Applied Mathematics & Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/amisl/060104

Construction of Strata Boundaries: A Review

S. E. H. Rizvi and Faizan Danish*

Division of Statistics and computer Science, Faculty of Basic Sciences, Main Campus SKUAST-J, Chatha Jammu-180009, India

Received: 4Aug. 2017, Revised: 8 Oct. 2017, Accepted: 12 Oct. 2017.
Published online: 1 Jan. 2018.

Abstract: The principal reason for stratification in the design of sample survey is to reduce the sample variance of the estimates. The main objective of stratification is to give a better cross section of the population so as to gain a higher degree of relative precision. Over the past number of years, many procedures have been developed to obtain optimum strata boundaries. After a brief discussion of fundamental theoretical contribution by Dalenius, a number of approximate rules which have been proposed as a result of the heavy computational work involved in solving the theoretical equations to obtain the optimum points of stratification were critically examined. In this manuscript, an attempt has been made to review and summarise the important works on some aspects of stratified sampling design, particularly the determination of optimum stratification points.

Keywords: Optimum boundaries, Optimum strata width, Minimal equation, Neyman allocation, Optimum allocation, Proportional allocation.

1 Introduction

The main problems in the use of stratified sample design are the choice of sampling design within strata, the choice of a stratification variable, the allocation of the total sample size to strata, the location of the optimum stratification points and the choice of number of strata. We may note that perhaps the best method of utilizing the technique of stratified sampling effectively consists in determining a composite choice of a possible number of solutions to the five problems mentioned above, since the solution to these points are interdependent. To gain the precision in estimates using stratified sampling one of the basic problem is the determination of the optimum strata boundaries (*OSB*) and the research carried out in this paper is to deal with this problem.

Dalenius and Gurney [1], Mahalanobis [2], Hansen et al. [3], Aoyama [4], Ekman [5],Dalenius and Hodges [6], Durbin [7], Sethi [8], Singh [9] used the frequency distribution of the main study variable for determining the strata boundaries under various allocations of the sample sizes. Most of these authors achieved the calculus equations for the strata boundaries which are not suitable to adopt for practical computations. They obtained only the approximate solutions under certain assumptions. Many authors such as Unnithan [10], Lavallée and Hidiroglou[11], Sweet and Sigman [12] and Rivest [13] suggested some iterative procedures to determine *OSB*. These algorithms require an initial approximate solution and also there is no guarantee that the algorithm will provide the global minimum. Moreover, the convergences of some of these algorithms are slow or non-existent .Gunning and Horgan [14] developed an approximate method of stratification for positively skewed populations. They showed that their algorithm is much easier and more efficient than the cum *f* method of Dalenius and Hodges [6] and Lavallee-Hidiroglou [11] method.

2. Optimum Stratification ForOne Character

Dalenius [15] 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 couldn't

*Corresponding author e-mail: danishstat@gmail.com

be evaluated unless these points were known. In deriving these equations Dalenius assumed both stratification and estimation variables to be same. Therefore even if the strata boundaries were known the values of parameters couldn't be estimated without having the knowledge of the distribution of estimation variable. The need having some methods for finding at least approximations to the optimum strata boundaries was therefore left. Mahalanobis [2] suggested that when the number of strata is predetermined, say L , a practical method of possible stratification is to stratify the whole population into an 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 couldn't be obtained in advance ,one has to fall back in some highly correlated auxiliary variable were made equal. These approach 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 is invariant. Hansen et al. [16] demonstrated that Mahalanobis rule would lead to efficient stratification if the coefficient of variations were same for all the strata. Kitagawa [17] named Mahalanobis rule as 'principle of equipartition'.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. An efficient method of constructing strata under Neyman allocation was given by Dalenius and Hodges [6].In this case also, it was assumed that both the estimation and the stratification variables are same. If the estimation variable is Y and F(y) it is 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

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

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 of the variable Y in each can be approximated by uniform distribution. Therefore the result are only asymptotically optimum. It was also shown that when L $\rightarrow\infty$,tyhe variance obtained by using the suggested strata boundaries tends to minimum variance, and $LS_0\rightarrow\frac{K^2}{\sqrt{12}}$ 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. Ekman [5] in another paper obtained the method of finding approximations to the optimum points of stratification 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 hth 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

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

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 later 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 [7] in a review of Dalenius doctoral thesis. Let F (y) be the cumulative density f(y0 of the estimation variable Y. Forming a rectangular distribution

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

over the same range ,the approximations to optimum boundaries are obtained by taking equal intervals on the cumulative of $\frac{1}{2}\left[r\left(y\right)-f\left(y\right)\right]$.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 .Cochran [18] compare 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 √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.

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 ([19] , [20]).In this case also it was assumed that we have the knowledge of the density function of the estimation variable, in the population which was taken to be infinite so that the density function could be taken to be continuous. If the exact form o 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

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

The solutions to these equations given 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. An alternative approach to tackle the problem of optimum stratification was tried by Sethi [8].In place of finding approximations to optimum strata boundaries 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. Para and Jan recently studied many probability distributions [21, 22 and 23]. 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 stratum totals as suggested by Mahalanobis [2] and Hansen et al. [3] doesn't 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. Des Raj [24] 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. Schneeberger [25] 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 proportional allocated to strata. For the case of triangular distribution with density f(x) 0.5 (1-x) , $-1\leq x\leq 1$,results were obtained for 2,3,...,7 strata. The accuracy increased when proposed optimal stratification was employed. Nilson [26] 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.

Singh and Sukhatme [27] 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) =Ø (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

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

and, under proportional; allocation ,the minimal equations were

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

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

$$g_1(t)=\frac{\phi'^2(t)+4\phi(t)c'^2(t)}{\left[\phi(t)\right]^{3/2}}$$

forNeyman 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 [27] ,Singh [28] made an investigation into the accuracy of various methodology relating to stratification on study variable. A new method for obtaining Approximately Optimum Strata Boundaries (AOSB) in case of proportional allocation, known as 'cum $\sqrt[3]{f(y)}$ rule' ,was proposed by Singh [29].The proposed method is easy to apply in practice and involves some 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.

In continuation to their earlier work ,Singh and Sukhatme [30] developed certain asymptotic properties of the AOSB following the method of Ekman [5].Singh and Sukhatme [31] 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. Wang and Aggarwal [32] 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.

3. Optimum Stratification Using Auxiliary Information

In [1] Dalenius and Gurney 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 samplewas 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 ith 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 nearly always be adequate in practice. Aoyama [4] 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. Block [33] 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 points 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.

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 [34].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 is given by which such an optimal stratification can be asymptotically obtained. The use of cum √f rule has been suggested by Serfling [35] 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 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 [36] 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. Singh and Prakash [37] considered the problem of optimum stratification on the auxiliary variable X for equal allocation. They proposed a cum f√∅ rule for obtaining AOSB. An alternative method of stratification was proposed by Singh [38]. He gave a new cum $\sqrt{f(c^{/2} + \theta\phi)^{1/2}}$ rule, where $\theta = \dfrac{12L^2}{(b-a)^2}$ as a generalization to Daleniuscum√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 indicates that the proposed rule and the cum $\sqrt[3]{p(x)}$ rule can, however, be used in more general situations.

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 [39].Anderson et al. [40] applied the theoretical framework developed by Singh and Sukhatme [27] 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.The problem of determining AOSB on the auxiliary variable X ,under equal allocation, using ratio and regression methods of estimation was considered by Singh [41].It is shown that equal allocation is equally efficient to Neyman allocation as regards the methods of estimation considered by him. Thomson [42] made an attempt to study the effect of stratification by using two stratifying variables, say, Y and Z. The method consists in stratifying the population into r strata along X-axis by using cun$\sqrt{f_1}$ method, and constructing s strata along Z-axis by an equal partitioning of cum$\sqrt{f_2}$ method. The results indicate that in many practical situations the gain from using two stratifying variables over one is nontrivial. Yadav and Singh [43] 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. Mandowara and Gupta [44] made an attempt to obtain points of stratification for two are more stage designs with equal primary stage units and subsequent stage units. Stratification on the auxiliary variable when the study variable is closely related tio the auxiliary variable has been made. The determination of OSB in these cases have been illustrated with the help of some known specific distributions. Another method of finding approximately the optimum points of stratification for Neyman's allocation in simple random sampling estimate was given by Ekman [5].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

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

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 within 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 $L \rightarrow \infty$,the approximate strata boundaries 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 for small values of L also, the approximate 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.

Danish et al. [45] considered the problem of obtaining optimum strata boundaries when we have two concomitant variables with one estimation variable and the regression line between them is assumed to be linear. Neyman allocation procedure has been made for obtaining optimum strata boundaries from minimal equations. Due to complexities of minimal equations, approximate to the variance of the study variable has been taken. The approximation depends only on the number of strata, the simultaneous density of stratifying variables and the correlation between the study variable and each of the stratifying variables.

4. Optimum Stratification For Two Characters

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 [46] 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. Samant [47] 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 be possible to neglect finite population correction as is done by him.

5. Other Methods Used for Obtaining Stratification Points

In [17] Kitagawa 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 criteria 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. 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. Hess et al.[16] 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. Also the basic concept regarding the construction of strata boundaries has been discussed by Tarray [48].Sethumadhavi [49] 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 [18] 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 temperature 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. When the study variable follows a log-normal distribution, optimum strata boundaries were derived by Schaffer [50] under the assumption that the variable of stratification is the same as variable taken into consideration. Isil and Taga [51] 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 [52] 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. Schneeberger and Gollar [53] 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=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 [54] 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 DaleniusNeyman allocation solution is not feasible. Taguri [55] has given OSB, minimum variance and efficiencies for five distributions on the following 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 [56] 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 is presented. Taguri [57] 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 ([58] , [59] , [60]) 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 [51] made an investigation into the gain in precision by optimum stratification in dependence on sampling fraction. Iachan [61] considered the problem of optimum stratification when stratification is on auxiliary variables and obtained the AOSB for Shellfish surveys.Sai and Taguri [62] 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. Mahajan et al. [63] considered the problem of optimum stratification in case of sensitive variable for which the data are collected by scrambled randomisedresponse technique. In this regard ,they proposed a rule of finding AOSB. Wywial [64] 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.* [65] considered the optimum stratification for two characters using proportional method of allocation by taking an auxiliary variable in both t5he characters used as stratification variable. Rivest [13] 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 [14] discussed a simple and practicable algorithm for constructing stratum boundaries in such a way that the coefficient of variation are equal in each stratum is derived for positively skewed populations. The new algorithm is shown to compare favourably with the cumulative root frequency method given by Dalenius and Hodges [6] and the Lavallee and Hidiroglou [11] approximation method for estimating the optimum stratum boundaries. Rizvi et al. [66] extends the Singh's [26] problem for proportional allocation, when two variates are under study. A cum $\sqrt[3]{R_3(x)}$ rule for obtaining approximately OSB has been provided. It has been shown theoretical 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. [67] 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 [68] track the progress of various methods used for obtaining optimum strata boundaries and ask 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 [69] 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 equation 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.

6 Conclusions

In this paper an attempt has been made to review and summarise the work done related to optimum strata boundaries. After the fundamental theoretical contribution by Dalenius ([15] , [6]), a number of approximation methods have been proposed like Dalenius and Gurney, Mahalanobis, Hansen, Dalenius and Hodges etc. Most of these proposed methods reveals that the precision could be increased as taking care of the choosing the stratification points while stratifying the heterogeneous population. For a given circumstances, there appears to be a unique optimum set of bounds if allowed to run long enough. Given the availability of high powered computing ,the number of iterations required to reach the optimum doesn't pose a problem. Here we can suggest that the adjustments suggested for Sethi's method and for cum√f rule be supported by empirical studies on a larger scale with various types of data.

References

[1] Dalenius, T. and Gurney, M..(1951).The problem of optimum stratification II. Skandinavisk Aktuarietidskrift. 34:133-148.

[2] Mahalanobis, P.C..(1952).Some aspect of design of sample surveys. Sankhya. 12 :1-17.

[3] 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.

[4] Aoyama, H..(1954). A study of stratified random sampling. Annals of Institute of Statistical Mathematrica.6 : 1-36.

[5] Ekman, G. (1959)..An approximation useful in univaritestratification.Ann.Math. Statist. , 30: 219-229.

[6] Dalenius, T. and Hodges ,J.L. (1957)..Minimum variance stratification. Journal of American Statistical Association. 54: 88-101.

[7] Durbin , J. (1959)..Review of Sampling in Sweden. J. R. Statist. Soc., A 122:246-248.

[8] Sethi, V.K. (1963). A note on optimum stratification of populations for estimating the population means. Australian Journal of Statistics. 5:20-23.

[9] Singh, R.. (1971b)..Approximately optimum stratification on auxiliary variable, Journal of American Statistical Association. 66:829-833.

[10] Unnithan, V. K. G. (1978). The minimum variance boundary points of stratification. Sankhya 40 (C):60–72.

[11] Lavallée, P., and Hidiroglou, M.A. (1988). On the stratification of skewed populations. Survey Methodology, 14, 33-43.

[12] Sweet, E. M., Sigman, R. S. (1995). Evaluation of model-assisted procedures for stratifying skewed populations using auxiliary data. In:Proceedings of the Survey Research Methods Section, Alexandria: American Statistical Association, pp. 491–496.

[13] Rivest, R. J. (2002).A generalization of Lavallee and Hidiroglou algorithm for stratification in survey. Survey Methodology. 28: 191-198.

[14] Gunning, P. and Horgan , J.M.. (2004).A new algorithm for construction of strata boundarties in skewed population. Survey Methodology 30(2): 159-166.

[15] Dalenius, T. (1950)..The problem of optimum stratification. SkandinaviskAktuarietidskrift. 33 : 203-213.

[16] Hess , I. ,Sethi ,V.K. and Balkrishnan, T.R.. (1966).Stratification : A practical Investigation. Journal of American Statistical Association, 61:74-90.

[17] Kitagawa, T. (1956)..Some contributions to the design of sample surveys. Sankhya., 17 :1-36.

[18] Cochran, W.G. (1961).. Comparison of methods for determining strata boundaries. Bull. Int. Statist. Inst. 38: 345-358.

[19] Ekman, G. (1959a).. Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution. Ann.Math.Statist. 30:1131-1134.

[20] Ekman,.(1960).. A limit theorem in connection with stratified sampling. part II. Skand. Akt., 42:1-16.

[21] Para, B. A., & Jan, T. R. (2014). Discretization of Burr-Type III distribution. Journal of Reliability Statistical Studies, 7(2), 87-94.

[22] Para, B. A., & Jan, T. R. (2017). Transmuted Inverse Loglogistic Model: Properties and Applications in Medical Sciences and Engineering, Mathematical Theory and Modeling, 7 (6), 2224-5804.

[23] Para, B.A and Jan, T.R (2017). Inverse Weibull Minimax Distribution: Properties and Applications, Journal of Statistics Applications and Probability, 6 (1), 205-2181.

[24] Des Raj. (1964). On forming strata of equal aggregate size. Journal of American Statistical Association, 59: 481-486.

[25] Schneeberger. (1971).Optimal stratification with optimal allocation. Journal of Applied Mathematics and Mechanics.46:106-108.

[26] Nilsson, G. (1967).. Optimal stratification according to the method of least squares.Skand.Akt. ,50:128-136.

[27] Singh, R. and Sukhatme, B.V. (1969).. Optimum stratification. Annals of Institute of Statistical Mathematrica , 21:515-528.

[28] Singh, R.(1971a). Determination of optimum strata boundaries. Journal of Indian Society Of Agricultural Statistics , 23:115-123.

[29] Singh, R. (1975a.).A note on optimum stratification in sampling with varying probabilities. Australian Journal of Statistics, 17:12-21.

[30] Singh, R. and Sukhatme, B.V.(1972a).A note on optimum stratification. J. Indian society of Agril. Statistics. 24.91-98.

[31] Singh, R. and Sukhatme, B.V.(1972b).Optimum stratification in sampling with varying probabilities.Ann.Inst.Statist.Math.24.485-494.

[32] Wang, M.C. and Aggarwal, V. (1984). Stratification under a particular pareto distribution. Communications in Statistics: Theory and Methods , 13 : 711-735.

[33] Block, E. (1958).Numerical considerations for the stratification of variables following a log-normal distribution. Skand. Akt., 41: 185-200.

[34] Taga , Y. (1967).On optimum stratification for the objective variable based on concomitant variable using prior information. Ann. Inst. Statist. Math. , 19:101-130.

[35] Serfling, R. J. (1968). Approximately optimum stratification. Journal of American Statistical Association. 63: 1298-1309.

[36] Singh, R. and Sukhatme, B.V. (1973). Optimum stratification with ratio and regression methods of estimation. Ann. Inst. Statist. Math. 25: 627-633.

[37] Singh, R. and Prakash, D. (1975). Optimum stratification for equal allocation. Annals of Institute of Statistical Mathematrica.27:273-280.

[38] Singh, R. (1975c).An alternative method of stratification 0n the auxiliary variable. Sankhya C 37.100-108.

[39] Thomsen, I. (1976).A comparison of approximately optimal stratification given proportional allocation with other methods of stratification and allocation. Metrika , 23:15-25.

[40] 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.

[41] Singh , R.. (1977).A note on equal allocation with ratio and regression methods of estimation. Australian Journal of Statistics, 19:96-104.

[42] Thomsen, I. (1977). On the effect of stratification when two stratifying variables are used. Journal of American Statistical Association, 72 : 149-153.

[43] Yadav m 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.

[44] Madowara, V.L. and Gupta, P.C. (1994).Optimum points of stratification for two or more stage designs. Commun. Statist.-Theory and Methods.23.947-958.

[45] Danish, F. and Rizvi, S.E.H. (2017). Optimum Stratification for Bi-variate Stratification variables with Single Study Variable, International Journal of Modern Mathematical Sciences, 15(2): 248-260.

[46] Gosh, S.P. (1963). Optimum stratification within two characters. Annals of Mathematical Statistics. 34: 866-872.

[47] Samanta, M. (1965). A note on the problem of optimum truncation of a bivarite population in stratified random sampling. Ann. Inst. Statist. Math. ,17 :363-375.

[48] Tarray T. A. (2016): Statistical Sample Survey Methods and Theory. Elite Publishers (onlinegatha.com), INDIA,ISBN: 978-93-86163-07-03.

[49] Sethumandhavi, R. (1966). Stratification in surveys on dfruit crops. Unpublished M.Sc thesis submitted to IARI, New Delhi.

[50] Schaffer, K. A.(1971). Optimum stratification for log-normal distribution (in German). Metrika, 17.98-115.

[51] Isil , K. and Taga, Y. (1969).On optimal stratification for multivariate distributions. Skand. Akt. 52 :24-38.

[52] Singh, R. (1975b).An alternative method of stratification on the auxiliary variable.SankhyaC ,37:100-108.

[53] Schneeberger , H. and Gollar , W. (1979)..On the problem of feasibility of optimal stratification points according to Dalenius. Statist. Hefte , 20:250-256.

[54] Schneeberger , H. and Drefahl , D. (1982).Limits of feasible sampling fractions in optimum stratification and optimal allocation in dependence on the sampling fraction. Statist. Hefte , 23: 228-236.

[55] Taguri, M.(1980). Optimum stratification points for five typical distributions. Rep. Statist. Appl. Res. JUSE , .27: 37-53.

[56] Jarque, C.M. (1981). A solution to the problem of optimum stratification in multivariate sampling. Appl. *Statist.* , **30**: 163-169.

[57] Taguri, M. (1982a).. Optimum stratification and its robustness, III. Rep. Statist. Appl. Res. JUSE , 29 :7-18.

[58] Taguri, M. (1982b).Optimum stratification and its robustness,II. Rep. Statist.Appl.Res.JUSE.29.19-29.

[59] Taguri,M. (1982c). Optimum stratification and its robustness,III. Rep. Statist.Appl.Res.JUSE.29.19-31.

[60] Taguri,M. (1982d). Optimum stratification and its robustness,IV. Rep. Statist.Appl.Res.JUSE.29.32-41.

[61] Iachan R.. (1985). Optimum strata boundaries for shellfish surveys, Biometrics , 41: 1053-1062.

[62] Sai , S. and Taguri , M. (1989).. Optimum stratification based on a concomitant variable and its application to the current statistics of commerce. Rep. Statist. Appl. Res. JUSE ,36: 22-31.

[63] Mahajan ,P.K. Gupta , J.P. and 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.

[64] Wywial , J. (1995).. On optimum stratification of population on the basis of auxiliary variables. Statist. Transition , 2: 831-837.

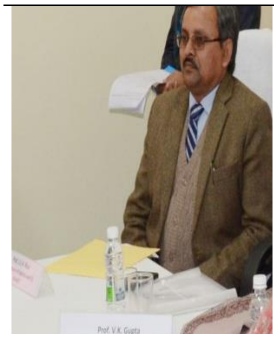
[65] Rizvi, S.E.H., Gupta, J.P. and Singh, R. (2000). Approximately optimum stratification for two study variables using auxiliary information. Journal of the Indian Society of the Agricultural Statistics , 53 (3):287-298.

[66] 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.

[67] Kozak, M. ,Verma, M.R. and Zielinski , A. (2007). Modern approach to optimum stratification: Review and perspectives. Statistics in Transition, 8(2): 223-250.

[68] Horgan , J.M. (2010). Chossing the strata boundaries: An elusive optima .Istanbul universities iisletmefakuitesi Dergisi ,39(2): 195-204.

[69] Mehta, S. and Mandowara, V.L. (2012)..An optmum stratification for stratified cluster sampling design when clusters are of varying sizes. International journal of scientific and technology research, 1(9): 74-79.



S. E. H Rizvi, presently working as Professor and Head, Division of Statistics and Computer Science, Faculty of Basic Sciences at Sher-e-Kashmir University of Agricultural sciences and Technology Jammu, J&K India.

He has around thirty years of teaching experience. He has published many research papers in reputed internal as well national journals. His research interests are Sampling theory, Mathematical Programming, Applied Statistics, Regression analysis and Statistical methods. He has been advisory committee of more than hundred Post Graduate students of Agriculture and Allied Science and is also guiding a number of M.Sc /Ph.D students. He has delivered various lectures/talks in several trainings / workshops /conferences.

Faizan Danish is pursuing Ph.D in Statistics at Division of Statistics and Computer Science, Faculty of Basic Sciences in Sher-e-Kashmir University of Agricultural sciences and Technology Jammu, J&K India.

He has done his M.Sc from Department of Statistics, University of Kashmir Srinagar, J&K India. He has received best paper awards too from many journals. Not only this but he has also attended a good number of workshops/trainings related to his research. His research interests are Sampling theory, Mathematical Programming, Applied Statistics and Inference. He has published research articles in reputed international/ national journals of Statistics and Mathematics. He has presented a number of papers in several national as well as international conferences.