution respectively

OSB (x_h, z_k)	**Variance (Proposed method)**	**Variance (Khan *et al*. 2005)**	**Variance** (Cum $\sqrt[3]{D_4(x,z)}$ Rule)
(1.331,0.500) (5.452,0.500) (10.000,0.500) (1.331,1.000) (5.452,1.000) (10.000,1.000)	0.005916	0.014708	0.01025863

A perusal of Tab 5.5.5 indicates that the variance obtained by the proposed method is much less than not on Khan *et al*. (2005).Thus, it may be concluded that using two auxiliary variables is better than using one auxiliary variable. In practice, the complete dataset of the study variable is unknown, which diminishes the uses of many stratification techniques. In such a situation, the proposed technique can be used as it requires only the values of parameters of the population, which can easily be available from the past studies

5.6 Neyman allocation

This allocation of the total sample size to different strata is called minimum variance allocation and given by Neyman (1934).In this case, the allocation of samples among different strata is based on a joint consideration of the stratum size and the stratum variance. In this allocation, it is assumed that the sampling cost per unit among different strata is constant and size of the sample is fixed. Sample sizes are allocated by

$$n_{hk} = \frac{nW_{hk}\sigma_{hky}}{\sum\sum W_{hk}\sigma_{hky}}$$

A formula for the minimum variance with fixed 'n' is obtained by substituting the value of n_{hk}

$$\sum_h \sum_k \frac{(1-f)}{n_{hk}} W_{hk}^2 \sigma_{hky}^2$$

then we get

$$V\left(\bar{y}_{st}\right)=\frac{\left(\sum_{h}\sum_{k}W_{hk}\sigma_{hky}\right)^{2}}{n}-\frac{\sum_{h}\sum_{k}W_{hk}\sigma_{hky}^{2}}{N} \tag{5.6.1}$$

There may be difficulty in using this as the value of σ_{hky} will usually be unknown. However, the stratum variance may be obtained from previous surveys or from a specially planned pilot survey. The other alternative is to conduct the main survey in a phased manner and utilize the data collected in the first phase for ensuring better allocation in the second phase.

If the finite population correction is ignored, equation (5.6.1) can be written as

$$\frac{1}{n}\left(\sum_{h}\sum_{k}W_{hk}\sigma_{hky}\right)^{2}$$

However, minimising this is equivalent to minimize (since ‘n’ is fixed constant)

$$\sum_{h}\sum_{k}W_{hk}\sigma_{hky}$$

Using (5.2.3),it can be written as

$$\sum_{h}\sum_{k}W_{hk}\sqrt{\beta^{2}\sigma_{hkx}^{2}+\gamma^{2}\sigma_{hkz}^{2}} \tag{5.6.2}$$

Thus, under Neyman allocation for obtaining OSB we need to minimize the objective function (5.6.2) for which MPP is written as:

Minimize $\sum_{h}\sum_{k}W_{hk}\sqrt{\beta^{2}\sigma_{hkx}^{2}+\gamma^{2}\sigma_{hkz}^{2}}$

Subject to

$$\begin{aligned}&\sum_{h}V_{h}=d_{x}\\&\sum_{k}U_{k}=t_{z}\\&\forall V_{h}\geq 0,U_{k}\geq 0 \quad , \quad \begin{matrix}h=1,2,...,L\\ k=1,2,...,M\end{matrix}\end{aligned} \tag{5.6.3}$$

5.6.1: Empirical study

Let the two auxiliary variables X and Z follow uniform and exponential distributions respectively having the pdf’s as:

$$f\left(x\right)=\begin{cases}\dfrac{1}{b-a} & ,a\leq x\leq b\\ o & ,otherwise\end{cases} \tag{5.6.4}$$

and

$$f(z)=\begin{cases} e^{-z+1} & ;\quad z\ge 0 \\ 0 & ; otherwise \end{cases} \tag{5.6.5}$$

In order to minimize (5.6.3) substitute pdf's (5.6.4) and (5.6.5) in (5.2.4),(5.2.5) and (5.2.6), we get

$$W_{hk}=\frac{V_h}{b-a}e^{-z_k+1}\left[e^{U_k}-1\right] \tag{5.6.6}$$

$$\sigma_{hkx}^2=\frac{4U_k\left(V_h^2+3x_hx_{h-1}\right)\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^3-3U_k^2\left(V_h+2x_{h-1}\right)^2}{12\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^2} \tag{5.6.7}$$

$$\sigma_{hkz}^2=\frac{b-a}{\left(e^{U_k}-1\right)^2}\left\{f_6-\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]^2\right\} \tag{5.6.8}$$

where $f_6=z_{k-1}^2\left(e^{U_k}-1\right)^2-\left(e^{U_k}-1\right)\left[U_k^2+2e^{U_k}\left(1+z_{k-1}\right)-2\left(1+z_k\right)\right]$

Case I: For Independent auxiliary variables

Using (5.6.6)-(5.6.8) in MPP (5.6.3), we have

Minimize

$$\sum_h\sum_k\frac{V_h}{b-a}e^{-z_k+1}\left[e^{U_k}-1\right]\left\{\sqrt{\begin{array}{l}\beta^2\dfrac{4U_k\left(V_h^2+3x_hx_{h-1}\right)\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^3-3U_k^2\left(V_h+2x_{h-1}\right)^2}{12\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^2}\\ +\gamma^2\dfrac{b-a}{\left(e^{U_k}-1\right)^2}\left\{f_6-\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]^2\right\}\end{array}}\right\}$$

Subject to

$$\begin{aligned}&\sum_h V_h=d_x\\&\sum_k U_k=t_z\end{aligned} \tag{5.6.9}$$

$$\forall V_h\ge 0, U_k\ge 0 \quad , \quad \begin{array}{l}h=1,2,\ldots,L\\k=1,2,\ldots,M\end{array}$$

By assuming $a = 1 = x_0, b = 2$ and $1 \le z \le 6$,i,e. $d_x = 1$ and $t_z = 5$.Having these conditions over the variables, a simulation study has been made in R software to obtain value of β and γ i,e. β = 0.56 and γ =0.72.To obtain total 6 (2×3) strata substitute these values in MPP (5.6.9).Finally, we get MPP as follows:

Minimize

$$\sum_h \sum_k V_h e^{-z_k+1}\left[e^{U_k}-1\right] \left\{ \sqrt{ 0.002175 \frac{4U_k\left(V_h^2+3x_h x_{h-1}\right)\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^3 - 3U_k^2\left(V_h+2x_{h-1}\right)^2}{\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^2} + \frac{1.56}{\left(e^{U_k}-1\right)^2}\left\{ f_6 - \left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]^2\right\} } \right\}$$

Subject to

$$\begin{aligned} &\sum_h V_h = 1 \\ &\sum_k U_k = 5 \qquad\qquad (5.6.10) \\ &\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{matrix} h = 1,2 \\ k = 1,2,3 \end{matrix} \end{aligned}$$

By executing a computer programme in LINGO to the optimization problem (5.6.10), we get the OSB points. These results have been given Tables

Table 5.6.1: OSB when the auxiliary variables X and Z have uniform and exponential distribution respectively

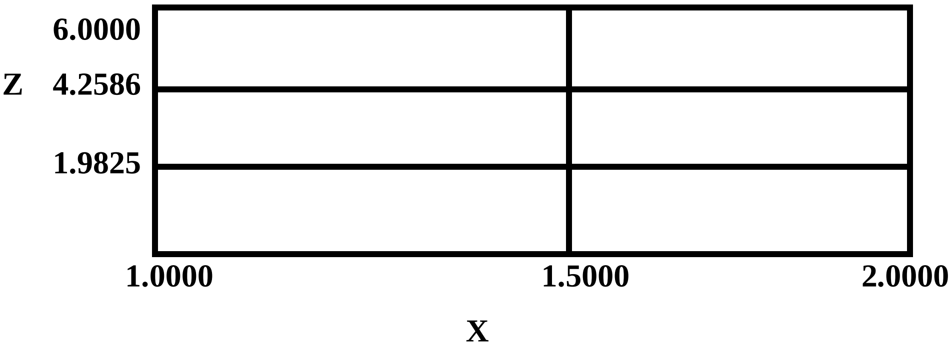


Table 5.6.2: OSB and Variance of proposed method and others when the auxiliary variables X and Z have uniform and exponential distribution respectively

OSB (x_h, z_k)	**Variance (Proposed method)**	**Variance (Singh and Sukhatme, 1969)**	**Variance (Khan *et al*. 2005)**
(1.5000,1.9825) (2.0000,1.9825) (1.5000,4.2586) (2.0000,4.2586) (1.5000,6.0000) (2.0000,6.0000)	0.009251	0.0740	0.02081

In order to obtain OSB for 2 strata along x variable and 3 along Z variable when X is uniformly and Z is exponentially distributed under Neyman allocation. The results so obtained in Table (5.6.1) and (5.6.2) reveals that variance obtained by the proposed method is much less than the variance obtained by Singh and Sukhatme (1969) and Khan *et al*. (2005).which concludes that the proposed method used for OSB is more preferable than the other methods.

Case II: For dependent auxiliary variables

Let ρ be the correlation coefficient between X and Z. Then for the case of dependence of X and Z having the same conditions and values as obtained above MPP (5.6.10) would take the form as

Minimize

$$\sum_h \sum_k V_h e^{-z_k+1}\left[e^{U_k}-1\right]$$

$$\left\{\sqrt{\begin{array}{l} 0.002175\dfrac{4U_k\left(V_h^2+3x_h x_{h-1}\right)\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^3-3U_k^2\left(V_h+2x_{h-1}\right)^2}{\left[e^{-z_k+1}\left[e^{U_k}-1\right]\right]^2} \\ +\dfrac{1.56}{\left(e^{U_k}-1\right)^2}\left\{f_6-\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]^2\right\} \\ +(1.4)(p_1)(p_2) \end{array}}\right\}$$

Subject to

$$\sum_{h} V_h = 1$$
$$\sum_{k} U_k = 5 \qquad (5.6.10)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2 \\ k = 1,2,3 \end{array}$$

where p_1 and p_2 are the standard deviation of X and Z respectively and are given as

$$p_1 = \sqrt{\frac{4U_k\left(V_h^2 + 3x_h x_{h-1}\right)\left[e^{-z_k+1}\left[e^{U_k} - 1\right]\right]^3 - 3U_k^2\left(V_h + 2x_{h-1}\right)^2}{12\left[e^{-z_k+1}\left[e^{U_k} - 1\right]\right]^2}}$$

$$p_2 = \sqrt{\frac{b-a}{\left(e^{U_k} - 1\right)^2}\left\{\begin{array}{l} z_{k-1}^2\left(e^{U_k} - 1\right)^2 - \left(e^{U_k} - 1\right)\left[U_k^2 + 2e^{U_k}\left(1 + z_{k-1}\right) - 2\left(1 + z_k\right)\right] \\ -\left[e^{U_k}\left(1 + z_{k-1}\right) - U_k - z_{k-1} - 1\right]^2 \end{array}\right\}}$$

By simulation in R using distribution functions of both the variables, the value of ρ=0.61.Executing the computer programme in LINGO keeping all the values same as above the OSB for the optimization problem (5.6.10) is obtained as

Table 5.6.3: OSB and Variance of proposed method and others having uniform and exponential distributed auxiliary variables when they are dependent

OSB (x_h, z_k)	**Total Variance (Proposed method)**	**Total Variance (Singh and Sukhatme, 1969)**	**Total Variance (Khan *et al*. 2005)**
(1.5000,1.5346) (2.0000,1.5346) (1.5000,3.9527) (2.0000,3.9527) (1.5000,6.0000) (2.0000,6.0000)	0.002562	0.0740	0.02081

However, when the two auxiliary variables are dependent the variance obtained by proposed method is still again less than other existing methods like Singh and Sukhatme (1969) and Khan *et al*. (2005).Thus we conclude that the proposed method is to be followed than the other methods when we are having such situations that follow the above conditions.

5.6.2: If the X variable has right triangular distribution with the pdf as

$$f\left(x\right)=\begin{cases}2(2-x) & ;0\le x\le 1\\ 0 & ;otherwise\end{cases} \tag{5.6.11}$$

and the variable Z follows Gamma Distribution with pdf as

$$f\left(z\right)=f\left(z,s,\theta\right)=\begin{cases}\dfrac{1}{\theta^{s}\Gamma\left(s\right)}z^{s-1}e^{-\frac{z}{\theta}} & ;\quad z\ge 0,s,\theta>0\\ 0 & ,otherwise\end{cases} \tag{5.6.12}$$

where'*s*' is the slope and 'θ' is the scale parameter and Γs is a gamma distribution function defined as

$$\Gamma s=\int_{0}^{\infty}e^{-z}z^{s-1}\partial z \qquad ;s>0$$

this function is also defined by the upper incomplete gamma function $\Gamma s,z$ and a lower incomplete gamma function γ(s,z) ,respectively as

$$\Gamma\left(s,z\right)=\int_{z}^{\infty}u^{s-1}e^{-u}\partial u$$

and
$$\gamma\left(s,z\right)=\int_{0}^{z}u^{s-1}e^{-u}\partial u$$

There also exists normalized incomplete gamma function which gives a value restricted between '0' and '1' that can be stated as

$$Q\left(s,z\right)=\frac{1}{\Gamma\left(s\right)}\int_{z}^{\infty}u^{s-1}e^{-u}\partial u \quad , \quad s,z>0$$

$$P\left(s,z\right)=\frac{1}{\Gamma\left(s\right)}\int_{0}^{z}u^{s-1}e^{-u}\partial u \quad , \quad s,z>0,\Gamma\left(s\right)\ne 0$$

Where Q(s, z) denotes the upper regularized incomplete gamma function with P(s,z) denotes regularized lower incomplete gamma**.**

In order to obtain OSB having distribution function of two auxiliary variables defined in (5.6.11) and (5.6.12) equations we have to find the value of W_{hk},σ_{hkx}^{2} and σ_{hkz}^{2} for that substitute above distribution functions in (5.2.4),(5.2.5) and (5.2.6),we have

$$W_{hk}=Q_{1}V_{h}\left(4-V_{h}-2x_{h-1}\right) \tag{5.6.13}$$

$$\sigma_{hkx}^2 = \frac{q_1 - U_k\left[2V_h\left(V_h + 2x_{h-1}\right) - \frac{2}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right)\right]}{Q_1^2 V_h^2\left(4 - V_h - 2x_{h-1}\right)^2} \tag{5.6.14}$$

$$\sigma_{hkz}^2 = \frac{\theta^2 s Q_1\left(s+1\right)\left[Q_2\right] - \theta^2 s^2\left(4 - V_h - 2x_{h-1}\right)\left[Q_3\right]^2}{Q_1^2\left(4 - V_h - 2x_{h-1}\right)} \tag{5.6.15}$$

where

$$q_1 = V_h U_k Q_1\left(4 - V_h - 2x_{h-1}\right)$$
$$\left\{\frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[\left(V_h + 2x_{h-1}\right)\left(V_h^2 + 2x_{h-1}^2 + 2V_h x_{h-1}\right)\right]\right\}$$

$$Q_1 = Q\left(s, \frac{z_{k-1}}{\theta}\right) - Q\left(s, \frac{z_k}{\theta}\right) \quad , \quad Q_2 = Q\left(s+2, \frac{z_{k-1}}{\theta}\right) - Q\left(s+2, \frac{z_k}{\theta}\right)$$

and $Q_3 = Q\left(s+1, \frac{z_{k-1}}{\theta}\right) - Q\left(s+1, \frac{z_{k-1} + U_k}{\theta}\right)$

By substituting values obtained in equations (5.6.13) to (5.6.15) in (5.6.3), we have

Minimize

$$\sum_h \sum_k Q_1 V_h\left(4 - V_h - 2x_{h-1}\right)\sqrt{\begin{array}{l}\beta^2 \dfrac{q_1 - U_k\left[2V_h\left(V_h + 2x_{h-1}\right) - \frac{2}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right)\right]}{Q_1^2 V_h^2\left(4 - V_h - 2x_{h-1}\right)^2} \\ +\gamma^2 \dfrac{\theta^2 s Q_1\left(s+1\right)\left[Q_2\right] - \theta^2 s^2\left(4 - V_h - 2x_{h-1}\right)\left[Q_3\right]^2}{Q_1^2\left(4 - V_h - 2x_{h-1}\right)}\end{array}}$$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \end{aligned} \tag{5.6.16}$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, ..., L \\ k = 1, 2, ..., M \end{array}$$

For computation details let us define the interval of both variables as $x \in [0,1]$ and $z \in [0,6]$ by generating the data in R-software using the pdf's (5.6.11) and (5.6.12) to estimate β=0.81 and γ= 0.72.Also the minimum likelihood estimates of the parameters of the gamma distribution while compiling generated data of distribution are s=4.3829 and θ=3.0124.By substituting these values in MPP (5.6.16), we get

Minimize

$$\sum_h \sum_k Q_1' V_h \left(4 - V_h - 2x_{h-1}\right) \sqrt{\begin{array}{l} (0.6561) \dfrac{q_1' - U_k \left[2V_h \left(V_h + 2x_{h-1}\right) - \dfrac{2}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right)\right]}{Q_1'^2 V_h^2 \left(4 - V_h - 2x_{h-1}\right)^2} \\ +(161.8539) \dfrac{Q_1' \left[Q_2'\right] - \theta^2 s^2 \left(4 - V_h - 2x_{h-1}\right)\left[Q_3'\right]^2}{Q_1'^2 \left(4 - V_h - 2x_{h-1}\right)} \end{array}}$$

Subject to

$$\begin{aligned} &\sum_h V_h = 1 \\ &\sum_k U_k = 6 \end{aligned} \qquad (5.6.17)$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, ..., L \\ k = 1, 2, ..., M \end{array}$$

where

$$\begin{aligned} q_1' = {} & V_h U_k Q_1' \left(4 - V_h - 2x_{h-1}\right) \\ & \left\{ \frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[\left(V_h + 2x_{h-1}\right)\left(V_h^2 + 2x_{h-1}^2 + 2V_h x_{h-1}\right)\right]\right\} \end{aligned}$$

$$Q_1' = Q\left(4.3829, \frac{z_{k-1}}{3.0124}\right) - Q\left(4.3829, \frac{z_k}{3.0124}\right), Q_2' = Q\left(6.3829, \frac{z_{k-1}}{3.0124}\right) - Q\left(6.3829, \frac{z_k}{3.0124}\right) \quad ,$$

and $Q_3' = Q\left(5.3829, \dfrac{z_{k-1}}{3.0124}\right) - Q\left(5.3829, \dfrac{z_{k-1} + U_k}{3.0124}\right)$

The MPP (5.6.17) is executed in the LINGO with total target of 6 (2×3) strata in which 2 are to be made along X variable and 3 along Z variable. The OSB so obtained is shown in the following table:

Table 5.6.4: OSB and Variance of proposed method when the auxiliary variables X and Z have right-triangular and gamma distribution respectively

OSB (x_h, z_k)	Variance (Proposed method)	Variance (Reddy *et al.* 2016)	% R.E.
(0.4231,1.9237) (1.0000,1.9237) (0.4231,3.6829) (1.0000,3.6829) (0.4231,6.0000) (1.0000,6.0000)	0.004869	0.0079251	162.77

The above table shows us the OSB under Neyman allocation when one of the auxiliary variables is having Right triangular distribution and the other is having Gamma distribution.

5.6.3: Let the variable X follows uniform distribution with pdf as

$$f(x) = \begin{cases} \dfrac{1}{b-a} & , a \le x \le b \\ o & , otherwise \end{cases} \tag{5.6.18}$$

and the other variable Z follows an exponential distribution with pdf as

$$f(z) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{z}{\theta}} & ; \quad z \ge 0 \\ 0 & ; \quad otherwise \end{cases} \tag{5.6.19}$$

In order to find OSB when the frequency distributions of X and Z are uniform and exponential having pdf as given in equations (5.6.18) and (5.6.19) respectively. We need to find the value of W_{hk} ,σ^2_{hkx} and σ^2_{hkz} for that substitute value of f(x) and f(z) in equations (5.2.4) to (5.2.6),we have

$$W_{hk} = \frac{V_h}{(b-a)} e^{-\frac{z_{k-1}}{\theta}} \left(1 - e^{-\frac{U_k}{\theta}} \right) \tag{5.6.20}$$

$$\sigma_{hkx}^2=\frac{2e^{-\frac{z_{k-1}}{\theta}}\left(1-e^{-\frac{U_k}{\theta}}\right)U_k\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-3U_k^2\left(V_h+2x_{h-1}\right)^2}{6\left[e^{-\frac{z_{k-1}}{\theta}}\left(1-e^{-\frac{U_k}{\theta}}\right)\right]^2} \quad (5.6.21)$$

$$\sigma_{hkz}^2=\frac{(b-a)\left(1-e^{-\frac{U_k}{\theta}}\right)p_3}{\left(1-e^{-\frac{U_k}{\theta}}\right)^2} \quad (5.6.22)$$

where

$$p_3=\left\{z_{k-1}^2-\left(U_k^2+z_{k-1}^2+2U_kz_{k-1}\right)e^{-\frac{U_k}{\theta}}\right\}-V_h^2\left[(\theta+z_{k-1})\left(1-e^{-\frac{U_k}{\theta}}\right)-U_ke^{-\frac{U_k}{\theta}}\right]^2$$

substitute (5.6.20)-(5.6.22) in (5.6.3),we have

Minimize

$$\sum_h\sum_k\frac{V_h}{(b-a)}a_1\left[\begin{array}{l}\beta^2\dfrac{2a_1U_k\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-3U_k^2\left(V_h+2x_{h-1}\right)^2}{6\left[e^{-\frac{z_{k-1}}{\theta}}\left(1-e^{-\frac{U_k}{\theta}}\right)\right]^2}\\ +\gamma^2\left[\dfrac{(b-a)\left(1-e^{-\frac{U_k}{\theta}}\right)p_3}{\left(1-e^{-\frac{U_k}{\theta}}\right)^2}\right]\end{array}\right]^{\frac{1}{2}}$$

Subject to

$$\sum_h V_h = d_x$$
$$\sum_k U_k = t_z \qquad (5.6.23)$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2,\ldots,L \\ k = 1,2,\ldots,M \end{array}$$

where $a_1 = e^{-\frac{z_{k-1}}{\theta}}\left(1-e^{-\frac{U_k}{\theta}}\right)$

Let us assume that the X is defined in the interval as in $\left[0,1\right]$ i,e. $x_0 = a = 0, x_L = b = 1$ and the Z defined in $\left[0,5\right]$ and θ = 1,such that $d_x = 1$ and $t_z = 5$.By simulation in R-software we get values of β = 0.58 and γ = 0.73.substituting these values in (5.6.23),we get the problem as follows:

Minimize

$$\sum_h \sum_k V_h e^{-z_{k-1}}\left(1-e^{-U_k}\right)$$

$$\left[(0.3364)\frac{2e^{-z_{k-1}}\left(1-e^{-U_k}\right)U_k\left(V_h^2+3x_{h-1}^2+3V_h x_{h-1}\right)-3U_k^2\left(V_h+2x_{h-1}\right)^2}{6\left[e^{-z_{k-1}}\left(1-e^{-U_k}\right)\right]^2} + (0.5324)\left[\frac{\left(1-e^{-U_k}\right)q_3'}{\left(1-e^{-U_k}\right)^2}\right] \right]^{\frac{1}{2}}$$

Subject to

$$\sum_h V_h = 1$$
$$\sum_k U_k = 5$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2,\ldots,L \\ k = 1,2,\ldots,M \end{array}$$

where

$$q_3' = \left\{ z_{k-1}^2 - \left(U_k^2 + z_{k-1}^2 + 2U_k z_{k-1} \right) e^{-U_k} \right\} - V_h^2 \left[\left(1 + z_{k-1}\right)\left(1 - e^{-U_k}\right) - U_k e^{-U_k} \right]^2$$

Executing a computer programme of this MPP in LINGO for 6(2×3) strata, i.e., 2 along X variable and 3 along Z variable. The results obtained are presented in Table 5.6.6 as follows.

Table 5.6.5: OSB and Variance of proposed method and other existing methods when the auxiliary variables X and Z follow uniform and exponential distributions, respectively.

OSB $\left(x_h, z_k\right)$	**Variance (Proposed method)**	**Variance (Khan *et al.* 2005)**	**Variance (Khan *et al.* 2014)**
(0.5000, 0.7528) (0.5000, 3.0451) (0.5000,5.0000) (1.0000,0.7528) (1.0000,3.0451) (1.0000,5.0000)	0.005291	0.0725	0.0539

The above table reveals that the variance obtained by the proposed method is much less than the variance obtained by Khan *et al*. (2005) as well as and Khan *et al*. (2014).This suggests us the superiority of the proposed method over other method.

5.6.4: Let the variable X is having a uniform distribution with the form as

$$f(x) = \begin{cases} \dfrac{1}{s-r} & , r \le x \le s \\ o & , otherwise \end{cases} \tag{5.6.24}$$

and the variable Z follows a triangular distribution with pdf as

$$f(z) = \begin{cases} \dfrac{2(z-p)}{(q-p)(c-p)} & ; \quad a \le z \le c \\ \dfrac{2(q-z)}{(q-p)(q-c)} & ; \quad c < z \le b \end{cases} \tag{5.6.25}$$

where p is the location parameter ,q is the scale parameter and c is the shape parameter. Since our aim is to find the stratification points at which the variance is minimum. In order to obtain them having the pdf's as given above we need to find the value of W_{hk} , σ^2_{hkx} and σ^2_{hkz} for that substitute value of f(x) and f(z) in equations (5.2.4) to (5.2.6),we have

$$W_{hk} = \frac{V_h U_k}{(q-p)(s-r)(c-p)(q-c)}[G_1] \tag{5.6.26}$$

$$\sigma^2_{hkx} = \frac{(q-p)(c-p)(q-c)}{4(G_1)^2}[G_4] \tag{5.6.27}$$

$$\sigma^2_{hkz} = \frac{(s-r)}{(G_1)^2}\left\{[G_2](G_1)-(s-r)(G_3)^2\right\} \tag{5.6.28}$$

where

$$G_1 = (c-p)(U_k + 2z_{k-1} - 2p) + (q-c)(2b - 2z_{k-1} - U_k)$$

$$G_2 = \frac{(U_k - 2z_{k-1})\left(U_k^2 + 2z_{k-1}^2 + 2U_k z_{k-1}\right)(q-2c+p)}{2} - \frac{2}{3}G_5$$

$$G_3 = \left[(q-c)\left(\frac{2}{3}U_k^2 + 2U_k z_{k-1} - rU_k + 2z_{k-1}^2 + 2rz_{k-1}\right) + (c-p)G_6\right]$$

$$G_4 = 4(G_1)\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - (q-p)(c-p)(q-c)(V_h + 2x_{h-1})^2$$

$$G_5 = \left(U_k^2 + 3z_{k-1}^2 + 3U_k z_{k-1}\right)\left(q(c-p-r) + rc\right)$$

$$G_6 = \left(sU_k - 2U_k z_{k-1} + sz_{k-1} - z_{k-1}^2 - \frac{2}{3}U_k^2\right)$$

Substituting the values obtained in (5.6.26) to (5.6.28) in MPP (6.4.3),we get

Minimize

$$\sum_h \sum_k \frac{V_h U_k}{(q-p)(s-r)(c-p)(q-c)}[G_1] \sqrt{\begin{array}{l} \beta^2 \dfrac{(q-p)(c-p)(q-c)}{4(G_1)^2}[G_4] \\ +\gamma^2 \dfrac{(s-r)}{(G_1)^2}\left\{[G_2](G_1)-(s-r)(G_3)^2\right\} \end{array}}$$

Subject to

$$\sum_h V_h = d_x$$
$$\sum_k U_k = t_z \qquad (5.6.29)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, ..., L \\ k = 1, 2, ..., M \end{array}$$

However, we assume that the variable X is defined in the interval [0,1] and z in [0,2] i,e. $x_0 = r = 0, x_L = s = 1, p = z_0 = o, c = 1, q = z_M = 2, a = 0, b = 2$ and by simulation done in R for the distribution function the estimate comes as β = 0.64 and γ = 0.56.By substituting all these values in MPP (5.6.29),we have

Minimize

$$\sum_h \sum_k V_h U_k \left[G_1'\right] \sqrt{\frac{0.8192}{\left(G_1'\right)^2}\left[G_4'\right] + (0.3136)\frac{1}{\left(G_1'\right)^2}\left\{\left[G_2'\right]\left[G_1'\right] - \left(G_3'\right)^2\right\}}$$

Subject to

$$\sum_h V_h = 1$$
$$\sum_k U_k = 2 \qquad (5.6.29)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, ..., L \\ k = 1, 2, ..., M \end{array}$$

where

$$G_1' = \left(U_k + 2z_{k-1}\right) + \left(2 - 2z_{k-1} - U_k\right)$$

$$G_2' = -\frac{2}{3}\left(G_5'\right)$$

$$G_3' = \left[\left(\frac{2}{3}U_k^2 + 2U_k z_{k-1} + 2z_{k-1}^2\right) + G_6'\right], G_4' = 4\left(G_1'\right)\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \left(V_h + 2x_{h-1}\right)^2$$

$$G_5' = 2\left(U_k^2 + 3z_{k-1}^2 + 3U_k z_{k-1}\right), G_6' = \left(U_k - 2U_k z_{k-1} + z_{k-1} - z_{k-1}^2 - \frac{2}{3}U_k^2\right)$$

For obtaining total 6 (2×3) strata execute a computer programme in LINGO, we get

Table 5.6.6: OSB when the auxiliary variables X and Z have uniform and right triangular distribution respectively

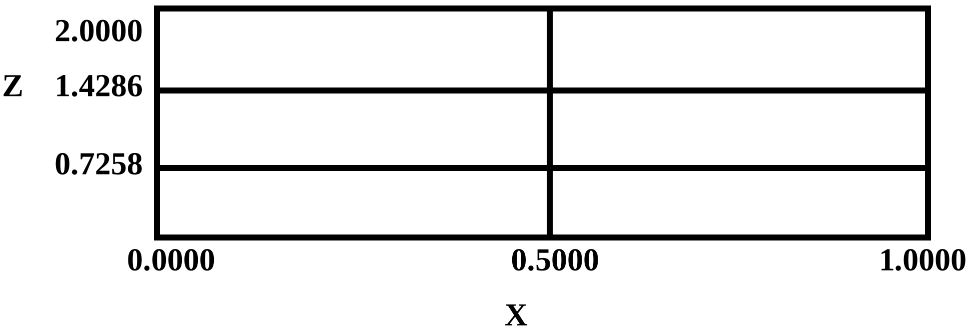


Table 5.6.7: OSB and Variance of proposed method having distribution of X and Z as uniform and right triangular distributions, respectively.

OSB (x_h, z_k)	**Variance (Proposed method)**	**Variance (Khan *et al.* 2008,2014)**	**% R.E.**
(0.5000,1.7258) (1.0000,1.7258) (0.5000,1.4286) (1.0000,1.4286) (0.5000,2.0000) (1.0000,2.0000)	0.0027925	0.009294	332.82

Table 5.6.8 shows the stratification points under Neyman allocation when the auxiliary variables X and Z are Uniformly and Right triangularly distributed defined in the interval of [0,1] and [0,2].Table 5.6.9 shows us the boundary points of the strata when 2 strata are to be made along X variable and 3 along Z variable and also it reveals that the variance obtained by the proposed method is less than the method obtains by Khan *et al.* (2008, 2014).Thus we can conclude that the proposed method of obtaining stratification points is better than the existing methods.

5.7 Optimum allocation

In stratified sampling, variance of the estimator depends on values of n_{hk} (h = 1,2,...,L ;1,2,3,...,M) apart from values of Y. Even for a fixed n, the values of the $V\left(\bar{y}_{st}\right)$ will differ for different configurations $\left(n = n_{11},...,n_{LM}\right)$.The combination 'n' for which $V\left(\bar{y}_{st}\right)$ is minimum among all the values of variances for different possible combinations 'n' is an optimal allocation for a fixed n. Since sampling involves cost in a survey , consider the linear cost function

$$C = C_0 + \sum_h \sum_k C_{hk} n_{hk} \tag{5.7.1}$$

where C_0 is an overhead cost (e.g cost of setting up and maintaining an office ,recruiting survey personal and other capital expanses etc.) C_{hk} is the cost of sampling unit from the $(h,k)^{th}$ stratum and C is the total cost.(5.7.1) equation is a very simple and reasonable cost function expressing the total cost of field operation. The variance under stratified random sampling is

$$V\left(\bar{y}_{st}\right) = \sum_h \sum_k \frac{W_{hk}^2 \sigma_{hky}^2}{n_{hk}} - \sum_h \sum_k \frac{W_{hk} \sigma_{hky}^2}{N_{hk}} \tag{5.7.2}$$

To determine the value of n_{hk} ,we consider the function

$$\psi = V\left(\bar{y}_{st}\right) + \lambda C \tag{5.7.3}$$

where λ is the known constant.

Using the calculus method of Langrangian multiplier, we select n_{hk} and a constant λ to minimize ψ .

Differentiating (5.7.3) w.r.t. n_{hk} ,we have

$$-\frac{W_{hk}^2 \sigma_{hky}^2}{n_{hk}^2} + \lambda c_{hk} = 0 \qquad \begin{array}{l} , h = 1, 2, \ldots, L \\ K = 1, 2, \ldots, M \end{array}$$

$$\therefore n_{hk} = \frac{W_{hk} \sigma_{hky}}{\sqrt{\lambda c_{hk}}} \tag{5.7.4}$$

while taking summation on both sides ,we get

$$n = \sum_h \sum_k \left(\frac{W_{hk} \sigma_{hky}}{\sqrt{\lambda c_{hk}}} \right) \tag{5.7.5}$$

From (5.7.4) and (5.7.5), we can obtain

$$n_{hk} = n \frac{W_{hk} \sigma_{hky} / \sqrt{c_{hk}}}{\sum_h \sum_k \left(W_{hk} \sigma_{hky} / \sqrt{c_{hk}} \right)} \tag{5.7.6}$$

Thus (5.7.6) relation leads to the following important conclusion that, in a given strata, we have to take a large sample size if

i. the stratum size is large.
ii. the stratum has large variability.
iii. the cost per unit is cheaper in the stratum.

If c_{hk}'s are the same from stratum to stratum, relation (5.7.7) will lead to Neyman allocation. Similarly if c_{hk}'s and σ_{hky}'s do not vary from stratum to stratum, relation (5.7.6) leads to proportional allocation.

The sample size 'n_{hk}' from $(h,k)^{th}$ stratum required for estimating the population with a specified cost 'C' is given by

$$n_{hk} = \frac{(C-c_0)\, W_{hk}\sigma_{hky} / \sqrt{c_{hk}}}{\sum_h \sum_k \left(W_{hk}\sigma_{hky} / \sqrt{c_{hk}} \right)} \tag{5.7.7}$$

substituting this value in (5.7.2),we get

$$V\left(\bar{y}_{st}\right) = \frac{\left(\sum_h \sum_k W_{hk}\sigma_{hky}\sqrt{c_{hk}} \right)^2}{(C-c_0)} - \sum_h \sum_k \frac{W_{hk}\sigma_{hky}^2}{N_{hk}} \tag{5.7.8}$$

If the finite population correction is ignored, then minimizing the expression on right hand side of (5.7.8) is the same as minimizing

$$\sum_h \sum_k W_{hk}\sigma_{hky}\sqrt{c_{hk}} \tag{5.7.9}$$

Thus in order to obtain OSB in this allocation type , objective function that we need to optimize (5.7.9) and the constraints are same as (5.1.15) and (5.1.16).Thus to obtain OSB the MPP that we have to minimize is

Minimize $\sum_h \sum_k W_{hk}\sigma_{hky}^2\sqrt{c_{hk}}$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \end{aligned} \tag{5.7.10}$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2,\ldots,L \\ k = 1,2,\ldots,M \end{array}$$

However, if the regression line is linear in two variables X and Z then σ^2_{hky} is to be replaced by (5.2.3) as

$$\begin{aligned} &\text{Minimize } \sum_h \sum_k W_{hk}\left(\beta^2\sigma^2_{hkx} + \gamma^2\sigma^2_{hkz}\right)\sqrt{c_{hk}} \\ &\text{Subject to} \\ &\qquad \sum_h V_h = d_x \\ &\qquad \sum_k U_k = t_z \qquad\qquad (5.7.11) \\ &\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{matrix} h = 1,2,\ldots,L \\ k = 1,2,\ldots,M \end{matrix} \end{aligned}$$

5.7.1: Empirical study

If the variable is having right triangular distribution with pdf as

$$f(x) = \begin{cases} 2(2-x) & ; \quad 0 \le x \le 1 \\ 0 & ; otherwise \end{cases} \tag{5.7.12}$$

and the other variable follows an exponential distribution as

$$f(z) = \begin{cases} \lambda e^{-\lambda z} & ; \quad z \ge 0, \lambda > 0 \\ 0 & ; otherwise \end{cases} \tag{5.7.13}$$

In order to obtain OSB when the auxiliary variables have as given above we need to estimate the values of W_{hk}, σ^2_{hkx} and σ^2_{hky} .For estimating this use above pdf's in (5.2.4) to (5.2.6),we get

$$W_{hk} = V_h g_1 \tag{5.7.14}$$

$$\sigma^2_{hkx} = \frac{U_k g_1\left\{\frac{4}{3}(f_7) - \frac{1}{2}[f_8]\right\} - p_4}{g_1^2} \tag{5.7.15}$$

$$\sigma^2_{hkz} = \frac{f_8\left\{z^2_{k-1} - (U_k + z_{k-1})^2 e^{-\lambda U_k} + 2\lambda f_9\right\} - \left[\left(z_{k-1} + \frac{1}{\lambda}\right)\left(1 - e^{-\lambda U_k}\right) - U_k e^{-\lambda U_k}\right]^2}{(f_8)^2}$$

(5.7.16)

where $g_1 = e^{-\lambda z_{k-1}}\left(1-e^{-\lambda U_k}\right)\left(4-V_h-2x_{h-1}\right)$ $p_4 = 4U_k^2\left[V_h\left(1-\frac{1}{3}V_h\right)+x_{h-1}\left(2-x_{h-1}-3V_h\right)\right]^2$

$f_6 = V_h^2+3x_{h-1}^2+3V_hx_{h-1},\ f_7 = \left(V_h+2x_{h-1}\right)\left(V_h^2+2x_{h-1}^2+2V_hx_{h-1}\right)$

$f_8 = \left(1-e^{-\lambda U_k}\right)\left(4-V_h-2x_{h-1}\right),\ f_9 = \left[z_{k-1}+\left(U_k+z_{k-1}\right)e^{-\lambda U_k}\right]+2\left(1-e^{-\lambda U_k}\right)$

Case I: For independent auxiliary variables

Using values obtained in equations (5.7.14) to (5.7.16) in MPP (5.7.11), we have Minimize

$$\sum_h\sum_k V_h g_1 \left(\begin{array}{l} \beta^2 \dfrac{U_k g_1\left\{\frac{4}{3}(f_6)-\frac{1}{2}[f_7]\right\}-p_4}{g_1^2} \\ +\gamma^2 \dfrac{z_{k-1}^2-\left(U_k+z_{k-1}\right)^2 e^{-\lambda U_k}+2\lambda f_9-\left[\left(z_{k-1}+\frac{1}{\lambda}\right)\left(1-e^{-\lambda U_k}\right)-U_k e^{-\lambda U_k}\right]^2}{(f_8)^2} \end{array}\right)\sqrt{c_{hk}}$$

Subject to

$$\begin{gathered} \sum_h V_h = d_x \\ \sum_k U_k = t_z \\ \forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{array}{l} h=1,2,...,L \\ k=1,2,...,M \end{array} \end{gathered} \tag{5.7.17}$$

By executing a programme in R-software relating to generating random numbers for estimating the parameters we get β = 0.52 and γ = 0.42.

Now let us assume that the auxiliary variable X is defined in $[0,1]$ and Z in $[0,6]$ i,e. $x_0 = 0, x_L = 1, z_0 = 0, z_M = 6$ and λ = 1.we can write MPP (5.7.17) as

Minimize

$$\sum_h\sum_k V_h g_1' \left(\begin{array}{l} (0.27) \dfrac{U_k g_1'\left\{\frac{4}{3}f_6-\frac{1}{2}f_7\right\}-p_4}{g_1^2} \\ +(0.17) \dfrac{z_{k-1}^2-\left(U_k+z_{k-1}\right)^2 e^{-U_k}+2f_9'-\left[\left(z_{k-1}+1\right)\left(1-e^{-U_k}\right)-U_k e^{-U_k}\right]^2}{(f_8)^2} \end{array}\right)\sqrt{c_{hk}}$$

Subject to

$$\begin{aligned} &\sum_h V_h = 1 \\ &\sum_k U_k = 6 \\ \forall V_h \geq 0, U_k \geq 0 \quad , \quad &h = 1,2,\ldots,L \\ &k = 1,2,\ldots,M \end{aligned} \tag{5.7.18}$$

Where $g_1' = e^{-z_{k-1}}\left(1-e^{-U_k}\right)\left(4-V_h-2x_{h-1}\right)$ $f_9' = \left[z_{k-1}+\left(U_k+z_{k-1}\right)e^{-U_k}\right]+2\left(1-e^{-U_k}\right)$

while executing a computer programme of the MPP (5.7.18) for obtaining OSB of 6 (2×3) strata assuming the cost values as $c_{11}=2, c_{12}=3, c_{13}=4, c_{21}=5, c_{22}=6, c_{23}=7$,we get OSB as presented in the following tables:

Table 5.7.1: OSB when the auxiliary variables X and Z having right triangular and exponential distribution respectively

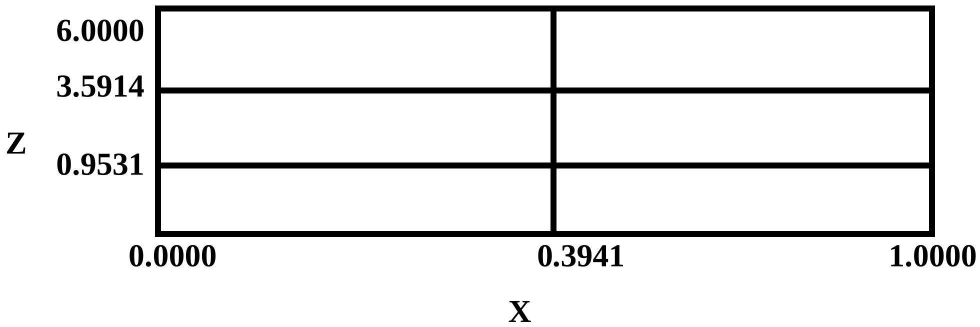


Table 5.7.2: OSB and Variance of proposed method and others when the auxiliary variables X and Z are having right triangular and exponential distributions, respectively

OSB (x_h, z_k)	**Variance (Proposed method)**	**Variance (Fonolahi and Khan, 2014)**	**% R.E.**
(0.3941,0.9531) (2.0000,0.9531) (0.3941,3.5914) (2.0000,3.5914) (0.3941,6.0000) (2.0000,6.0000)	0.03519	0.04348	321.62

Table 5.7.1 shows the points of stratification when 2 strata are to be made along X variable and 3 along Z variable under optimum allocation while as Table 5.7.2 shows the optimum stratification points $\left(x_h, z_k\right)$ for each stratum. It also shows the variance obtained through the proposed method and by Fonolahi and Khan (2014). The percent relative efficiency reveals that the proposed method is more preferable rather than the method proposed by Fonolahi and Khan (2014).

Case II: For dependent auxiliary variables

However, when the two auxiliary variables are dependent to each other then let **ρ** be the correlation coefficient. By calculating correlation coefficient of the data obtained by simulation we get **ρ**=0.51.Thus keeping all the conditions same as taken above, we can write MPP (5.7.18) as

Minimize

$$\sum_h \sum_k V_h g_1' \left(\begin{array}{l} (2.31) \dfrac{U_k g_1' \left\{ \dfrac{4}{3} f_6 - \dfrac{1}{2} f_7 \right\} - p_4}{g_1^2} + (0.222768) g_3 g_4 \\ + (0.17) \dfrac{z_{k-1}^2 - \left(U_k + z_{k-1}\right)^2 e^{-U_k} + 2 f_9' - \left[\left(z_{k-1}+1\right)\left(1-e^{-U_k}\right) - U_k e^{-U_k} \right]^2}{\left(f_8\right)^2} \end{array} \right) \sqrt{c_{hk}}$$

Subject to

$$\begin{aligned} &\sum_h V_h = 1 \\ &\sum_k U_k = 6 \end{aligned} \qquad (5.7.19)$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, ..., L \\ k = 1, 2, ..., M \end{array}$$

Where $g_1' = e^{-z_{k-1}} \left(1 - e^{-U_k}\right)\left(4 - V_h - 2x_{h-1}\right)$

g_3 and g_4 denotes the standard deviation of X and Z and are given as

$$g_3 = \sqrt{\frac{U_k g_1 \left\{ \dfrac{4}{3} f_6 - \dfrac{1}{2} f_7 \right\} - p_4}{g_1^2}}$$

and

$$g_4 = \frac{1}{f_8}\sqrt{f_8\left\{z_{k-1}^2 - \left(U_k + z_{k-1}\right)^2 e^{-\lambda U_k} + 2\lambda f_9\right\} - \left[\left(z_{k-1} + \frac{1}{\lambda}\right)\left(1 - e^{-\lambda U_k}\right) - U_k e^{-\lambda U_k}\right]^2}$$

Executing a computer programme to the MPP given in equation (5.7.19), we get the OSB as

Table 5.7.3: OSB and Variance of proposed method when the auxiliary variables X and Z are having right triangular and exponential distribution respectively

OSB $\left(x_h, z_k\right)$	**Total Variance (Proposed method)**
(0.2946,0.8249)	0.00924619
(2.0000,0.8249)	
(0.2946,4.2759)	
(2.0000,4.2759)	
(0.2946,6.0000)	
(2.0000,6.0000)	

Thus while in case of dependence of the two auxiliary variables increase the total variance but not as much as we received from the existing methods.

5.7.2: Let us assume that the auxiliary variable X is having a uniform distribution with pdf as

$$f(x) = \begin{cases} \dfrac{1}{b-a} & , a \le x \le b \\ o & , otherwise \end{cases} \tag{5.7.20}$$

and assumed that the other variable Z is following standard normal distribution with pdf as

$$f(z) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} & ; -\infty < z < \infty \\ 0 & ; otherwise \end{cases} \tag{5.7.21}$$

To obtain OSB for the study variable which is having two auxiliary variables following uniform and standard normal distributions, we need to find the values of W_{hk}, σ_{hkx}^2 and σ_{hky}^2 .For estimating this use above pdf's in (5.2.4),(5.2.5) and (5.2.6),we get

$$W_{hk}=\frac{V_h}{2(b-a)}(E_1) \quad (5.7.22)$$

$$\sigma_{hkx}^2=\frac{2U_kE_1\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-12U_k^2\left(V_h+2x_{h-1}\right)^2}{3(E_1)^2} \quad (5.7.23)$$

$$\sigma_{hkz}^2=2\sqrt{2}(b-a)\begin{bmatrix} z_{k-1}\exp\left(-\frac{z_{k-1}^2}{2}\right)E_2-(U_k+z_{k-1})\exp\left(-\frac{(U_k+z_{k-1})^2}{2}\right)E_2 \\ -z_{k-1}\exp\left(-\frac{z_{k-1}^2}{2}\right)E_3+(U_k+z_{k-1})\exp\left(-\frac{(U_k+z_{k-1})^2}{2}\right)E_3 \end{bmatrix}+\pi(E_1)^2 \quad (5.7.24)$$

where $E_1=erf\left(\frac{U_k+z_{k-1}}{\sqrt{2}}\right)-erf\left(\frac{z_{k-1}}{\sqrt{2}}\right)$, $E_2=erf\left(\frac{U_k+z_{k-1}}{\sqrt{2}}\right)$, $E_3=erf\left(\frac{z_{k-1}}{\sqrt{2}}\right)$

$$E_4=(U_k+z_{k-1})\exp\left(-\frac{(U_k+z_{k-1})^2}{2}\right)$$

and erf is known as the error function which is given as

$$erf(z_k)-erf(z_{k-1})=\left(\frac{2}{\sqrt{\pi}}\right)\int_{z_{k-1}}^{z_k}e^{-u^2}\partial u$$

using values obtained in (5.7.22) to (5.7.24) in equation (5.7.11),we have

Minimize

$$\sum_h\sum_k\frac{V_h}{2(b-a)}(E_1)\left\{\begin{matrix} \beta^2\frac{2U_kE_1\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-12U_k^2\left(V_h+2x_{h-1}\right)^2}{3(E_1)^2} \\ +\pi(E_1)^2+\gamma^2 2\sqrt{2}(b-a)\begin{bmatrix} E_2\left(z_{k-1}\exp\left(-\frac{z_{k-1}^2}{2}\right)-E_4\right) \\ -E_3\left(z_{k-1}\exp\left(-\frac{z_{k-1}^2}{2}\right)+E_4\right)\end{bmatrix}\end{matrix}\right\}\sqrt{c_{hk}}$$

Subject to

$$\begin{aligned}&\sum_h V_h=d_x \\ &\sum_k U_k=t_z \\ &\forall V_h\geq 0, U_k\geq 0 \quad , \quad \begin{matrix} h=1,2,\ldots,L \\ k=1,2,\ldots,M\end{matrix}\end{aligned} \quad (5.7.24)$$

Let us assume that the interval of X is defined as $x \in [0,1]$ and the variable Z to be truncated at 4 i,e $z \in [0,4]$ and by simulation in R-software estimates β = 0.07 and γ = 0.73.Thus ,we have (5.7.24) as

Minimize

$$\sum_h \sum_k \frac{V_h}{2}(E_1)\left\{\begin{array}{l} (0.2401)\dfrac{U_k E_1\left(V_h^2+3x_{h-1}^2+3V_h x_{h-1}\right)-6U_k^2\left(V_h+2x_{h-1}\right)^2}{3(E_1)^2} \\ +\pi(E_1)^2+(1.5)\left[\begin{array}{l} E_2\left(z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right)-E_4\right) \\ -E_3\left(z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right)+E_4\right)\end{array}\right]\end{array}\right\}\sqrt{c_{hk}}$$

Subject to

$$\sum_h V_h = 1$$

$$\sum_k U_k = 4 \qquad (5.7.24)$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1,2,...,L \\ k = 1,2,...,M \end{array}$$

For obtaining OSB we assume that the values of $c_{11} = 2, c_{12} = 3$, $c_{13} = 4, c_{21} = 5$, $c_{22} = 6$, $c_{23} = 7$ for total 6 (2×3) strata i,e 2 strata along X variable and 3 along Z variable, execute a computer programme in LINGO by assuming all these conditions given above, we have

Table 5.7.4: OSB when the auxiliary variables X and Z follow uniform and standard normal distributions, respectively

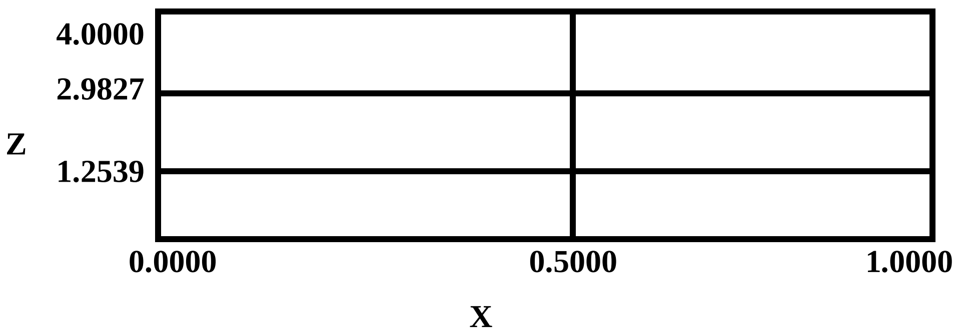


Table 5.7.5: OSB and Variance of proposed method and others when the auxiliary variables X and Z follow uniform and standard normal distribution respectively

OSB $\left(x_h, z_k\right)$	Variance (Proposed method)
(0.50000,1.2539)	0.00098473
(2.0000,1.2539)	
(0.5000,2.9827)	
(2.0000,2.9827)	
(0.5000,4.0000)	
(2.0000,4.0000)	

The Table 5.7.4 shows the stratification points along X variable and Z variable under optimum allocation when variable along X variable is uniformly distributed and Z variable is normally distributed. However table 5.7.5 presents both stratification points as well as variance obtained through the proposed method.

5.7.3: Let the variable X follows a distribution with pdf as

$$f(x)=\begin{cases}2(2-x) & ;1\le x\le 2\\ 0 & ;otherwise\end{cases} \tag{5.7.25}$$

and the other auxiliary variable Z follows truncated exponential distribution with pdf

$$f(x)=\begin{cases}e^{-z+1} & ;\quad 1\le z\le 6\\ 0 & ;otherwise\end{cases} \tag{5.7.26}$$

In order to obtain OSB under proportional allocation having the pdf's of the auxiliary variables defined in (5.7.25) and (5.7.26) we need to obtain the value of W_{hk} and σ^2_{hky} for that we have to substitute these pdf's in equations (5.2.4) to (5.2.6),we get

$$W_{hk}=m_1 V_h e^{-z_k+1} \tag{5.7.27}$$

$$\sigma_{hkx}^{2}=\frac{m_1U_ke^{-z_k+1}}{\left(m_1e^{-z_k+1}\right)^2}\left\{\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-\frac{1}{2}\left[m_2\right]\right\}-4U_k^2 \tag{5.7.28}$$

$$\sigma_{hkz}^{2}=\frac{m_1\left\{m_3\right\}-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{m_1^2} \tag{5.7.29}$$

where $m_1=\left(e^{U_k}-1\right)\left(4-V_h-2x_{h-1}\right)$

$m_2=\left(V_h+2x_{h-1}\right)\left[V_h\left(V_h+2x_{h-1}\right)+x_{h-1}\left(1+x_{h-1}\right)\right]$

$m_3=z_{k-1}^2e^{U_k}-U_k^2-z_{k-1}^2-U_kz_{k-1}+2\left[e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1\right]$

Using the values obtained above in equation (5.7.11), we have MPP as

Minimize

$$\sum_h\sum_k\sqrt{\left(m_1V_he^{-z_k+1}\right)}\left(\begin{array}{l}\beta^2\dfrac{m_1U_ke^{-z_k+1}}{6\left(m_1e^{-z_k+1}\right)^2}\left\{\left[8\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-3\left[m_2\right]\right]-4U_k^2\right\}\\+\gamma^2\dfrac{m_1\left\{m_3\right\}-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{m_1^2}\end{array}\right)$$

Subject to

$$\begin{aligned}&\sum_h V_h=d_x\\&\sum_k U_k=t_z\\&\forall V_h\geq 0,U_k\geq 0 \quad , \quad \begin{array}{l}h=1,2,...,L\\k=1,2,...,M\end{array}\end{aligned} \tag{5.7.30}$$

By using the given pdf's a simulation has been done in R-software and the values of β=0.576 and γ = 0.48 and have been obtained. Thus using values of β and γ, d_x =1 and t_z =5 as given above the defined interval for X and Z respectively for total 6 (2×3 strata. Thus (5.7.30) can be written as

Minimize

$$\sum_h\sum_k\sqrt{\left(m_1V_he^{-z_k+1}\right)}\left(\begin{array}{l}(0.055)\dfrac{m_1U_ke^{-z_k+1}}{\left(m_1e^{-z_k+1}\right)^2}\left\{\left[8\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-3\left[\left[m_2\right]\right]\right]-4U_k^2\right\}\\+(0.234)\dfrac{m_1\left\{\left[m_3\right]\right\}-e^{U_k}\left(1+z_{k-1}\right)-U_k-z_{k-1}-1}{m_1^2}\end{array}\right)$$

Subject to

$$\sum_{h} V_h = 1$$
$$\sum_{k} U_k = 5 \qquad (5.7.31)$$

$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2 \\ k = 1, 2, 3 \end{array}$$

Executing a computer programme for the MPP (5.7.31) using LINGO software, we get the OSB as given in below table

Table 5.7.6: OSB and Variance when the auxiliary variables X and Z are having right triangular and exponential distribution respectively

OSB (x_h, z_k)	**Variance (Proposed method)**
(1.4681,1.5269)	
(2.0000,1.5269)	
(1.4681,3.0791)	0.004217
(2.0000,3.0791)	
(1.4681,6.0000)	
(2.0000,60.0000)	

This table shows the stratification points if the variables are having the distribution functions of X as Right triangular and Z as exponential, under optimum allocation, for 2 strata along X variable and 3 strata along Z variable.

CHAPTER 6

DISCUSSION

The problem of optimum stratification was first discussed by Dalenius (1950) who gave sets of equations giving optimum strata boundaries for optimum and proportional methods of allocating the sample to different strata. Theses equations were given for stratified simple random sampling estimate of population mean. The equations involved population parameters which were functions of the optimum strata boundaries. Due to implicit nature of these minimal equations, their exact solutions can not be obtained. Subsequently, various authors gave methods of obtaining approximations to the exact solutions. Another important aspect of this problem was that both stratification and estimation variable were taken to be same. Later on several authors worked on obtaining optimum strata boundaries using mathematical programming approach and developed a number of methods for the construction of stratification points.

In the present investigation we have tried to consider the problem of optimum stratification on the basis of some suitably chosen auxiliary variables. The optimum strata boundaries so obtained correspond to the minimum variance of the stratified estimate for the population mean of the estimation variable. The results obtained during the present investigation entitled "Optimum stratification with auxiliary information using mathematical programming" in preceding chapters have been discussed under the following heads:

i) Classical optimization technique

ii) Mathematical programming technique

i) Classical optimization technique

In this approach it has been assumed that we know the form of the regression estimation variable Y on the auxiliary variables X and Z and also the form of the conditional variance function $V(y \mid x,z)$.Furthermore methods have been developed in different allocation methods.

In the empirical studies discussed under the heading of 4.11 we have assumed the linear regression line Y on X and Z have been taken as, of the form $y = \alpha + \beta x + \gamma z + e$. For the conditional variance function $\eta(x,z)$ we have taken two forms viz. $\eta(x,z) = \alpha$

and $\eta(x,z)=\lambda xz$ where α *and* λ are constants. The origin is deliberately excluded from the range of the auxiliary variables X and Z otherwise $\eta(x,z)=\lambda xz$ we have $m_1(x,z)=\infty$ at x=0,z=0 and the function $m_1(x,z)\,f(x,z)$ in that case doesn't belong to the class Ω of functions. We could have also avoided this difficulty by taking some other suitable forms to the functions. For the empirical studies under optimum allocation let us assume values of $\alpha=0.0214, \lambda=0.00437$ which are quite small so that the effect of taking $\eta(x,z)=\alpha$ and $\eta(x,z)=\lambda xz$ is negligibly small. In order to obtain AOSB let us assume that the correlation coefficient between X and Z is denoted by ρ and is equal to 0.65 For this purpose the following density functions of the stratification variables X and Z have been considered.

Table 4.11.1 shows the stratification points of the auxiliary variables X and Z in which it is assumed that both the variates are standard normally distributed. However, the conditional variance is taken of the form $\eta(x,z)=\alpha$. Further Table 4.11.2 shows us both the stratification points (x_h, z_k) of each stratum and the total variance of all the 16 strata too, where 4 strata are to be made along the X variable and 4 strata are to be made along the Z variable in which the variable Z is defined in [0,6] and Z in [0,4].While comparing the proposed method with the Singh and Sukhatme (1969) the percent relative efficiency comes out to be 243.89.However ,when the conditional variance is taken of the form $\eta(x,z)=\lambda xz$. The optimum strata boundaries (OSB) so obtained are shown in Table 4.11.3 and 4.11.4 when the auxiliary variables are not independent. Table 4.11.4 shows us the stratification points and the variance obtained by the developed Cum$\sqrt[3]{D_1(x,z)}$ rule assuming the same number of strata. The percent relative efficiency with respected to the method proposed by Singh and Sukhatme (1969) of the proposed method comes out to be 268.85 which is clearly an indication of superiority of the proposed method.

In empirical study 4.11.5 it is assumed that the variable X follows uniform distribution defined in the closed interval [1,2] and the variable Z follows exponential truncated at z = 6 i,e defined in [1,6].Table 4.11.5 displays the OSB and the variance obtained by using Cum$\sqrt[3]{D_1(x,z)}$ rule and the variance obtained by Singh and Sukhatme (1969) when the conditional variance $\eta(x,z)$of is taken of the form$\eta(x,z)=\alpha$. The table also worked out the percent relative efficiency of proposed method over the method developed by Singh and Sukhatme (1969) which is 160.98 that indicates the positive effect of results by using the proposed method. Furthermore, when the conditional variance $\eta(x,z)$ is taken of the form $\eta(x,z)=\lambda xz$ the results obtained are displayed in Table 4.11.6. when the two auxiliary variables are dependent to each other. By changing the form of conditional variance the percent relative efficiency has reached 275.76 from 160.98 when having the same comparison with the method developed by Singh and Sukhatme (1969).

The empirical studies discussed to illustrate the developed Cum $\sqrt[3]{D_2(x,z)}$ Rule under the heading of 4.14 in which for the sake of simplicity the linear regression line Y on X and Z have been taken as, of the form $y=\alpha+\beta x+\gamma z+e$ .For the conditional variance function $\eta(x,z)$ we have taken two forms viz.

1. $\eta(x)=\alpha_1, \eta(z)=\alpha_2$ *where* $\eta(x,z)=\eta(x)\eta(z)$
2. $\eta(x)=\lambda_1 x, \eta(z)=\lambda_2 z$ *where* $\eta(x,z)=\eta(x)\eta(z)$

where $\alpha_1, \alpha_2, \lambda_1, \lambda_2$ are constants. The origin is deliberately excluded from the range of the auxiliary variables X and Z otherwise $\eta(x)=\lambda_1 x$ *and* $\eta(z)=\lambda_2 z$ we have $m_1(x)=\infty$ at x = 0 and $m_1(x)=\infty$ at z = 0 and the functions $m_1(x)f(x)$ and $m_2(z)f(z)$ in that case does not belong to the class Ω of functions. We could have also avoided this difficulty by taking some other suitable forms to the functions. Now let us assume values of $\alpha_1=0.0432, \alpha_2=0.0334, \lambda_1=0.00432$, $\lambda_2=0.00331$ which are quite small so that the effect of taking $\eta(x)=\alpha_1, \eta(z)=\alpha_2$ and $\eta(x)=\lambda_1 x$, $\eta(z)=\lambda_2 z$ is negligibly small. In empirical study 4.14.1,Table 4.14.1 shows the OSB when X follows uniform distribution defined in [1,2] and Z follows exponential distribution defined in [1,6] by assuming the form of condition variance of the form $\eta(x)=\alpha_1, \eta(z)=\alpha_2$ .Table 4.14.2 shows both the stratification points (x_h, z_k) as well as the variance obtained using Cum $\sqrt[3]{D_2(x,z)}$ Rule and method developed by Singh and Sukhatme (1969).The percent relative efficiency comes out to be 190.49 in this case under the optimum allocation when the two auxiliary variables are independent. The Table 4.14.3 represents the stratification points along the X variable and Z variable when the total strata to be made are 6 in which 2 along X variable and 3 along Z variable are to be made. When the conditional variance is taken of the form $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$ under optimum allocation for the case of independent auxiliary variables for uniform and exponential auxiliary variables, the OSB obtained using the Cum$\sqrt[3]{D_2(x,z)}$ Rule are shown in Table 4.14.4.The table is also having the variance of the proposed method and an existing method along with the percent relative efficiency which comes to be 324.82, that is the indication of the superiority of the proposed method over the existing method.

The OSB obtained using the Cum$\sqrt[3]{D_2(x,z)}$ rule when the auxiliary variables X and Z follow right triangular and uniform distributions respectively under the optimum allocation for the two variables used as basis of stratification are independent to each other by assuming the conditional variance of the form are shown in the Table 4.14.5 along with total variance obtained in 6 strata of which 3 along the X variable and 2 along the Z variable have been made. Table 4.14.6 includes the OSB obtained by using the Cum$\sqrt[3]{D_2(x,z)}$ rule for right triangular and uniform auxiliary variables when the form of conditional variance is assumed to be $\eta(x)=\lambda_1 x \quad and \quad \eta(z)=\lambda_2 z$.The table also shows the variance obtained in the given condition.

Under the proportional allocation for the case of dependent auxiliary variables the Cum$\sqrt[3]{D_3(x,z}$ rule has been developed. By assuming the linear regression among the estimation variable Y and the stratification variables X and Z as discussed above, the OSB obtained when both the variables are standard normally distributed and are truncated at = 6 and z = 4 are shown in Table 4.19.1 in which 16 strata are to be made, 4 along the X variable and 4 along the Z variable. Further Table 4.19.2 shows both the stratification points and the total variance obtained through the Cum$\sqrt[3]{D_3(x,z)}$ rule. While comparing the proposed method with the method given by Singh (1974a) by computing percent relative efficiency that results in 268.21 which indicates that the proposed method is more preferable.

Table 4.19.3 shows the OSB (x_h, z_k) under proportional allocation, when the variable X follows right triangular distribution and the variable Z follows exponential distribution and they are dependent to each other.

Using the Cum$\sqrt[3]{D_3(x,z)}$ rule for obtaining OSB when one of the auxiliary variable X follows standard log-normal distribution and the other auxiliary variable follows uniform distribution, the stratification points are presented in Table 4.19.5 when 3 strata are to be made along X variable and 2 along Z variable. Table 4.19.6 displays both the OSB and variance obtained in creating 6 strata, when the variable X is truncated at x = 10 and Z is defined in [0, 1].

Table 4.22.1 shows the stratification points when the auxiliary variables X and Z are having right triangular and exponential distribution respectively while as Table 4.22.2 displays both the OSB and the variance obtained by using Cum $\sqrt[3]{D_4(x,z)}$ rule in case of proportional allocation when the auxiliary variables are independent to each other. The percent relative efficiencies between proposed method and Yadava and Singh (1984) comes out to be 169.89 which is clearly indication of the superiority of the proposed method.

The OSB obtained for the right triangularly and exponentially distributed auxiliary variables using Cum $\sqrt[3]{D_4(x,z)}$ rule are presented in Table 4.22.3.The variance is also displayed in the same table when the variable X is defined in [1,4] and Z in [1,2] under the proportional allocation for the case of independent auxiliary variables. The method has been compared with Khan *et al*. (2008) and the percent relative efficiency comes out to be 233.64 which prefer the proposed method.

For the case when the auxiliary variables are standard log-normally and uniformly distributed under proportional allocation using Cum $\sqrt[3]{D_4(x,z)}$ rule, the results are presented in Table 4.22.4 after making 3 strata along X variable and 2 along Z variable. The results include both OSB and variance. While comparing it with Khan et al. (2015) the parentage of relative efficiency comes out to be 516.05.

ii) Mathematical programming technique

In order to obtain OSB when we are having two stratification variables under consideration one of the most advanced method for the purpose is using mathematical programming approach in which we optimize the variance subject to the range of the variables with the non negative condition. In this heading several allocation procedures have been discussed. In the subheading of empirical study 5.3 several studies have been made to illustrate the developed method. In empirical study 5.3.1 we have obtained the OSB when the auxiliary variables X and Z are having exponential and right triangular distribution respectively. The OSB are presented in Table 5.3.1 in which the range of X is taken as 10 and Z as 2.In this table 4 strata have been made along the X variable and 3 strata along the Z variable. Table 5.3.2 displays stratification points, stratum weight and total variance. In Fig 1, we considered 25 number of strata but it can be observed that (which displays the graph between number of strata and variance) there is no substantial gain in efficiency for more than 20 strata. The Fig.1 shows that the variance remains constant for some strata when the number of strata reached 20 and then it shows an increasing trend. Thus the strata boundaries for each of the auxiliary variables would be 4 or 5.In order make the comparison of the propose method with several other existing method we have assumed a population of 5000.The results obtained from the example are presented in Table 5.3.3.It can be concluded that proposed method performs better than the methods developed by Dalenius and Hodges (1959) cum $\sqrt{f}$ method, Gunning and Horgan (2004) geometric method, Lavallee-Hidiroglou (1988) method using Kozak's (2004) method and Khan *et al*. (2008) mathematical programming approach.

For the case of dependent auxiliary variables when they follow exponential and right triangular distribution, the OSB obtained using the proposed method are presented in Table 5.3.4.However,Table 5.3.5 presents both stratification points and variance when 3 strata are to be made along X variable and 2 along the Z variable.

Table 5.3.6 displays the OSB (x_h, z_k) and the variance obtained by proposed method and by Singh (1977).From the table the percent relative efficiency is 260.42 which indicates the better performance of the proposed method rather than the method proposed by Singh (1977).

Using the proposed method to obtain OSB when the distribution of X is exponential and Z has pareto distribution, the results obtained by assuming the interval of x in [1,4] and z is [1,10] are presented in Table 5.3.7 having 3 strata along the X variable and 4 strata along the Z variable when the two auxiliary variables are independent.

Table 5.4.1 shows the stratification points when the sample size is distributed equally to each stratum under the proposed method. Further, Table 5.4.2 presents the OSB and variances which reveals that that the variance obtained by the proposed method is much less than the method proposed by Singh (1977). Also the percent Relative Efficiency of the proposed method over Singh (1977) is 276.66 which show the efficiency of the proposed method. Hence we conclude that the proposed method of obtaining OSB when the two independent auxiliary variables are uniformly and right triangularly distributed is preferable when the auxiliary variables are independent. However, when the auxiliary variables are dependent the OSB obtained with total variance are presented in Table 5.4.3.

Table 5.5.1 shows the OSB under proportional allocation when the distributions of auxiliary variables are right triangular and exponential. In this table 2 strata have been made along X variable and 3 along Z variable when both then variables are independent. Further Table 5.5.2 represents OSB as well as variances obtained by proposed method and Thomson (1973) method when the range of X is 1 and Z has 5. Thereby, it is revealed that using two auxiliary variables is better than using univarite auxiliary variable. The percent relative efficiency comes to be 476.85.

Under proportional allocation when the auxiliary variables are dependent and are having right triangular and exponential distributions, the results obtained using the proposed method are presented in Table 5.5.3 and it also shows the variance obtained by the proposed method.

In the empirical study 5.4.2 we have assumed the distribution of X is standard log-normal and Z has uniform distribution. The OSB obtained using the proposed method under proportional allocation for the case of dependent auxiliary variables are presented in Table 5.5.4.Moreover, Table 5.5.5 shows the OSB and variances obtained using proposed method, Khan *et al*. (2005) and Cum $\sqrt[3]{D_4(x,z)}$ Rule and it can be concluded that the variance obtained by proposed method is less than the method obtained by Khan *et al.* (2005) and Cum $\sqrt[3]{D_4(x,z)}$ Rule too. Thus the proposed method is not only preferable than Khan *et al.* (2005) but also rather than using classical optimization technique (Cum $\sqrt[3]{D_4(x,z)}$ Rule).

In order to obtain OSB when the allocation of samples among different strata is based on a joint consideration of the stratum size and the stratum variance. Table 5.6.1 shows the stratification points when the auxiliary variables are uniformly and exponentially distributed with the assumption that the sampling cost per unit among different strata is constant and size of the sample is fixed. However, Table 5.6.2 presents both the OSB and the variance obtained by proposed method and other two existing methods which include Singh and Sukhatme (1969) and Khan *et al*. (2005) methods. From the table the variance of the proposed method is less than other two methods which is followed by Khan *et al.* (2005) .Thus we can conclude here that proposed method is preferable rather than the method developed by Singh and Sukhatme (1969) and Khan *et al*. (2005) when the auxiliary variables are independent.

Similarly, Table 5.6.3 presents the OSB and the variances obtained through proposed method and other two methods for the case of dependent auxiliary variables under the Neyman allocation. Thus, we infer that while making two strata along X variable and 3 along Z variable in this allocation the proposed method performs better than Singh and Sukhatme (1969) and Khan *et al*. (2005).

When the auxiliary variable X follows right triangular distribution defined in closed interval [0,1] and Z follows gamma distribution defined in closed interval [0,6] under Neyman allocation, then the OSB and variance so obtained in this case are presented in Table 5.6.4 having 6 strata. The variance through the proposed method is less than the variance obtained by Reddy *et al*. (2016) and the percent relative efficiency comes out to be 162.77.

Table 5.6.5 shows the OSB and variances obtained by proposed method, Khan *et al*. (2005) and Khan *et al*. (2014) when the auxiliary variable X and Z are independent under the neyman allocation. From the table it can be concluded that the proposed method performs better than the method developed by Khan *et al*. (2005) and Khan *et al*. (2014).

Table 5.6.6 shows the stratification points while as Table 5.6.7 presents both the OSB and variance obtained by using the proposed method and Khan *et al*. (2008,20014) method in which percent relative efficiency is 332.82. Here it is to be concluded that proposed method is more precise than Khan *et al*. (2008, 2014) under Neyman allocation when the stratification variables are independent.

Under the sub heading of 5.7 several empirical studies have been done. Empirical study 5.7.1 has been done when the auxiliary variable X has right triangular distribution and Z has exponential distribution and the stratification points obtained in this case are presented in Table 5.7.1.However, both OSB and variances are presented in Table 5.7.2 obtained using proposed method and Fonolahi and Khan (2014) under optimum allocation for the case of independent auxiliary variables. The percent relative efficiency comes out to be 321.62 which is clearly

indication of better performance of the proposed method over Fonolahi and Khan (2014).While as for the case of dependent auxiliary variables having the same condition, the OSB and variance is shown in Table 5.6.3.

Table 5.7.4 shows the stratification points when the auxiliary variables X and Z are having uniform and standard normal distribution respectively when x is defined in [0,1] and z in [0,4] and Table 5.7.5 presents both the OSB and variance obtained by proposed method.

Table 5.7.6 presents OSB and variance having 2 strata along X variable and 3 along Z variable when the variable X follows right triangular distribution and Z follows exponential distribution defined in [1,2] and [1,6] respectively while using the proposed method under optimum allocation for the case of independent auxiliary variables.

CHAPTER 7

SUMMARY AND CONCLUSIONS

In sample surveys, stratification results in the reduction of the variance in sample estimates. Optimum stratification is a technique of obtaining strata boundaries so that the variance of a particular estimator is minimized for the type of allocation envisaged. Dalenius (1950) first considered the problem of obtaining optimum strata boundaries (OSB) for one study variable in case of stratified random sampling estimates by using study variable for the purpose of stratification. Subsequently, this work was carried out by several other research workers who considered an auxiliary variable closely related to study variable as the basis of stratification. Singh and Sukhatme (1969) considered this problem in more general form and gave various methods of finding approximately optimum strata boundaries (AOSB).Ghosh (1963) considered the problem of optimum stratification for two study variables, under proportional allocation, assuming stratification variables as identical to the estimation variables under consideration. For practical utility of this theory, it is desirable to stratify the population on the basis of some auxiliary variables. The problem of construction optimum stratification for two study variables based on auxiliary variable that follow respectively a uniform and a right-triangular distribution has been discussed by Khan *et al*. (2014a).The problem of determining the OSB has been formulated as NLPP ,which turned out to be multistage decision problems and were solved using dynamic programming techniques. The comparative study has been done on simulated data of two sets and the results were obtained by cum√f method of Dalenius and Hodges (1959), geometric method by Gunning and Horgan (2004),the generalized method of Lavallee and Hidiriglous (1988) and proposed method, and it was found that construction of strata using auxiliary variable for the populations with uniform and right-triangular distributions, leads to substantial gains in precision of the estimates while using the proposed technique.

In the present investigation, we have considered the problem of optimum stratification for one study variable Y, by taking two auxiliary variables X and Z as stratification variables. During the present investigation it is assumed that the finite population under consideration is a random sample from an infinite super population with the same population characteristics. Further, priori knowledge of the forms of regression of Y on X and Z and also the form of the conditional variance functions $V(y \mid x,z)$ are assumed. In the study we have assumed that regression of Y on X in the population is given by $Y = c(X,z)+e,$ where e is the error term such that $E(e \mid x,z)=0 \quad and \quad V(e \mid x,z)=\eta(x,z)>0$ for all $x \in [a,b]$ with $(b-a)<\infty$ and $z \in [c,d]$ with $(d-c)<\infty$.The functions $c \quad and \quad \eta$ of x and z are assumed to be known. Under this regression model we have obtained sets of minimal equations giving optimum points of stratification for different allocation methods. These equations are explicit in nature and are very difficult to solve. Hence the methods of finding AOSB have been proposed and a Cum$\sqrt[3]{D_1(x,z)}$ rule, where $D_1(x,z)=m_1(x,z)\,f(x,z)$ has been provided when the two auxiliary variables X and Z are dependent under optimum allocation. Table 4.11.4 shows us the stratification points and the variance obtained by the developed Cum$\sqrt[3]{D_1(x,z)}$ rule assuming the same number of strata. The percent relative efficiency with respected to the method proposed by Singh and Sukhatme (1969) of the proposed method comes out to be 268.85 which is sufficient indication of superiority of the proposed method. Also, for the case of independence under optimum allocation Cum$\sqrt[3]{D_2(x,z)}$ rule, where $D_1(x,z)=m_1(x)m_1(z)\,f(x)\,f(z)$ has been proposed to obtain AOSB. Table 4.14.2 shows both the stratification points (x_h, z_k) as well the variance obtained using Cum$\sqrt[3]{D_2(x,z)}$ Rule and method developed by Singh and Sukhatme (1969).The percent relative efficiency comes out to be 190.47 in this case under the optimum allocation when the two auxiliary variables are independent. The problem of optimum stratification has also been discussed when we take stratum variances of the estimators of the population characteristics into consideration and use proportional allocation. In this case also, a Cum$\sqrt[3]{D_3(x,z)}$ rule, where $D_3(x,z)=c^{'2}(x,z)\,f(x,z)$ has been proposed to obtain AOSB when the auxiliary variables X and Z are dependent to each other and Cum$\sqrt[3]{D_4(x,z)}$ rule, where $D_4(x,z)=c^{'2}(x)c^{'2}(z)\,f(x)\,f(z)$ when they are independent. While comparing the proposed method with the method proposed by Singh (1974a) by computing percent relative efficiency that results in 268.21 which indicates that the proposed method id more preferable. The results obtained through all these proposed techniques under classical optimization techniques shows gain in precision over the conventional methods proposed by several authors like Singh (1971b), Ekman (1959a).

Apart from the classical optimization technique, mathematical programming technique has been used to obtain stratification points. Different techniques have been proposed at different allocation methods The problem of determining the optimum strata boundaries (OSB) has been redefined as the problem of determining optimum strata width (OSW) and is formulated as a NLPP that seeks minimization of the variance of the estimated population mean under different allocations subject to the constraint that the sum of the widths of all the strata is equal to the range of the distribution. The formulated NLPP turns out to be a multistage decision problem that was solved by dynamic programming technique. The results obtained in Table 5.3.3concludes that proposed method performs better than the methods developed by Dalenius and Hodges (1959) cum $\sqrt{f}$ method, Gunning and Horgan (2004) geometric method, Lavallee-Hidiroglou (1988) method using Kozak's (2004) method as well as Khan *et al*. (2008) mathematical programming approach. The proposed technique under Proportional allocation leads to the substantial increase in precision while comparing with the method proposed by Thomson (1973) under the independence of the two auxiliary variables. However, Table 5.5.5 shows the OSB and variances obtained using proposed method, Khan *et al*. (2005) and Cum $\sqrt[3]{D_4(x,z)}$ Rule and it can be concluded that the variance obtained by proposed method is less than the method obtained by Khan *et al*. (2005) and Cum $\sqrt[3]{D_4(x,z)}$ Rule too. Thus the proposed method is not only preferable than Khan *et al.* (2005) but also rather than using classical optimization technique (Cum $\sqrt[3]{D_4(x,z)}$ Rule).

However, through the computational results obtained through the proposed method under the Neyman allocation when the two auxiliary variables are dependent the variance is less than other existing methods like Singh and Sukhatme (1969) and Khan *et al*. (2005) that shows us the impact of using two auxiliary variables rather than single auxiliary variable as the basis of stratification. Also the variance obtained through the proposed method is much less than the variance obtained by Khan *et al*. (2005) as well as and Khan *et al*. (2014). Table 5.6.7 presents both the OSB and variance obtained by using the proposed method and Khan *et al*. (2008, 2014) method in which percent relative efficiency is 332.82. Furthermore the method has been proposed under the optimum allocation using mathematical programming approach and it is revealed from the empirical study that percent relative efficiency of proposed method over Fonolahi and Khan (2014) is 321.62.While comparing the methods obtained through classical optimization approach with the methods proposed through mathematical programming approach using the empirical studies, it is to be revealed that the approach of mathematical programming shows substantial increase in precision rather than the classical optimization technique.

REFERENCES

Abramowitz, M. and Stegun, I. A. 1972. *Handbook of Mathematical Functions* , Dover, New York.

Aggarwal, O. P. 1974a.On mixed integer quadratic problems. *Naval Research Logistics Quarterly,* **21**: 284-297.

Aggarwal, 0. P. 1974b. On integer solution to quadratic problems by a branch and bound technique. *Trabajas de Estatistica Y De Investigation Operation,* **25**: 65-70.

Anderson, D.W., Kish, L. and Cornell ,R.G. 1976.Quantifying gains from stratification for optimum and approximately optimum strata using Bivariate normal model. *Journal of American Statistical Association,* **71**: 887-892.

Aoyama, H.1954. A study of stratified random sampling. *Annals of Institute of Statistical Mathematics,* **6** : 1-36.

Arshad, M., Khan, S. U. and Ahsan, M. J. 1981.A procedure for solving concave quadratic programs. *Aligarh Journal of Statistics,* **1**(2):106-112.

Baillargeon, S. and Rivest ,L.P. 2009. A General Algorithm for Uni-variate Stratification. *International Statistical Review* , **77**(3): 331-344.

Baillargeon, S. and Rivest , L. P. 2011. The construction of stratified designs in R with the package stratification. *Survey Methodology,* **37**(1): 53-65.

Bellman, R.E.1957. *Dynamic Programming*. Princeton University Press, Princeton.

Block, E. 1958.Numerical considerations for the stratification of variables following a log-normal distribution. *Skandinavisk Aktuarietidskrift* **,** **41**: 185-200.

Buhler, W. and Deutler, T. 1975. Optimal stratification and grouping by dynamic programming. *Metrik,* **22**: 161-175.

Charnes, A. and Cooper, W. W. 1957. Nonlinear power of adjacent extreme point methods of linear programming. *Econometrica,* **25**: 132-153.

Chen, H.D., Hearn, D.W. and Lee, C.Y. 1994.A dynamic programming algorithm for dynamic lot size models with piecewise linear cost. *Journal of Global Optimization*, **4**(4): 397-413.

Cochran, W.G. 1961. Comparison of methods for determining strata boundaries. *Bulletin of the International Statistical Institute* , **38**: 345-358.

Cochran ,W.G 1963 .*Sampling Technique.* John wiley and Sons,Inc.

Dalenius, T. 1950. The problem of optimum stratification. *Skandinavisk Aktuarietidskrift,* **33** : 203-213.

Dalenius ,T. 1957. *Sampling in Sweden contribution to the methods and theories of sample survey practice*. Almqvist and Wicksell, Stockholm.

Dalenius, T. and Gurney, M. 1951.The problem of optimum stratification II.*Skandinavisk Aktuarietidskrift* **, 34**: 133-148.

Dalenius, T. and Hodges, J.L. 1959. Minimum variance stratification. *Journal of American Statistical Association ,* **54**: 88-101.

Danish , F. and Rizvi, S.E.H. 2017. On Optimum Stratification Using Mathematical Programming Approach. *International Research journal of Agricultural Economics and Statistics* , **8**(2):435-439.

Danish, F., Rizvi, S.E.H. Jeelani, M. I and Reashi J.A. 2017a. Obtaining Strata Boundaries under Proportional Allocation with Varying Cost of Every Unit. *Pakistan Journal of Statistics and Operations Research*, **13**(3): 567-574

Danish, F., Rizvi, S.E.H., Sharma, M.K. and Jeelani, M. I. 2017b.Optimum Stratification Using Mathematical Programming Approach: A Review. *Journal of Statistics Applications & Probability Letters*, 4(3): 123-129

Dantzig, G. B., Johnson, S. M. and White, W. B. 1958.A linear programming approach to the chemical equilibrium problem. *Management Science*, **5**: 38-43.

Des Raj 1956 .A note on the determination of optimum probabilities in sampling without replacement. *Sankhya ,* **17**: 197-200.

Des Raj 1964. On forming strata of equal aggregate size. *Journal of American Statistical Association,* **59**: 481-486.

Deutler, T. 1976. The determination of optimum stratification-a comparison of solution procedures. *Jahrbucher Fur Nationalokonomie Und Statistik*, **191**: 153-173.

Du, D.Z. , Wu, F. and Zhang, X.S. 1990. On Rosen's gradient projection methods**.** *Annals of Operations Research,* **24**(1-4): 11-28.

Durbin , J. 1959.Review of Sampling in Sweden. *Journal of the Royal Statistical Society: Series A,* **122**: 246-248.

Ekman, G. 1959a. An approximation useful in univarite stratification. *Annals of Mathematical Statistics,* **30**: 219-229.

Ekman, G. 1959b. Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution. *Annals of Mathematical Statistics*, **30**: 1131-1134.

Ekman, G. 1960. A limit theorem in connection with stratified sampling. part II. *Skandinavisk Aktuarietidskrift*, **42**: 1-16.

Fonolahi, A.V. and Khan, M.G.M. 2014. Determing the optimum strata boundaries with constant cost factor. *Conference: IEEE Asia-Pacific World Congress on Computer Science and Engineering (APWC), At Plantation Island, Fij.*

Finkbeiner, B. and Kail, P. 1978. Direct algorithm in quadratic programming. *Zeitschrift fur Operation Research,* **17**: 45-54.

Fletcher, R. 1971. A general quadratic programming algorithm. *Journal of the Institute of Mathematics and its Applications,* **7**: 76-91.

Goldfarb, D. 1969. Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints. *SIAM Journal on Applied Mathematics,* **17**: 739-764.

Ghosh, S.P. 1963: Optimum stratification within two characters. *Annals of Mathematical Statistics,* **34**: 866-872.

Graves, R. L. 1967. A principal pivoting simplex algorithm for linear and quadratic programming. *Operation Research*, **15**: 482-494.

Gunning, P. and Horgan , J.M. 2004. A new algorithm for construction of strata boundaries in skewed population. *Survey Methodology*, **30**(2): 159-166.

Hansen, M.H., Hurwitz, W.N. and Madow,W.K.1953. *Sample Survey Methods and Theory, Vol. I & II*. John Wiley and Sons, New York.

Hess , I. ,Sethi ,V.K. and Balkrishnan, T.R. 1966.Stratification : A practical Investigation. *Journal of American Statistical Association*, **61**:74-90.

Hidiroglou, M. A., and Kozak, M. 2017. Stratification of Skewed Populations: A Comparison of Optimisation-based versus Approximate Methods. *International Statistical Review*, doi: 10.1111/insr.12230.

Horgan , J.M. 2010.Chossing the strata boundaries: An elusive optima .*Istanbul universitiesi isletme fakuitesi Dergisi*, **39**(2): 195-204.

Iachan R. 1985. Optimum strata boundaries for shellfish surveys. *Biometrics* , **41**: 1053-1062.

Isil , K. and Taga, Y.1969. On optimal stratification for multivariate distributions. *Skandinavisk Aktuarietidskrift* , **52** : 24-38.

Jarque, C.M. 1981. A solution to the problem of optimum stratification in multivariate sampling. *Journal of Applied Statistics*, **30**: 163-169

Kelley, J. E. 1960. The cutting-plane method for solving convex programs. *Journal* of the *Society* for *Industrial* and *Applied Mathematics*, **8**:703-712.

Khan, E. A., Khan, M. G. M. and Ahsan, M. J. 2002. Optimum stratification: a mathematical programming approach. *Calcutta Statistical Association bulletin*, **52**: 323-333.

Khan, M.G.M. , Khan, E.A. and Ahsan , M.J. 2003. An optimal multivariate stratified sampling design using dynamic programming**.** *Australian and New Zealand Journal of Statistics*, **45**(1): 107–113.

Khan, M. G. M. , Khan, E. A. and Ahsan, M.J. 2011. Optimum allocation in multivariate stratified sampling in presence of non response. *Journal of the Indian Society of Agricultural Statistics*, **62**(1):42-48.

Khan, M. G. M. , Khan, E. A. and Jahan, N. 1994. Optimum of strata for surveys with small sample size. *Aligarh Journal of Statistics,* **14**: 45-53.

Khan, M. G. M., Najmussehar and Ahsan, M. J. 2005. Optimum stratification for exponential study variable under Neyman allocation. *Journal of the Indian Society of Agricultural Statistics,* **59**(2):146-150.

Khan, M. G. M., Nand, N. and Ahmad, N. 2008. Determining the optimum strata boundary points using dynamic programming. *Survey Methodology,* **34**(2): 205-214.

Khan, M.G.M. , Prasad, D.K. and Rao, D.K. 2014a.On optimum stratification. *International Journal of Mathematical, Computational, Natural and Physical Engineering*, **8**(3): 513-517.

Khan, M. G. M. , Prasad, V. D. and Rao, D. K. 2014b. Optimum stratification using auxiliary variables. *Aligarh Journal of Statistics*, **34**: 85-104.

Khan, M. G. M. , Rao, D., Ansari , A.H. and Ahsan, M.J. 2015.Determining optimum strata boundaries and sample sizes for skewed population with log-normal distribution. *Communication in Statistics –Simulation and Computation*, **44**:1364-1387.

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, Z. A., Ahsan. M.J. and Khan, E.A. 1983.On mixed integer concave quadratic programs. *Soochov Journal of Mathematics,* **9**: 111-116.

Kish, L. 1965. *Survey Sampling*. Wiley New York.

Kitagawa, T. 1956. Some contributions to the design of sample surveys. *Sankhya* , **17** :1-36.

Kozak, M. 2004.Optimal Stratification using Random search Method in Agricultural Surveys. *Statistics in Transition,* **6**(5): 797-806.

Kozak, M. , Verma, M.R. and Zielinski , A. 2007.Modern approach to optimum stratification: Review and perspectives. *Statistics in Transition*, **8**(2): 223-250.

Kuhn, H. W. and Tucker, A. W. 1951.Nonlinear programming. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 481-492.

Lai, Y. L., Gao, Z.Y. and He, G.P. 1993 A generalized gradient projection algorithm of optimization with nonlinear constraints. *Science China Mathematics, Series A*, **36**(2): 170-180.

Lavallee, P. and Hidiriglou, M. 1988. On the stratification of skewed populations. *Survey Methodology,* **14**: 33-43.

Lemke, C. E. 1962. A method for solution of quadratic programs. *Management Science,* **8**: 442-453.

Lin, C.C. 1994. A systolic algorithm for dynamic programming. *Computers and Mathematics with Applications,* **27** (l): 1-10.

Mahajan , P.K. Gupta J.P. , Mandowara, V.L. and Gupta, P.C. 1994. Optimum points of stratification for two or more stage designs. *Communication in Statistics-Theory and Methods*, **23**: 947-958.

Mahalanobis, P.C. 1952. Some aspect of design of sample surveys. *Sankhya* , **12** : 1-17.

Mandowara, V. L. and Gupta, P.C. 1994. Optimum points of stratification for two or more stage designs. *Communication in Statistics –Theory and Methods,* **23**: 947-958.

Markowit , H. M. and Manne. A. S. 1957.On the solution of discrete programming problems. *Econometrica,* **25**: 84-110.

Mathew, O. O., Sola, A. F., Oladiran, B. H. and Amos, A. A. 2013. Efficiency of Neyman allocation procedure over other allocation procedures in stratified random sampling. *American Journal of Theoretical and Applied Statistics,* **2**(5): 122-127.

Mehta, S. and Mandowara, V.L. 2012. An optimum stratification for stratified cluster sampling design when clusters are of varying sizes. *International journal of scientific and technology research*, **1**(9): 74-79.

Nilsson, G. 1967. Optimal stratification according to the method of least squares. *Skandinavisk Aktuarietidskrift* , **50**: 128-136.

Panne, C. and Whinston, A 1964a. Simplified methods for quadratic programming. *Navel Research Logistics*, **11**(3-4): 273-302.

Panne, C. van de and Whinston, A 1964b.The Simplex and the Dual Method for Quadratic Programming. *Journal of the Operational Research Society*, **15**(4):355-388.

Pla , L .1991.Determining strata boundaries with multivariate real data. *Biometrics,* **47**: 1409-1422.

Rao, D.K., Khan, M.G.M. and Khan, S. 2012. Mathematical programming on multivariate calibration estimation in stratified sampling. *World academy of science, engineering and technology*, **6**:12-27.

Reddy, K.G. , Khan, M.G.M. and Rao, D.K. 2016.A procedure for computing optimal strata boundaries and sample sizes for multivariate surveys. *Journal of Software,* **11**(8):816-832.

Rao, D.K. , Khan, M.G.M. and Reddy, K.G. 2014.Optimum stratification of a skewed population. *International Journal of Mathematical, Computational, Natural and Physical Engineering*, **8**(3): 497-500.

Rivest, R. J. 2002. A generalization of Lavallee and Hidiroglou algorithm for stratification in survey. *Survey Methodology*, **28**: 191-198.

Rizvi, S.E.H., Gupta, J.P. and Singh, R. 2000. Approximately optimum stratification for two study variables using auxiliary information. *Journal of the Indian Society of the Agricultural Statistics,* **53** (3): 287-298.

Rizvi , S.E.H. Gupta, J.P. and Bhargava, M. (2002). Optimum stratification based on auxiliary variable for compromise allocation. *Metron*, **28**(1):201–215.

Rizvi, S.E.H. , Gupta , J.P. and Bhargava , M. 2004. Effect of optimum stratification on sampling with varying probabilities under proportional allocation. *Statistica* , **4**: 721-733.

Rosen, J. B. 1960.The gradient projection method for nonlinear programming, Part I : linear constraints. *Journal of the Society* for *Industrial and Applied Mathematics,* **8**: 181-217.

Rosen, J. B. 1961.The gradient projection method for nonlinear programming, Part II : nonlinear constraints. *Journal* of the *Society* for *Industrial* and *Applied Mathematics,* **9**: 514-532.

Sai , S. and Taguri , M. 1989. Optimum stratification based on a concomitant variable and its application to the current statistics of commerce. *Reports of Statistical Application Research ,Union of Japanese Scientists and Engineers, **36***: 22-31.

Samanta, M. 1965. A note on the problem of optimum truncation of a bivarite population in stratified random sampling. *Annals of Mathematical Statistics,* **17** :363-375.

Schaffer, K. A. 1971. Optimum stratification for log-normal distribution. *Metrika*, **28**: 179-189.

Schneeberger. 1966. Optimal stratification with optimal allocation. *Journal of Applied Mathematics and Mechanics,* **46**:106-108.

Schneeberger , H. and Drefahl , D. 1980. Limits of feasible sampling fractions in optimal stratification. *Statistische Hefte* , **21**: 61-65.

Schneeberger , H. and Drefahl , D. 1982. Limits of feasible sampling fractions in optimum stratification and optimal allocation in dependence on the sampling fraction. *Statistische Hefte* , **23**: 228-236.

Schneeberger , H. and Gollar , W. 1979. On the problem of feasibility of optimal stratification points according to Dalenius. *Statistische Hefte* , **20**:250-256.

Sebnem, Er. 2011. Computation methods for optimum stratification: A overview. *International Statistical institute: Proceeding 58th world statistical congress.* Dulbin (STS058).

Serfling, R.J. 1968. Approximately optimum stratification. *Journal of American Statistical Association,* **63**: 1298-1309.

Sethi, V.K. 1963. A note on optimum stratification of populations for estimating the population means. *Australian Journal of Statistics.* **5**: 20-23.

Sethumadhavi, R. 1966. Stratification in surveys on fruit crops. Unpublished M.Sc thesis submitted to IARI, New Delhi.

Singh, R. 1971a. Determination of optimum strata boundaries. *Journal of Indian Society Of Agricultural Statistics* , **23**: 115-123.

Singh, R. 1971b. Approximately optimum stratification on auxiliary variable. *Journal of American Statistical Association,* **66**: 829-833.

Singh, R. 1975a. An alternative method of stratification on the auxiliary variable. *Sankhya C* , **37**: 100-108

Singh, R. 1975b. A note on optimum stratification in sampling with varying probabilities. *Australian Journal of Statistics,* **17**: 12-21.

Singh , R. 1977. A note on equal allocation with ratio and regression methods of estimation. *Australian Journal of Statistics,* **19**: 96-104.

Singh, R. and Prakash, D. 1975. Optimum stratification for equal allocation. *Annals of Institute of Statistical Mathematics,* **27**: 273-280.

Singh, R. and Sukhatme, B.V. 1969. Optimum stratification. *Annals of Institute of Statistical Mathematics,* **21**: 515-528.

Singh, R. and Sukhatme, B.V. 1972a. A note on optimum stratification. *Journal of Indian Society of Agricultural Statistics,* **24**: 91-98.

Singh, R. and Sukhatme, B.V. 1972b. Optimum stratification in sampling with varying probabilities. *Annals of Institute of Statistical Mathematics***, 24**: 485-494.

Singh, R. and Sukhatme, B.V. 1973. Optimum stratification with ratio and regression methods of estimation. *Annals of Institute of Statistical Mathematics,* **25**: 627-633.

Taga , Y. 1967. On optimum stratification for the objective variable based on concomitant variable using prior information. *Annals of Institute of Statistical Mathematics,* **19**: 101-130.

Taguri, M. 1980. Optimum stratification points for five typical distributions. *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers ,* **27**: 37-53.

Taguri, M. 1882a. Optimum stratification and its Robustness, I. *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers ,* **29**: 7-18.

Taguri, M. 1882b. Optimum stratification and its Robustness, II. *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers,* **29**: 19-29.

Taguri, M. 1882c. Optimum stratification and its Robustness, III. *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers,***29**: 19-31.

Taguri, M. 1882d. Optimum stratification and its Robustness, IV. *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers,***29**: 32-41.

Thomsen, I. 1976. A comparison of approximately optimal stratification given proportional allocation with other methods of stratification and allocation. *Metrika,* **23**:15-25.

Wang, C.L. and Xing, A.Q.1990. Dynamic programming and penalty functions. *Journal* of *Mathematical Analysis* and *Applications,* **150** (2):562-573.

Wang, M.C. and Aggarwal, V. 1984. Stratification under a particular pareto distribution. *Communications in Statistics-Theory and Methods*, **13** : 711-735.

Wolfe, P. 1959. The simplex method for quadratic programming. *Econometrica ,* **27**: 382-398.

Wywial, J. 1995. On optimum stratification of population on the basis of auxiliary variables. *Statistics in Transition*, **2**: 831-837.

Yadava, S.S. and Singh, R. 1984.Optimum stratification for allocation proportional to strata totals for simple random sampling scheme. *Communications in Statistics – Theory and Methods*, **13**: 2793-2806.

CURRICULUM VITA

Name of the student	:	Faizan Danish
Father's name	:	Gull Mohd Ganie
Mother's name	:	Gulshana
Nationality	:	Indian
Date of birth	:	June 12, 1992
Permanent home address	:	Heff Shopian Distt: Shopian, J&K-192303
Email	:	danishstat@gmail.com

EDUCATIONAL QUALIFICATION

Bachelor's degree	:	Mathematics and Statistics
University and year of award	:	University of Kashmir, 2012
Percentage of Marks	:	61.44%
Masters Degree	:	M.Sc. Statistics
University and year of award	:	University of Kashmir, 2014
Percentage of Marks	:	78.29%

CERTIFICATE-IV

Certified that all the necessary corrections as suggested by the external examiner and the Advisory Committee have been duly incorporated in the thesis entitled "Optimum Stratification with Auxiliary Information using Mathematical Programming" submitted by **Mr. Faizan Danish**, Regd. No. **J-14-D-04-BS.**

05/03/18
Major Advisor and Chairman
Advisory Committee

Place: Jammu
Date: 05/03/18

05/03/18
Head
Division of Statistics and Computer Science