Applied Mathematics & Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/amisl/060105

Construction of Optimum Strata Boundaries for Uniform and Exponential Auxiliary Variable

*Faizan Danish**, *S. E. H. Rizvi, Manish Kr. Sharma and M. Iqbal Jeelani*

Division of Statistics and Computer Science, Faculty of Basic Sciences SKUAST-Jammu, Main campus, Chatha-180009 (J&K), India.

Received: 17 Jan. 2017, Revised: 8 May 2017, Accepted: 10 May 2017
Published online: 1 Jan. 2018

Abstract: 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 which is possible only when we have the best way of stratification. In this paper, we discuss the problem of determining the OSB of a study variable based on auxiliary variable under proportional allocation. The problem is formulated as a Mathematical Programming Problem (MPP) which minimizes the objective function subject to the constraint that the sum of the widths of all the strata is equal to the whole range of the distribution. The distributions of the auxiliary variable are considered continuous with uniform and exponential density functions separately. The formulated MPPs, which turn out to be multistage decision problems, could then be solved using dynamic programming approach. The theoretical results have been illustrated empirically for both the distributions which show that there is substantial gain in efficiency with the increase in the number of strata.

Keywords: Optimum stratification, Mathematical programming approach, Dynamic programming technique, Proportional allocation.

1 Introduction

Stratification is an approach usually used in sample surveys for heterogeneous population under study through which it is divided into homogeneous sub-populations known as strata. One of the reasons for stratification is that it may produce a gain in precision in the estimates of the characteristics of the population under study. This paper will focus on the gain in the precision due to stratification using auxiliary variable as the basis of stratification. The basic points in stratified sampling about which care must be taken in order to increase precision are Choosing stratification variable, Number of strata, Determination of points of stratification, Allocation of units from each stratum and Choosing of sampling design for selection of units in each stratum. Stratification should be done on the basis of study variable itself, however, it might be possible that information on study variable is not known. In such cases the variable which is highly correlated with the study variable, known as auxiliary variable, could be used as stratification variable.

The optimum stratification of a population is defined as dividing the joint domain of the stratification variable in such a way that the precision of the estimate is maximized. Considering the determination of optimum strata boundaries (OSB) as an important problem in survey sampling, many authors developed different methods to solve it with varying degrees of mathematical rigour. It was first studied by Dalenius [1] who obtained the minimal equations which were very difficult to solve because of their implicit nature. Dalenius and Gurney [2] showed that in some cases increasing the number of strata can lead to a loss in precision if stratification is not chosen carefully. Using approximation rules other classical methods were determined by Mahalanobis [3] Hansen and Hurwitz [4] and Aoyoma [5]. Dalenius [6], Taga [7], Singh and Sukhatme [8], Singh [9], Singh and Prakash [10] ,Rizvi *et al*. [11] and Gupta *et al*. [12] are some of the other authors who used frequency distribution of auxiliary variable and come up with different approximation methods for determining OSB. Khan *et al*, [13] proposed the technique of determining OSB for the study variable with different frequency functions through formulate the problem as

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

equivalent of determining optimum strata width (OSW), using non linear programming problem (NLPP) and finally solved it by dynamic programming approach. Gunning and Horgan [14] proposed an alternative approach to approximate stratification based on a geometric progression and the assumption of uniform stratification variable within strata.Kozak and Verma [15] has proved that the optimization approach is more efficient than the approximate stratification, and disputed applying approximate stratification procedures in order to obtain final stratification points.. Khan *et al*. [16] determined OSB using auxiliary information and formulated the problem as a mathematical programming problem when the number of strata is fixed in advance. By suitable transformation they converted the problem into a multistage decision problem and solve it using dynamic programming problem.Also Khan *et al.* [17] proposed a technique in stratified sampling design for economic surveys based on auxiliary information which can be used for constructing optimum stratification and determining optimum sample allocation to maximize the precision in estimate. In this paper, a technique using dynamic programming approach has been developed to determine the OSB when the frequency distribution of the auxiliary variable is used as basis for stratification under proportional allocation. Section 3 provides the detailed procedure for finding OSW as an NLPP and the procedure for solution of it is discussed. Under numerical illustration the results obtained are presented in table 1 and table 2.

2 The problem of OSB

Let the heterogeneous population under study be divided into L strata based on a single auxiliary X in order to estimate the mean of the study variable Y under interest. Suppose that from h^{th} stratum (h=1,2,...,L), containng N_h units , a sample of n_h units is drawn using simple random sampling. Then the stratified sample mean, $\overline{y}_{st}$,is given by

$$\overline{y}_{st} = \sum_{h=1}^{L} W_h \overline{y}_h \tag{2.1}$$

where,$W_h = N_h / N$ is the proportion of the population which is contained in the hth stratum and $\overline{y}_h$ is the sample mean of the hth stratum.Under proportional allocation, the $V(\overline{y}_{st})$ is given as

$$V(\overline{y}_{st}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{h=1}^{L} W_h \sigma_{hy}^2 \tag{2.2}$$

where $\sigma_{hy}^2 = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (y_{hj} - \mu_h)$ is the variance of the hth stratum;h=1,2,...L and n is the total sample size selected from the whole population.

Thus, if the finite population correction (fpc) is ignored then (2.2) can be written as

$$V(\overline{y}_{st}) = \frac{1}{n} \sum_{h=1}^{L} W_h \sigma_{hy}^2 \tag{2.3}$$

Thus, our aim is to minimize (2.3),or equivalently to minimize

$$\sum_{h=1}^{L} W_h \sigma_{hy}^2 \tag{2.4}$$

Now, let the study variable has the regression model of the form

$$y = \Phi(x) + e \tag{2.5}$$

where $\Phi(x)$ could be either linear or non-linear function of x and e is the error term such that $E(e/x) = 0$ and $V(e/x) = \Psi(x)$ which is positive for all values of x.

Thus under above model, the stratum mean μ_{hy} and the stratum variance σ_{hy}^2 can be expressed as

$$\mu_{hy} = \mu_{h\Phi} \tag{2.6}$$

And

$$\sigma_{hy}^2 = \sigma_{h\Phi}^2 + \mu_{h\Psi}^2 \tag{2.7}$$

where $\mu_{h\Phi}$ and $\mu_{h\Psi}$ are the expected values of function of $\Phi(x)$ and $\Psi(x)$ respectively and $\sigma_{h\Phi}^2$ represents the variance of $\Phi(x)$ in the hth stratum.However,if Φ and e are not correlated then σ_{hy}^2 in (2.5) can be expressed as T.Dalenius et al. (1951) has done as

$$\sigma_{hy}^2 = \sigma_{h\Phi}^2 + \sigma_{he}^2 \tag{2.8}$$

where σ_{he}^2 denotes the variance of the hth stratum of e and also it could be verified that (2.7) and (2.8) are equal.

Let f(x) be the probability density function of the auxiliary variable X defined in [a,b] used as the basis for stratification. If the population mean of the study variable y is estimated under proportional allocation then the problem of determining OSB is to partition the range of X into L strata as $[a, x_1]$, $[x_1 , x_2]$, ...,$[x_{L-1} , b]$ such that $a = x_1 \le x_2 \le \cdots \le x_{L-1} \le x_L = b$, in such a way (2.5) is minimum.

Now let

$$x_L - x_0 = d \tag{2.9}$$

Using (2.7) in (2.4), we need to minimize

$$\sum_{h=1}^{L} W_h (\sigma_{h\Phi}^2 + \sigma_{he}^2) \tag{2.10}$$

When the functions f (x) ,Φ (x) and Ψ (x) are known and are imtegrable, then W_h , σ_{hy}^2 and σ_{he}^2 can be expressed as the function of the strata boundary points.Thus, we can write

$$W_h = \int_{x_{h-1}}^{x_h} f(x)\,dx \tag{2.11}$$

$$\sigma_{h\Phi}^2 = \frac{1}{W_h} \int_{x_{h-1}}^{x_h} \Phi^2(x) f(x)\,dx - \mu_{h\Phi}^2 \tag{2.12}$$

and

$$\sigma_{he}^2 = \frac{1}{W_h} \int_{x_{h-1}}^{x_h} \Psi(x) f(x)\,dx \tag{2.13}$$

where $$\mu_{h\Phi} = \frac{1}{W_h} \int_{x_{h-1}}^{x_h} \Phi(x) f(x)\,dx \tag{2.14}$$

where Wh , $\mu_{h\Phi}$ and $\sigma^2_{h\Phi}$ are the stratum weight , stratum mean and stratum variance, respectively for hth stratum with boundary points (xh-1 , xh) are the boundary points. Thus, objective function (2.10) can be expressed as the function of boundary points x_h and x_{h-1} as given below

Let

$$\Psi (x_h , x_{h-1}) = W_h (\sigma^2_{h\Phi} + \sigma^2_{he})$$

Then the optimization problem for determining the OSB can be expressed as follows Minimize

$$\sum_{h=1}^{L} \Psi (x_h , x_{h-1})$$

Subject to constraint

$$x_0 \le x_1 \le x_2 \le, \dots, \le x_{L-1} \le x_L \tag{2.15}$$

Let us define

$$V_h = x_h - x_{h-1} \ ; h = 1,2,\dots,L \tag{2.16}$$

where $V_h \ge 0$ denotes the width of the hth stratum

Thus (2.9) can be expressed as a function of stratum width

$$\sum_{h=1}^{L} V_h = \sum_{h=1}^{L} (x_h - x_{h-1})$$

$$= b\text{-}a = x_L - x_0 = d \tag{2.17}$$

Thus kth stratification point x_k ; (k = 1,2,...,L-1) can be expressed as

$$x_k = x_0 + v_1 + v_2 + \dots + v$$

$$= x_{k-1} + v_k$$

which is the function of the kth stratum and (k-1)th stratum boundaries.

Considering (2.17) as a constraint, then (2.15) can be treated as similar problem of obtaining OSW as V_1 , V_2 ,...,V_L and is expressed as fallowing optimization problem

Minimize $\sum_{h=1}^{L} \Psi_h (x_h , x_{h-1})$

subject to constraint

$$\sum_{h=1}^{L} V_h = d \tag{2.18}$$

and $V_h \ge 0$; h= 1,2,...,L

Initially,x_0 is known.Therefore,the first term $\Psi_h (V_1 , x_0)$ in the objective function of (2.18) is a function of V_1 only. The next stratification point $x_1 = x_0 + V_1$ will be known once V_1 is known and the second term of the above objective function, that is, $\Psi_2 (V_2 , x_1)$ will be the function of V_2 alone.

Then, as stating the above objective function as a function of V_h alone. Thus we can rewrite the NLPP (2.18) as

Minimize $\sum_{h=1}^{L} \Psi_h (V_h)$

subject to constraint

$$\sum_{h=1}^{L} V_h = d \tag{2.19}$$

and

$$V_h \ge 0; \ h= 1,2,\dots,L$$

3 The Solution Procedure

It is obvious that (2.19) is a NLPP which could be solved as multistage decision problem using dynamic programming approach. Dynamic programming is an approach which determines the optimal solution of the multi variable problem by splitting it into different stages in which each stage is comprising a sub-problem consisting of single variable.Dynamic programming model is actually a recursive equation based on principle of optimality given by Bellman [18].This principle of optimality connects the different stages of the given problem in such a manner which guarantees for optimal feasible solutions at each stage as also optimal and feasible solutions for the whole problem(Taha,[19]).

Now, consider a sub-problem of (2.19) for first m (< L) strata. Then (2.19) can be rewritten as

Minimize $\sum_{h=1}^{m} \Psi_h (V_h)$

subject to constraint

$$\sum_{h=1}^{m} V_h = d_m \tag{3.1}$$

And

$$V_h \ge 0; \ h= 1,2,\dots,m$$

where d_m < d is the total width available to us from the division into first m strata. Note that $d_m = d$ for m = L

The transformation functions could be explained

$$d_m = V_1 + V_2 + \cdots + V_m$$

$$d_{m-1} = V_1 + V_2 + \cdots + V_{m-1} = d_m - V_m$$

$$d_{m-2} = V_1 + V_2 + \cdots + V_{m-2} = d_{m-1} - V_{m-1}$$

.

.

.

$$d_2 = V_1 + V_2 = d_3 - V_3$$

$$d_1 = V_1 = d_2 - V_2$$

Let the minimum value of the objective function obtained in equation (3.1) be denoted by $\varphi_h (V_h)$,then using Bellman's principle of optimality we can write a forward recursive equation of dynamic programming as

$$\varphi_m (V_m) = \min_{0 \le V_m \le d_m} [\Psi_m (V_m) + \Psi_{m-1}(d_m - V_m)], \ m \ge 2 \tag{3.2}$$

similarly, for $m \ge 3$

$$\varphi_{m-1}(V_{m-1}) = \underset{0\le V_{m-1}\le d_{m-1}}{\text{Min}} [\Psi_{m-1}(V_{m-1}) + \Psi_{m-2}(d_{m-1} - V_{m-1})], \quad (3.3)$$

Now, for the first stage ,that is, m = 1

$$\varphi_1(V_1) = \Psi_1(V_1) = V_1^* = d_1 \quad (3.4)$$

Equation (3.2) ,(3.3) and (3.4) when solved recursively for each m = 1,2, ...,L and $0 \le d_m \le d$, $\varphi_L(d)$ is obtained. From $\varphi_L(d)$ the optimum width of the L[th] stratum, that is, V_L^* is obtained. Similarly for $\varphi_{L-1}(d - V_1^*)$,that is, V_{L-1}^*, is obtained and continue this procedure till V_1^* is obtained.

4 Determination of OSB in Case Of Uniform Distribution

4.1 Formulation of the problem for uniformly distributed auxiliary variable

Let the auxiliary variable X follows uniform distribution with pdf given as

$$g(x)=\begin{cases} \frac{1}{b-a} & , \quad a \le x \le b \\ 0 & , \quad otherwise \end{cases} \quad (4.1.1)$$

Where ‘a’ denotes the location parameter and the difference of ‘b’ and ‘a’ is the scale parameter.

Using it in (2.11) and (2.12),we get

$$W_h = \frac{V_h}{b-a} \quad (4.1.2)$$

and

$$\sigma_{h\Phi}^2 = \frac{V_h^2}{12} \quad (4.1.3)$$

Let the given model be linear, then the linear regression model be

$$\Phi(x) = a + bx \quad (4.1.4)$$

Thus (2.10) can be written as

$$\sum_{h=1}^{L} W_h(\sigma_{hx}^2 + \sigma_{he}^2) \quad (4.1.5)$$

substituting values obtained in equation (4.1.2) and (4.1.3) in equation (4.1.5),we get NLPP as

$$\text{Minimize} \quad \sum_{h=1}^{L} \frac{V_h}{b-a}\left(\beta^2 \frac{V_h^2}{12} + \sigma_{eh}^2\right)$$

subject to constraint

$$\sum_{h=1}^{L} V_h = d \quad (4.1.6)$$

and $V_h \ge 0$; h= 1,2,...,L

where β is the regression coefficient and d is the range of the distribution.

4.2 Variance of error term and formulation of problem

Table 1: OSW, OSB and optimum value of the objective function for uniform distribution

No. of strata L	Strata width V_h^*	Strata boundary points $x_h^* = x_{h-1}^* + V_h^*$	Optimum value of objective function $\sum_{h=1}^{L} W_h\sigma_h^2$
2	V_1^*= 0.5000 V_2^*= 0.5000	$x_1^* = x_0 + V_1^* =$ 1.5000	1.1259
3	V_1^*= 0.3333 V_2^*= 0.3333 V_3^* = 03333	$x_1^* = x_0 + V_1^* =$ 1.3333 $x_2^* = x_1^* + V_2^*$ = 1.6666	1.0634
4	V_1^*= 0.2500 V_2^*= 0.2500 V_3^* = 0.2500 V_4^* = 0.2500	$x_1^* = x_0 + V_1^*$ = 1.2500 $x_2^* = x_1^* + V_2^*$ = 1.5000 $x_3^* = x_2^* + V_3^*$ = 1.7500	1.0356
5	V_1^*= 0.2000 V_2^*= 0.2000 V_3^* = 0.2000 V_4^* = 0.2000 V_5^* = 0.2000	$x_1^* = x_0 + V_1^*$ = 1.2000 $x_2^* = x_1^* + V_2^*$ = 1.4000 $x_3^* = x_2^* + V_3^*$ =1.6000 $x_4^* = x_3^* + V_4^*$ = 1.8000	1.0127
6	V_1^*= 0.1666 V_2^*= 0.1667 V_3^* = 0.1666 V_4^* = 0.1667 V_5^* = 0.1666 V_6^* = 0.1667	$x_1^* = x_0 + V_1^*$ = 1.1666 $x_2^* = x_1^* + V_2^*$ =1.3333 $x_3^* = x_2^* + V_3^*$ =1.4999 $x_4^* = x_3^* + V_4^*$ = 1.6665 $x_5^* = x_4^* + V_5^*$ = 1.8332	1.0062

It is assumed in the regression model (2.5) that the variance of the error term is V(e/x) = Ψ(x) for all x in the range [a,b] and the expectation of Ψ(x) is σ_{he}^2 is obtained by (2.13).Many authors like Singh and Sukhatme (1969), Rizvi *et al*. (2000) and Khan *et al*.(2009) have assumed that Ψ(x) may be of the form

$$\Psi(x) = c\,x^g \;;\; c > 0,\; g \ge 0 \quad (4.2.1)$$

where c and g are constants and the value of g lies in [0,2].substituting value of W_h , f(x) and Ψ(x) in (2.13), we get

$$\sigma_{he}^2 = \frac{cV_h^{g+1}}{V_h\,(b-a)(g+1)} \quad (4.2.2)$$

Thus,(4.2.2) shows that σ_{he}^2 can be obtained only once the constants are known. Thus substituting (4.2.2) in (4.1.5),we hav

$$\text{Minimize} \quad \frac{V_h}{b-a}\left(\frac{\beta^2 V_h^2 (b-a)(g+1)+12c}{12\,(b-a)(g+1)}\right)$$

subject to constraint

$$\sum_{h=1}^{L} V_h = d \quad (4.2.3)$$

And $V_h \ge 0$; h= 1,2,...,L

we know (h-1)th stratification point is as $x_{h-1} = d_h - V_h$

By substituting this value, the recurrence relation (3.2) and (3.4) can be written as

For first stage where m = 1 and at $V_1^* = d_1$

$$\Psi_1(V_1) = \frac{1.44d_1^3 + 12d_1}{12} \quad (4.3.2)$$

For m ≥ 2

$$\Psi_m(V_m) = \min_{0\le V_m \le d_m}\left[\frac{1.44V_m^3 + 12V_m}{12} + \Psi_{m-1}(d_m - V_m)\right] \quad (4.3.3)$$

Solving equations (4.1.2) and (4.3.3), the NLPP of (4.4.1) is easily solved while running the computer programme OSW(V_h^*) and OSB (x_h^*) are obtained and the results are presented in Table **1** for different number of strata L = 2 , 3 , 4 , 5 , 6.

5 Determination of OSB under exponential auxiliary variable

5.1 Formulation of problem from exponential auxiliary variable

Let the auxiliary variable X fallows exponential distribution with pdf as

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & , \quad x \ge 0, \theta > 0 \\ 0 & , \quad otherwise \end{cases} \quad (5.1.2)$$

By substituting (5.1.1) in (2.11) and (2.12),we get

$$W_h = e^{\frac{-y_{h-1}}{\theta}}(E) \quad (5.1.2)$$

$$\sigma_{hx}^2 = \frac{(\theta E)^2 - V_h^2(1-E)}{(E)^2} \quad (5.1.3)$$

where $E = 1 - e^{\frac{-V_h}{\theta}}$

Using equation (5.1.2) and (5.1.3) in equation (4.1.5) while in case of exponential auxiliary variable, we get

$$\text{Minimize} \quad \sum_{h=1}^{L} e^{\frac{-y_{h-1}}{\theta}}(E)[\beta^2 \frac{(\theta E)^2 - V_h^2(1-E)}{(E)^2} + \sigma_{he}^2]$$

Subject to constraint

$$\sum_{h=1}^{L} V_h = d \quad (5.1.4)$$

and

$$V_h \ge 0;\ h = 1,2,...,L$$

where β denotes regression coefficient and d = $x_L - x_0$ is the range of the distribution and σ_{he}^2 denotes the variance of the error term given in (2.13) from the error term in the regression model (2.5).

5.2 Estimating the variance of error term

Assuming the same condition for exponential auxiliary variable as assumed for uniform auxiliary variable in equation (4.2.1), we can write the variance of the error term as

$$\sigma_{he}^2 = c \qquad \text{for } g = 0 \quad (5.2.1)$$

Thus,(5.2.1) shows that σ_{he}^2 can be obtained when the value of the constant c is known. Substituting (5.2.1) in (5.1.4) , we get

$$\text{Minimize} \quad \sum_{h=1}^{L} e^{\frac{-y_{h-1}}{\theta}}(E)[\beta^2 \frac{(\theta E)^2 - V_h^2(1-E)}{(E)^2} + c]$$

Subject to constraint

$$\sum_{h=1}^{L} V_h = d \quad (5.2.2)$$

And

$$V_h \ge 0;\ h = 1,2,...,L$$

where β denotes regression coefficient and d = $x_L - x_0$ indicates the range of the distribution and σ_{he}^2 is the variance of the error term defined in (2.13) from the error term in the regression model given in equation (2.5).

5.3 Numerical illustration for exponential auxiliary variable

For Numerical illustration the method for obtaining solution has already been discussed in section 3 using dynamic programming approach to get OSB.By substituting values of c = 1, θ=1 and β = 1.2 in (5.2.2) , we get

$$\text{Minimize} \quad \sum_{h=1}^{L} e^{-y_{h-1}}(E_1)[2.44 + V_h^2(1-E_1)]$$

Subject to constraint

$$\sum_{h=1}^{L} V_h = 1 \quad (5.3.1)$$

and

$$V_h \ge 0;\ h = 1,2,...,L$$

where $E_1 = 1 - e^{-V_h}$

we know that x_{h-1} denotes (n-1)th stratification point then the recurrence relation (3.2) and (3.4) can be written as;

For first stage, when m = 1

$$\Psi_1(V_1) = (1 - e^{-d_1})(2.44 + d_1^2 e^{-d_1}) \quad (5.3.2)$$

For the stage, when m ≥ 2

$$\Psi m(Vm) = \min_{0\le V_m \le d_m}[e^{-y_{m-1}}(1 - e^{-V_m})(2.44 + V_m^2 e^{-V_m})] \quad (5.3.3)$$

The solution of the NLPP (5.3.1) can be obtained by resolving the recurrence relations (5.3.2) and (5.3.3) and would get the desired results of OSW and OSB.The obtained results are presented in Table 2 for different values of L = 2 , 3 , 4 , 5 and 6.

Table 2: Osw, Osb and optimum value of the objective function for exponential distribution

No. of strata L	Strata width V_h^*	Strata boundary points $x_h^* = x_{h-1}^* + V_h^*$	Optimum value of objective function $\sum_{h=1}^{L} W_h\sigma_h^2$

2	V_1^*= 1.3725 V_2^*= 18.6275	$x_1^* = x_0 + V_1^*$ = 1.3725	0.6249
3	V_1^*= 0.8592 V_2^*= 1.3561 V_3^* = 17.7847	$x_1^* = x_0 + V_1^*$ = 0.8592 $x_2^* = x_1^* + V_2^*$ = 2.2153	0.4137
4	V_1^*= 0.6233 V_2^*= 0.8952 V_3^* = 1.4272 V_4^* = 17.0543	$x_1^* = x_0 + V_1^*$ = 0.6233 $x_2^* = x_1^* + V_2^*$ = 1.5185 $x_3^* = x_2^* + V_3^*$ = 2.9457	0.3724
5	V_1^*= 0.5381 V_2^*= 0.7234 V_3^* = 0.9342 V_4^* = 1.4923 V_5^* = 16.3120	$x_1^* = x_0 + V_1^*$ = 0.5381 $x_2^* = x_1^* + V_2^*$ = 1.2615 $x_3^* = x_2^* + V_3^*$ =2.1957 $x_4^* = x_3^* + V_4^*$ = 3.6880	0.2478
6	V_1^*= 0.4231 V_2^*= 0.5013 V_3^* = 0.7394 V_4^* = 0.9423 V_5^* = 1.4321 V_6^* = 15.9618	$x_1^* = x_0 + V_1^*$ = 0.4231 $x_2^* = x_1^* + V_2^*$ =0.9244 $x_3^* = x_2^* + V_3^*$ =1.6638 $x_4^* = x_3^* + V_4^*$ = 2.6061 $x_5^* = x_4^* + V_5^*$ = 4.0382	0.1829

6 Conclusion

In this paper, we proposed the technique of determining the OSB while using auxiliary variable as stratification variable. OSB have been obtained when the frequency distribution of the auxiliary variable is uniformly distributed and also for the case when it is exponentially distributed. The problem is formulated as NLPP that seek minimization of the estimated population parameter under proportional allocation. The NLPP is then solved by using dynamic programming approach. The main advantage of dynamic programming is that it can determine OSB very efficiently when the density function of the stratification variable is known, and it also gives the global minimum of the objective function. Numerical illustration of the proposed method has also been done and the results obtained for uniform auxiliary variable and for exponential auxiliary variable are presented in Table I and Table II respectively. In both the cases ,the numerical results reveal that variance decreases substantially with the increase in number in strata.

References

[1] Dalenius,T.(1950).The problem of optimum stratification. Skandinavisk Aktuarietidskrift. 33,203-213.

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

[3] Mahalanobis, P.C. (1953). Some aspect of design of sample surveys. Sankhya.12 ,1-17.

[4] Hansen, M. H.,Hurwitz, W. N. and Madow, W.G. (1953).Sample survey methods and theory.Vol. I & II ,John Wiley and Sons,Inc.,New York.

[5] Aoyama, H. (1954): A study of stratified random sampling. Ann. Ins. Stat. Math. 6,1-36.

[6] Dalenius, T. (1957). Sampling in Sweden. Almqvist & Wiksell.Stockholm.

[7] Taga, Y. (1967).On optimum stratificationfor the objective variable based on concomitant variables using prior information. Annals of the institute of Statistical Mathematics. 19, 101-129.

[8] Singh, R., and Sukhatme, B. V. (1969).Optimum stratificatrion. Annals of the institute of Statistical Mathematics. 21(3), 515-528.

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

[10] Singh, R., and Prakash , D. (1975). Optimum stratification for equal allocation. Annals of the institute of Statistical Mathematics. 27, 273-280.

[11] Rizvi, S.E.H., Gupta, J.P. and Singh, R. (2002).Approximately optimum stratification for two study variable using auxiliary information. Journal of Indian Society of Agricultural Statistics.53(3), 287-298.

[12] Gupta, R. K., Singh, R. And Mahajan, P. K. (2005).Approximate optimum strata boundaries for ratio and regression estimators.Aligarh Journal of Statistics. 25, 49-55.

[13] 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.

[14] Gunning, P. and Horgan, J. M. (2004).A new algorithm for the construction of stratum boundaries in skewed populations. Survey Methodology. 30(2), 159-166.

[15] Kozak, M. and Verma, M. R. (2006). Geometric versus optimization stratification: a comparison of efficiency. Survey Methodology 32(2),157-163.

[16] Khan,M.G.M.,Ahmad,N.and Khan,S.(2009). Determination the optimum stratum boundaries using mathematical programming. J. Math. Model. Algor. 8,409-423.

[17] 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):2080-209.

[18] Bellman, R.(1957).Dynamic programming. Princeton University Press Princeton,New Jersey.

[19] Taha, H. A. Operations research: An introduction. 9th edition. Pearson.

[20] Cochran, W.G. (1977). Sampling technique, 3rd ed. New York: John Wiley and Sons,Inc.

[21] Abramowitz, M. and Stegun, I. A. (1972), Handbook of Mathematical functions, Dover New York.

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. 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.

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.

Manish Kr. Sharma is working as Associate Professor (Statistics) at Sher-e-Kashmir university of Agricultural Sciences & technology, Jammu. He is recipient of gold medal at M.Sc level and has 17 years of teaching and research experience. Alongside teaching he is working in many projects. He has reputed number of research papers in various international & national journals. His research interests are Sampling theory, Regression theory, and Applied Statistics. He has also delivered more than a hundred invited lectures in reputed institution and presented research papers in more than two dozen international and national conferences.

M. Iqbal Jeelani Bhat is presently working as Assistant Professor, 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 a good number of Research publications in many reputed journals. He has been awarded as gold medal at M.Sc level for topping the batch at the concerned University. He has delivered various lectures/talks in several trainings / workshops /conferences. His thrust areas of research are Sampling Theory, Applied Statistics and Mathematical Programming. He is expert in many statistical software's.