APPROXIMATE OPTIMUM STRATA BOUNDARIES FOR PROPORTIONAL ALLOCATION USING RANKED SET SAMPLING

Khalid Ul Islam Rather[1]*, S.E.H Rizvi[2], Manish Sharma[3], M. Iqbal Jeelani[4], Faizan Danish[5]

•

Division of Statistics and Computer Science, SKUAST-Jammu, India[1,2,3]
Division of Social and Basic Sciences Faculty of Forestry, SKUAST-Kashmir, India [4]
Department of Mathematics, School of Advanced Science(SAS) VIT-AP India[5]
khalidstat34@gmail.com
sehrizvistat@gmail.com
manshstat@gmail.com
jeelani.miqbal@gmail.com
danishstat@gmail.com

Abstract

Ranked set sampling is an approach to data collection originally combines simple random sampling with the field investigator's professional knowledge and judgment to pick places to collect samples. Alternatively, field screening measurements can replace professional judgment when appropriate and analysis that continues to stimulate substantial methodological research. The use of ranked set sampling increases the chance that the collected samples will yield representative measurements. This results in better estimates of the mean as well as improved performance of many statistical procedures. Moreover, ranked set sampling can be more cost-efficient than simple random sampling because fewer samples need to be collected and measured. The use of professional judgment in the process of selecting sampling locations is a powerful incentive to use ranked set sampling. Optimum stratification is the method of choosing the best boundaries that make strata internally homogeneous, given some sample allocation. In order to make the strata internally homogenous, the strata should be constructed in such a way that the strata variances for the characteristic under study be as small as possible. This could be achieved effectively by having the distribution of the study variable known 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 (auxiliary information) obtained at a recent study. The present investigation deals with paper the problem of optimum stratification on an auxiliary variable for proportional allocation under ranked set sampling (RSS), when the form of the regression of the estimation variable on the stratification variable given the variance function is known. A cum rule of finding approximately optimum strata boundaries has been developed. Further, empirical study has been made and presented along with relative efficiency which showed remarkable gain in efficiency as compared to unstratified RSS.

Keywords: Ranked set Sampling, approximately optimum strata boundaries, auxiliary variable optimum strata width

1. Introduction

The main aim of stratification is to produce estimators with more precision when a population characteristic is under study. The problem of obtaining optimum strata boundaries (OSB) taking study variable itself as stratification variable was first considered [1]. The study showed that for a given method of allocation, the variance is clearly a function of the strata boundaries. By minimizing the variance of the estimate of population mean sets of equations were obtained, solutions to which gave optimum strata boundaries. Due to the implicit nature of the minimal equations, their exact solutions could not be obtained. Subsequently, various authors gave methods of obtaining approximations to the exact solutions of the minimal equations. However, the ideal situation is that the distribution of such a study variable is known and the OSB can be determined by placing boundaries on the range of this distribution at suitable cut points. For an excellent account of these investigations reference may be made [2]. When the information about the study variables are not known, the utilization of auxiliary variable as stratification variable may be considered and obtained approximate OSB under simple random sampling (SRS) studied [3]. The problem of optimum stratification for two characters under study using auxiliary information [4]. A methodology developed for obtaining AOSB using auxiliary information under compromise method of allocation [5]. The problem of obtaining OSB under proportional allocation with varying cost of each unit studied [6]. A technique proposed under Neyman allocation when the stratification is done on the two auxiliary variables under consideration [7]. Recently, the problem considered of optimum stratification for a model-based allocation under a super population model [8]. Situation considered of optimum stratification of heteroscedastic populations in stratified selection for a known allocation under SRS strategy [9]. Estimator for the population Mean under Ranked Set Sampling [10] and [11]. Situation of optimum stratification under RSS considered by [12] and [13]. Further, the selection of sample from each stratum could be taken by utilizing any sampling technique. Therefore, in the present investigation we have taken the procedure of ranked set sampling (RSS) as a method of selection of units from each stratum, which is more efficient as compared to SRS. A stratified ranked set sample (SRSS) is a sampling plan in which a population is divided into mutually exclusive strata and a RSS of n elements is quantified within each stratum. The sampling is performed independently across the strata. Therefore, one can think of an SRSS scheme as a collection of L separate ranked set samples. Originally, RSS was first suggested to estimate mean pasture and forage yields [14]. The necessary mathematical theory provided [15].

Let the population under consideration be divided into L strata and a sample $n_{0h} = (R_h \times n_h)$ units which is selected from h^{th} stratum is drawn using RSS, where R_h is the number of cycles and n_h is sample size of each cycle. Each sample element is measured with respect to some variable Y, and estimator of the population mean is given by

$$\bar{y}_{SRSS} = \sum_{h=1}^{L} \frac{W_h}{n_{0h}} \left[\sum_{j=1}^{R_h} \sum_{i=1}^{n_h} \bar{y}_{ij(r)} \right] \tag{1.1}$$

where W_h is the weight of the h^{th} stratum and $\bar{y}_{ij(r)}$ is the sample mean based on n_{0h} units drawn from the h^{th} stratum.

If the finite correction is ignored, the variance of the estimate will be:

$$V\left(\bar{y}_{SRSS}\right)_{prop} = \frac{1}{n}\left[\sum_{h=1}^{L} W_h \sigma^2_{h(r)}\right] \tag{1.2}$$

$\sigma^2_{h(r)} = \left(\sigma^2{}_{hc} - \frac{1}{n}\left(\mu_i - \mu\right)^2\right)$ denotes the variance of r^{th} order statistics in h^{th} stratum of the random sample of size n_h.

In most of these investigations related to optimum stratification, both the estimation and stratification variables are taken to be the same. Since the distribution of the estimation variable $'Y'$ is rarely known in practice, it is desirable to stratify on the basis of some suitably chosen concomitant variable $'X'$. An investigation has considered the general problem of optimum stratification based on auxiliary variables for the case of proportional allocation [16]. We consider the problem of optimum stratification on the auxiliary variable 'X', assuming knowledge about the form of the regression of 'Y' on 'X' and the variance function $V(y \mid x)$, minimal equations giving optimum strata boundaries have been obtained for proportional allocations under ranked set sampling. Since these equations cannot be solved easily, various methods of finding approximations to the exact solutions have been given.

In this paper, the problem of construction of strata boundaries will be dealt using classical approach when the sample is selected from the strata using RSS.

MINIMAL EQUATIONS UNDER PROPORTIONAL ALLOCATION

If the regression of the estimation variable 'Y' on the stratification variable 'X', in the infinite super population is given by

$$y = c(x) + e \tag{2.1}$$

Where '$c(x)$' is a function of auxiliary variable, 'e' is the error term such that $E(e \mid x) = 0$ and $V(e \mid x) = \eta(x) > 0 \ \forall x \in (a,b)$ with $(b-a) < \infty$. Let $f(x,y)$ and $f(x)$ be the joint density function and marginal density function of (x,y) and x respectively. Then, we have

$$W_h = \int_{x_{h-1}}^{x_h} f_i(x) \ , \ \mu_{hc} = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} c(x) f_i(x) \ \text{ and } \ \sigma^2{}_{hy} = \sigma^2{}_{hc} + \mu_{h\eta} \ , \ (h = 1,2,3\ldots,L) \tag{2.2}$$

where (x_{h-1}, x_h) are lower and upper boundaries of the h^{th} stratum with $(x_0 = a)$ and $(x_L = b)$, $\mu_{h\eta}$ is the expected value of $\eta(x)$ and σ^2_{hc} is the variance of $c(x)$ in the h^{th} stratum.

Using these relations, the variance expression (1.2), under proportional allocation

$$V\left(\bar{y}_{SRSS}\right)_{prop} = \frac{1}{n}\left[\sum_{h=1}^{L} W_h \sigma^2_{hc(r)} + \mu_{h\eta}\right] \tag{2.3}$$

Let $[x_h]$ denote the set of optimum points of stratification on the range (a,b), for which the $V\left(\bar{y}_{SRSS}\right)$ is minimum. These points $[x_h]$ are the solutions of the minimal equations which are obtained by equating to zero the partial derivatives of $V\left(\bar{y}_{SRSS}\right)$ with respect to $[x_h]$.

Minimization of the variance expression given in (2.3), is equivalent to the minimization of the expression $\sum_{h=1}^{L} W_h \sigma^2_{hc(r)}$, since $\sum_{h=1}^{L} W_h \mu_{h\eta} = \mu_{h\eta}$ is a population parameter and is a constant. On equating to zero the partial derivative of this expression with respect to $[x_h]$, we get

$$\frac{W_h f(x_h)\left[\left(c(x_h) - \mu_{hc(r)}\right)^2 - \sigma^2_{hc(r)}\right]}{W_h + f(x_h)\sigma^2_{hc(r)}} = \frac{W_i f(x_h)\left[\left(c(x_h) - \mu_{ic(r)}\right)^2 - \sigma^2_{ic(r)}\right]}{W_i + f(x_h)\sigma^2_{ic(r)}}$$

Therefore, using these results we get the minimal equations on simplification as

$$c(x_h) = \left(\frac{\mu_{hc(r)} + \mu_{ic(r)}}{2}\right), i = h+1, h = 1,2,\ldots,L-1 \tag{2.4}$$

These equations are implicit functions of the strata boundaries $[x_h]$ and their exact solutions are somewhat difficult to find. Therefore, we proceed to find the method of solving these minimal equations by conducting approximations.

APPROXIMATE EXPRESSIONS FOR CONDITIONAL MEAN AND VARIANCE

Let the functions $f_i(x)$, $c(x)$ and $\eta(x)$ are bounded away from zero and possess first two derivatives continuous $\forall x \in (a,b)$. Then, we have the following identities due to [17].

$$I_i(y,x) = \int_Y^X (t-y)^i f_i(t)dt = \sum_{j=0}^{3} \frac{(k)^{i+j+1}}{j!(i+j+1)} f^{(j)} + O(k^{i+5}) \tag{3.1}$$

where $f^{(j)}$ is the j^{th} derivative of $f_i(t)$ at $t = y$ and $k = x - y$

$$I_i(y,x) = \int_Y^X (t-y)^i f_i(t)dt = \sum_{j=0}^{3} \frac{(-k)^{i+j+1}}{j!(i+j+1)} f^{(j)} + O(k^{i+5}) \tag{3.2}$$

$O(k^i)$ is the higher order terms with power $\geq i$

Let $\mu_\eta(y,x)$ denote the conditional expectation of function $\eta(t)$ in the interval (y,x), so that

$$\mu_\eta(y,x) = \frac{\int_y^x \eta(t) f_i(t)dt}{\int_y^x f_i(t)dt} \tag{3.3}$$

we have from the definition of $\mu_\eta(y,x)$

$$\mu_\eta(y,x)\int_y^x f_i(t)dt = \int_y^x \eta(t) f_i(t)dt)$$

therefore, we have

$$\mu_\eta(y,x).I_0(y,x) = \left[\eta I_0(y,x) + \sum_{i=1}^{3} \eta^{(i)} I_i(y,x)/i!\right] + O(k^5)$$

Using the Taylor series expansions for $I_i(y,x)$ and $I_i(x,y)$ from (3.2) and simplifying the result at point $t = y$, we have

$$\mu_\eta(y,x) = \eta\left[1 + \frac{\eta'}{2\eta}k + \frac{(\eta' f' + 2f\eta'')}{12 f\eta}k^2 + \frac{(ff''\eta' + ff'\eta'' + f^2\eta''' - \eta' f'^2)}{24 f^2 \eta}k^3 + O(k^4)\right] \tag{3.4}$$

Proceeding in the same fashion using Taylor series expansions about the point $'x'$, the expression for $\mu_\eta(y,x)$ is obtained as

$$\mu_\eta(y,x) = \eta\left[1 - \frac{\eta'}{2\eta}k + \frac{(\eta' f' + 2f\eta'')}{12 f\eta}k^2 - \frac{(ff''\eta' + ff'\eta'' + f^2\eta''' - \eta' f'^2)}{24 f^2 \eta}k^3 + O(k^4)\right] \tag{3.5}$$

Let $\sigma^2{}_\eta(y,x)$ denotes the conditional variance of the function $\eta(t)$ in the interval (y,x), we have

$$\sigma^2{}_\eta(y,x) = \mu_{\eta^2}(y,x) - (\mu_\eta(y,x))^2$$

substituting values, we get

$$\sigma^2{}_\eta(y,x) = \frac{k^2(\eta'(y))^2}{12}\left[1 + (\eta''(y)/\eta'(y))k + O(k^2)\right] \tag{3.6}$$

Using the above results, several other approximations can be obtained. Multiplying the series expansions for $\mu_\eta(y,x)$ about the points $t = y$, $t = x$ and taking the square root, we obtain

$$\mu_\eta(y,x) = \sqrt{\eta(y)\eta(x)}\left[1 + O(k^2)\right] \tag{3.7}$$

From (3.3), we have

$$\mu_\eta(y,x).I_0(y,x) = \int_y^x \eta(t) f_i(t)dt$$

Taking $\eta(t) = t^2$ and using (3.7), we get

$$\int_y^x f_i(t)dt = \frac{1}{xy}\int_y^x t^2 f_i(t)dt\left[1 + O(k^2)\right] \tag{3.8}$$

Similarly expanding $\sqrt[\lambda]{f_i(t)}$ about the point t=y, we have

$$\left[\int_Y^X \sqrt[\lambda]{f_i(t)}dt\right]^\lambda = k^{\lambda-1}\int_Y^X f_i(t)dt\left[1 + O(k^2)\right] \tag{3.9}$$

APPROXIMATE SOLUTIONS OF THE MINIMAL EQUATIONS

To find approximate solutions to the minimal equation (2.4), we shall obtain the series expansions of system of equations about the point $[x_h]$, the common boundary of h^{th} and $(h+1)^{th}$ strata. The expansions for the two sides of the equation (2.4) are obtained by using various results proved in the preceding section. For the expansion of the right hand side about the point x_h, (y,x) is replaced by (x_{h-1}, x_h) while for the left hand side we replace (y,x) by (x_{h-1}, x_h).

we have

$$[\mu_{hc(r)} - c(x_h)]^2 - [\mu_{ic(r)} - c(x_h)]^2 = 0$$

We have from (2.13) after replacing $'y'$ and $'x'$ by x_h and x_{h+1} respectively.

$$\mu_{ic(r)} = c\left[1 + \frac{c'}{2c}k_i + \frac{(c'f' + 2fc'')}{12fc}k_i^2 + \frac{(ff''c' + ff'c'' + f^2c''' - c'f'^2)}{24f^2c}k_i^3 + O(k_i^4)\right]$$

The derivatives of c and f are evaluated at x_h. Therefore, we get

$$[\mu_{ic(r)} - c(x_h)] = \frac{k_i}{2}\left[c' + \frac{(c'f' + 2fc'')}{6f}k_i + O(k_i^2)\right] \tag{4.1}$$

Similarly, we get

$$\left[\mu_{hc(r)} - c(x_h)\right] = \frac{k_i}{2}\left[c' + \frac{\left(c'f' + 2fc''\right)}{6f}k_i + O(k_i^{\,2})\right] \tag{4.2}$$

Therefore from (4.1) and (4.2), we have

$$\frac{k_i}{2}\left[c' + \frac{\left(c'f' + 2fc''\right)}{6f}k_i + O(k_i^{\,2})\right] = \frac{k_i}{2}\left[c' + \frac{\left(c'f' + 2fc''\right)}{6f}k_i + O(k_i^{\,2})\right]$$

Now let us consider an expansion of the function

$$B_h = \int_{x_{h-1}}^{x_h} c'^2 f_i(t)dt$$

$$\therefore B_h = c'^2 f k_h\left[1 - \frac{\left(c'f' + 2fc''\right)}{fc'}\frac{k_h}{2} + O(k_h^2)\right]$$ (Using Taylors expansion)

multiplying it by $\left(\frac{k_h^2}{8}\frac{c'}{f}\right)$ and taking cube root both sides, we get

$$\left[\frac{k_h^2}{8}\frac{B_h c'}{f}\right]^{\frac{1}{3}} = \frac{c'k_h}{2}\left[1 - \frac{\left(c'f' + 2fc''\right)}{6fc'}k_h + O(k_h^2)\right] \tag{4.3}$$

similarly, we have

$$\left[\frac{k_i^2}{8}\frac{B_i c'}{f}\right]^{\frac{1}{3}} = \frac{c'k_i}{2}\left[1 - \frac{\left(c'f' + 2fc''\right)}{6fc'}k_i + O(k_i^2)\right] \tag{4.4}$$

From the equations (4.3) and (4.4), we have

$$\left[\frac{k_h^2}{8}\frac{B_h c'}{f}\right]^{\frac{1}{3}} = \left[\frac{k_i^2}{8}\frac{B_i c'}{f}\right]^{\frac{1}{3}} \tag{4.5}$$

Or

$$k_h^2 B_h = \text{constant} \tag{4.6}$$

In case it is possible to find a function $Q_2\left(x_{h-1}, x_h\right)$ such that

$$k_h^2 B_h = k_h^2 \int_{x_{h-1}}^{x_h} c'^2 f_i(t)dt$$

$$= Q_2\left(x_{h-1}, x_h\right)\left[1 + O(k_h)^2\right] \tag{4.7}$$

Thus the system of equations (4.6) to the same degree of accuracy can be put as

$Q_2\left(x_{h-1}, x_h\right) =$ Constant

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

THEOREM :-

If the regression of the estimation variable $'Y'$ on the stratification variable $'X'$, in the infinite super population is given by

$$y = c(x) + e$$

where $'c(x)'$ is a function of auxiliary variable, $'e'$ is the error term such that $E(e\,|\,x)=0$ and $V(e\,|\,x)=\eta(x)>0\ \forall x\in(a,b)$ with $(b-a)<\infty$, and further if the function $g_1(x)\,f_i(x)\ \in\Omega$: then the system of equations (2.4) give strata boundaries (x_h) which correspond to the minimum of $V(\overline{Y}_{stRSS})_{prop}$ can be written as

$$\left[k_h^2\int_{x_{h-1}}^{x_h} c'^2 f_i(t)dt.\left[1+O(k_h^2)\right]\right]^{\frac{1}{3}}=\left[k_i^2\int_{x_h}^{x_{h+1}} c'^2 f_i(t)dt.\left[1+O(k_i^2)\right]\right]^{\frac{1}{3}}$$

Neglecting the terms of order $O\left(Sup_{(a,b)}(k_h)\right)^3$ can be neglected; these equations can be replaced by the approximate system of equations

$$k_h^2\int_{x_{h-1}}^{x_h} c'^2 f_i(t)dt = \text{constant}$$

Or equivalently by

$$Q_2(x_{h-1},x_h) = \text{constant}\quad,\quad k_h=(x_h-x_{h-1})$$

$$Q_2(x_{h-1},x_h)\left[1+O(k_i^2)\right]=k_h^2\int_{x_{h-1}}^{x_h} c'^2 f_i(t)dt\quad,\ i=h+1, h=1,2,\ldots,L$$

The similar results can also be obtained by minimizing the function

$$\sum_{h=1}^{L}\int_{x_{h-1}}^{x_h} c'^2 f_i(t)dt\left[1+O(k_h^2)\right]$$

Thus we find that if the function $c'^2 f_i(x)$ belongs to Ω, the minimum value of $\sum_{h=1}^{L} W_h\sigma_{hc}^2$ and therefore $V(\overline{Y}_{stRSS})_{prop}$, exists and the solutions of the system of equations (2.4) or equivalently of (4.5). These equations as such are very difficult to solve and therefore it is essential to find some way out of this difficulty. It is done by replacing these systems of equations by other systems 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. We have obtained these systems of equations after neglecting the terms of order $O\left(Sup_{(a,b)}(k_h)\right)^4=O(m^4)$ where $m=Sup_{(a,b)}(k_h)$, on both sides of the equation (4.5). If the number of strata is large and therefore terms of order $O(m^4)$ are quite small, the error involved in the approximate systems of equations is expected to be quite small and the set of points (x_h) obtained from them shall be quite near the optimum values.

Now we proceed to develop the approximate systems of equations given in (4.6) and (4.7). Here, in finding various forms of the function $Q_2(x_{h-1}, x_h)$, we shall keep in mind that the function $Q_2(x_{h-1}, x_h)$ is such that

$$k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt = Q_1(x_{h-1}, x_h)\left[1 + O(k_i^2)\right]$$

If in (2.4) we retain only the first term on both sides of the equation and neglect the others, the two sides are equalized if

$$k_h = \text{constant, } \left(\frac{b-a}{L}\right)h \text{ , } \quad \text{for} \quad h = 1,2,\ldots,L \tag{4.8}$$

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

This set of solutions cannot be expected to yield very good results as we have neglected terms of order $O(m^3)$ on both sides of the exact minimal equations. This solution holds for all $c'^2 f_i(x)$ provided they belong to Ω and all density functions with finite range. Due to its universality of application it can be recommended in case of less information about c'^2 and $f_i(x)$. 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' is infinite, but one can resort to truncation of the density function to any suitable probability level before using this approximation.

We obtain next approximate systems of equations, the optimum points of stratification are such that

$$k_h^2 \int_{x_{h-1}}^{x_h} c'^2 f_i(t) dt = c_1 \text{ , } \quad h = 1,2,\ldots,L \tag{4.9}$$

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

From (3.8) and equation (4.9) is obtained the following class of approximate equations. The approximations to optimum (x_h) are obtained from

$$\left[k_h^{(3\lambda-1)} \int_{x_{h-1}}^{x_h} c'^2 f_i(t)^\lambda dt\right]^{\frac{1}{\lambda}} = \text{constant,} \quad h = 1,2,\ldots,L$$

For λ = 1/2 and 1/3 we have

$$k_h\left[\sqrt{\int_{x_{h-1}}^{x_h} c'^2 f_i(t)^\lambda dt}\right] = c_2 = \text{constant,} \quad h = 1,2,\ldots,L \tag{4.10}$$

For λ =1/3, we have the system of equation as

$$\sqrt[3]{\int_{x_{h-1}}^{x_h} \left[c'^2 f_i(t)dt\right]} = c_3\,, \qquad h = 1,2,...,L \tag{4.11}$$

In all these systems of equations, $c's$ are constants to be determined and in some cases, the few equations may be meaningless such as for $h = 1$, i,e $x_{h-1} = 0$.

Cum$\sqrt[3]{K_1(x)}$ Rule

If the $K_1(x) = c'^2 f_i(t)$ is bounded and its first two derivative exists $\forall x \in [a,b]$, then for given value of L taking equal intervals on the cumulative cube root of $K_1(x)$ will give AOSB.

EMPIRICAL STUDY:

We shall consider following distributions of auxiliary variable for evaluating the efficiency of the proposed method for obtaining optimum points of stratification.

Let us assume the auxiliary variable 'x' follows certain distributions as follows:

I. Rectangular $f(x) = \begin{cases} \dfrac{1}{b-a} & , \quad a \le x \le b \\ 0, otherwise \end{cases}$

II. Right-triangular $f(x) = \begin{cases} \dfrac{2(2-x)}{(b-a)^2} & , \quad a \le x \le b \\ 0, otherwise \end{cases}$

III. Exponential $f(x) = e^{-x+1}$, $\quad 1 \le x \le \infty$

IV. Standard normal $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $\quad -\infty \le x \le \infty$

If the stratification variable follows the uniform distribution with pdf $f(x) = \dfrac{1}{b-a}$ $x \in [1,2]$, utilizing the cum $\sqrt[3]{K_1(x)}$ rule, we get the stratification points as given in table I.

Table I: *AOSB and Variance for uniformly distributed auxiliary variable*

L	AOSB	Total variance $n\left\{V\left(\bar{y}_{stRSS}\right)_{\mathrm{Prop}}\right\}$	R.E.%
2	1.4967	0.75521	102.069
3	1.3257, 1.6624	0.75232	102.461
4	1.2483, 1.4959, 1.7466	0.75151	102.572
5	1.1983, 1.3965, 1.5898, 1.7967	0.75084	102.663
6	1.1631, 1.3289, 1.4922, 1.6543, 1.8236	0.75058	102.698

If the stratification variable follows the right triangular distribution with pdf $f(x) = 2(2-x)$ $x \in [1,2]$, utilizing the cum $\sqrt[3]{K_1(x)}$ rule, we get the stratification points as given in table II.

Table II: *AOSB and Variance for Right-triangular distributed auxiliary variable*

L	AOSB	Total Variance $n\left\{V\left(\bar{y}_{stRSS}\right)_{\mathrm{Prop}}\right\}$	R.E.%
2	2.9662	1.01136	111.979
3	2.2935, 3.6419	0.97727	115.886
4	2.0325, 3.0001, 3.9986	0.96505	117.353
5	1.7954, 2.5879,3.4289, 4.2247	0.95773	118.250
6	1.6565, 2.3143, 2.4771, 3.6057, 4.2593	0.95589	118.477

If the stratification variable follows the Exponential distribution with pdf $f(x)=e^{-x+1}\ x\in[1,5]$, utilizing the cum $\sqrt[3]{K_1(x)}$ rule, we get the stratification points as given in table III.

Table III: *AOSB and Variance when the auxiliary variable is exponentially distributed:*

L	AOSB	Total Variance $n\left\{V\left(\bar{y}_{stRSS}\right)_{\mathrm{Prop}}\right\}$	R.E.%
2	1.4949	0.67123	101.389
3	1.3257, 1.6558	0.66879	101.760
4	1.2589, 1.5173, 1.7814	0.66801	101.878
5	1.1981, 1.3971, 1.6452, 1.8494	0.66758	101.944
6	1.1638, 1.3277, 1.4911, 1.6562, 1.8217	0.66722	101.999

If the stratification variable follows the standard Normal distribution with pdf $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ $x\in[0,1]$, utilizing the $cum\sqrt[3]{K_1(x)}$ rule, we get the stratification points as given in table IV.

Table IV: *AOSB and Variance for Standard normally distributed*

L	AOSB	Total Variance $n\left\{V\left(\bar{y}_{stRSS}\right)_{\mathrm{Prop}}\right\}$	R.E.%
2	0.4939	0.08024	121.102
3	0.3282, 0.6531	0.07927	122.584
4	0.2448, 0.4947, 0.7509	0.07893	123.111
5	0.1973, 0.3943, 0.5993, 0.7986	0.07877	123.360
6	0.1687, 0.3286, 0.4944, 0.6596, 0.8269	0.07868	123.496

CONCLUSION

The AOSB are determined for this distribution by using cum $\sqrt[3]{K_1(x)}$ method. For each L = 2, 3, 4, 5 and 6 the variance $n\left\{V\left(\bar{y}_{stRSS}\right)_{\mathrm{Prop}}\right\}$ is calculated, which is used for the efficiency of the stratification. The results of this investigation are given in Table I-IV. When the auxiliary variable follows uniform, right triangular, exponential and standard normal distributions the stratification points obtained has been presented in Table I-IV and percent RE as compared unstratified RSS have been worked out. The standard normal distribution shows highest % R.E. Overall results show that the increase in the number of strata is directly proportional to the decrease in total variance. From the last column of tables it can be seen that the AOSB obtained by the proposed method are more efficient for all L = 2,3,...,6. Thus, the proposed method of cum $\sqrt[3]{K_1(x)}$ shows increases gain in precision in obtaining AOSB while selecting samples using RSS.

References

[1] Dalenius, T. (1950). The problem of optimum stratification. *Scandinavian Actuarial Journal, 33*(3-4): 203-213.

[2] Cochran, W. G. (1961). "Comparison of methods of determining strata boundaries," Bulletin of the International Statistical Institute, 38: 345-358.

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

[4] 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.

[5] 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.

[6] Danish, F., Rizvi, S. E. H., Jeelani, M. I. and Reashi, J. A. (2017). Obtaining strata boundaries under proportional allocation with varying cost of every unit. Pakistan Journal of Statistics and Operation Research 13 (3):567–74.

[7] Danish, F. and Rizvi, S.E.H. (2018) Optimum stratification in Bivariate auxiliary variables under Neyman allocation, *Journal of Modern Applied Statistical Methods*, 17(1).

[8] Gupt. K, B, and Ahamed. I, M, (2021). Construction of Strata for a Model-Based Allocation Under a Super population Model. *Journal of Statistical Theory and Applications.* 20(1): 46-60.

[9] Gupt. K. B., Swer, M, Ahamed, I. M., Singh, B. K. and Singh, K. H., (2022). Stratification Methods for an Auxiliary Variable Model-Based Allocation under a Super-population Model. *Mathematics and Statistics,* 10(1): 15-24.

[10] Rather, K.U.I. and Kadilar, C. (2021). Exponential Type Estimator for the population Mean under Ranked Set Sampling. *Journal of Statistics: Advances in Theory and Applications,* 25(1), 1–12.

[11] Rather, K.U.I., Eda, K. G. Unal, C. & Jeelani, M. I. (2022). New exponential ratio estimator in Ranked set sampling. *Pakistan Journal of Statistics and operation research,* 18(2), 403–409.

[12] Rather, K.U.I., Rizvi, S.E.H., Sharma, M. and Jeelani, M. I. (2022): Optimum Stratification for Equal Allocation using Ranked Set Sampling as Method of Selection. International Journal of Statistics and applied mathematics. 7(2): 63-67.

[13] Rather, K.U.I. and Rizvi, S.E.H. (2022): Approximate Optimum Strata Boundaries for Equal Allocation under Ranked Set Sampling. *Journal of Scientific Research,* 66(3): 302-307.

[14] Mcintyre, G. A. (1952). A method for unbiased selective sampling, using ranked set sampling and stratified simple random sampling. *Journal of Applied Statistics* 23, 231-255.

[15] Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the stratified sampling by means of ordering. Annals of the Institute of Statistical Mathematics, 20, 1-31.

[16] Hayashi, C., Maruyama F. and Isida, M. D. (1951). "On some criteria for stratification," *Annals of the Institute of Statistical Mathematics,* 2: 77-86.

[17] Ekman, G. (1959). "Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution," Annals of the Institute of Statistical Mathematics, 30: 1131-1134.