

---

## Estimating Bark Eating Caterpillars *Indarbela quadrinotata* (walker) in *Populus deltoides* Using Ranked Set Sampling

Arvind Kumar<sup>1</sup>, Girish Chandra<sup>2</sup>, and Sanjay Kumar<sup>3</sup>

<sup>1</sup>Forest Protection Division, Forest Research Institute, Dehradun, India

<sup>2</sup>Forest Statistics Division, Indian Council of Forestry Research and Education, Dehradun, India

<sup>3</sup>Ministry of Statistics and Program Implementation, Government of India, New Delhi, India

<sup>2</sup>gchandra23@yahoo.com

**Abstract.** The problem of bark eating caterpillar, *Indarbela quadrinotata* infestation has been observed from variety of horticulture and forest tree species in India. The estimation of infestation of this caterpillar using conventional sampling methods was found difficult because counting the number of caterpillar in each tree is practically not feasible. Ranked set sampling (RSS) is a cost efficient method which provides improved estimators of mean and variance when actual measurement of the observations is difficult to obtain but a reasonable ranking of the units in the sample is relatively easy. In the present study, poplar, *Populus deltoides* plantation of Western Uttar Pradesh and Uttarakhand was taken for the assessment of *Indarbela quadrinotata* infestation. The RSS estimator of population mean and variance have been discussed and compared with the corresponding estimators from simple random sampling (SRS). The relative precision (RP) of RSS procedure with respect to the SRS for four different set sizes of  $k = 3, 5, 7,$  and  $10$  has been deliberated. It was seen that RP increase with the increment in  $k$ . The method of RSS was found suitable for the assessment of insect pest infestation.

**Keywords:** *Indarbela quadrinotata*, *Populus deltoides*, simple random sampling, ranked set sample, order statistics.

### 1. Introduction

The bark eating caterpillar, *Indarbela quadrinotata*, is economically injurious insect pest. This insect pest infests to many plant species and damages varieties of fruit trees in central India [1] and thus becomes a pest of national significance [2]. Bark eating caterpillar is a polyphagous and comparatively considered of minor importance in forestry; however it is a major pest of Indian gooseberry *Emblica officinalis* [3, 4]. This caterpillar belongs to order Lepidoptera, family Cossidae with bigger larval size, about 35-45 mm, dirty brown in colour and smooth body. Presence of long winding, thick and brownish ribbon-like masses composed of small chips of wood and excreta on the tree is evidence of infestation of this pest [1].

The larva of *Indarbela quadrinotata* make shelter tunnels inside the trunk, feeding on plant tissues and make plant and stem weaker, resulting in the drying of the branches and finally the mortality of the tree itself [5]. The infestation of pest generally starts in the month of April with the emergence of moths [6]. Female lays egg below the bark or in between cracks and there after larvae bore into the tree continue feeding on the

---

bark for 9 to 11 months, however, the damage caused by this pest reached maximum during rainy season [1, 6, 7]. The larva of *I. quadrinotata* made a tunnel into the trunk of the tree for about 20 cm depth and take refuge during the day and eats the bark of the tree during night, making excretal ribbons. Feeding of bark and internal tissues by the larva resulted in inhibiting translocation of cell sap, reduces growth and top dying of the plant [8] and subsequently, death of the heavily infested plants. The pest has been reported to infest 70 plant species across fruits, forests and avenue plantations [9].

The enumeration of the population of active caterpillar *I. quadrinotata* per tree or estimation of proportion of trees infected by the caterpillar is one of the important challenges for entomologist to know density and distribution of the caterpillar over the tree or tree stands. The caterpillar makes tunnel into the tree and feed bark tissues covering with fecal ribbons. It is compulsory to remove the ribbon made by *I. quadrinotata* confirm the active insects inside which requires more resources in terms of effort, cost and time. Therefore, the conventional sampling technique like simple random sampling (SRS) failed here because the actual measurement of active insects present in each tree is not feasible.

An effort for the assessment of *I. quadrinotata* in poplar trees has been made in the present study with the help of ranked set sampling (RSS) which primarily required ranking of trees based on the number of ribbons and their sizes rather than the actual measurement of the number of active insects in each tree. In this method, the first step is to rank the small set of samples of trees by some rough gauging method (say visual inspection) without the actual measurement. After repeating the process in certain cycles, the unit (tree) associated with each rank order statistic is selected for measurement (see Section 2.2.1). Thus, the RSS selected units which are more likely to span over the full range of the population. The suitability of RSS lies in the fact that it yields more precise estimator of population mean (average number of insects per tree) as compared to SRS with lesser amount of measurement of units irrespective of possible raking errors.

The technique of RSS has been applied satisfactorily in various areas of agriculture and forestry. For example, in estimating yield of dry bark and quinine contents from *Cinchona* plants [10]; estimation of forage yields in a pine hardwood forest [11]; estimation of shrub phytomass in Appalachian Oak forest [12]; survey of conifer (*Pinus palustris*) trees [13] etc.

In the present study, the average number of caterpillars per poplar tree had been estimated by RSS and the same has been compared with the corresponding estimates of SRS. Section 2 deals with material and methods used in the present case study. The framework of RSS has been briefly described and the technique of RSS has been applied in estimating bark eating caterpillars in poplar tree. Section 3 discusses the results and finally the study has been concluded in Section 4.

## 2. Material and Methods

**2.1. Study area.** A total of 12 different sites which includes farmer's plantation areas of poplar, *Populus deltoides* tree in Haridwar, Saharanpur and Muzaffarnagar districts of the states of Uttarakhand and Uttar Pradesh were selected. The description of these sites along with the area under plantation is given in Table 1.

Table 1. Description of study sites of poplar plantation

Study Site	Site name	District (State)	Area of plantation (in ha.)
Site-1	Bijipur	Haridwar (Uttarakhand)	0.80
Site-2	Maqulpur	Haridwar (Uttarakhand)	0.60
Site-3	Narsalkala	Haridwar (Uttarakhand)	1.10
Site-4	Iqbalpur	Haridwar (Uttarakhand)	0.90
Site-5	Hosangpur	Haridwar (Uttarakhand)	1.20
Site-6	Salauli	Saharanpur (Uttar Pradesh)	0.50
Site-7	Meghrajpur	Saharanpur (Uttar Pradesh)	0.40
Site-8	Deoband	Saharanpur (Uttar Pradesh)	0.60
Site-9	Sisauli	Muzaffarnagar (Uttar Pradesh)	0.80
Site-10	Lalukhera	Muzaffarnagar (Uttar Pradesh)	0.40
Site-11	Baghra	Muzaffarnagar (Uttar Pradesh)	0.50
Site-12	Kajikhera	Muzaffarnagar (Uttar Pradesh)	1.30

**2.2. Ranked set sampling.** McIntyre [14] while estimating mean pasture and forage yields in Australia, introduced the concept of RSS. This method provided more precise estimator for the population mean and reduced the sample size for actual measurement. Takahasi and Wakimoto [15] proved that when ranking is perfect, the RSS mean is an unbiased estimator of the population mean and the variance of the RSS mean is always smaller than the variance of the mean of SRS of the same sample size. Dell and Clutter

[16] showed that RSS performs better than SRS even in presence of ranking error in the experiment. For more detail, one may refer to [17, 18, 19, 20].

**2.2.1 Obtaining a ranked set sample.** The RSS starts with drawing a simple random sample of size  $k$  from the population and these  $k$  units are ranked on the attribute of interest through some personal judgment or an expert advice or use of a concomitant variable, without actual measurement of the units. After ranking of these  $k$  selected units, the smallest unit i.e. unit with rank 1 is identified and included in ranked set sample for measurement; the remaining  $k-1$  units of the sample are discarded. Thereafter, another independent simple random sample of size  $k$  is drawn and the units are ranked, the unit with rank 2 is included in the ranked set sample; the remaining units are discarded. This process is repeated until the largest of  $k$  units in  $k^{th}$  random sample is included in ranked set sample. This whole process is referred to as a cycle and the number of units in each random sample ( $k$  here) is known as set size. The cycle is repeated  $m$  times to get a ranked set sample of size  $n = km$  from the population of size  $N = k^2m$ . This procedure in which each rank order statistics is measured equal number of times is termed as balanced RSS.

Let  $Y_{(i:k)_j}$ ,  $i = 1, 2, \dots, k$ ;  $j = 1, 2, \dots, m$ , denotes the measured unit for the  $i^{th}$  rank order in the  $j^{th}$  cycle. The  $k^2$  ordered observations under the  $j^{th}$  cycle can be displayed as

$$\begin{aligned} & Y_{(1:1)_j}, Y_{(1:2)_j}, \dots, Y_{(1:k)_j} \\ & Y_{(2:1)_j}, Y_{(2:2)_j}, \dots, Y_{(2:k)_j} \\ & \cdot \\ & \cdot \\ & \cdot \\ & Y_{(k:1)_j}, Y_{(k:2)_j}, \dots, Y_{(k:k)_j} \end{aligned}$$

Suppose, for fixed  $i$ ,  $Y_{(i:k)_j}$ ,  $j = 1, 2, \dots, m$ , are independent and identically distributed with mean  $\mu_{(i:k)}$  and variance  $\sigma_{(i:k)}^2$ . An unbiased estimator of  $\mu$  is given as [14].

$$\bar{Y}_{(k)RSS} = \frac{1}{km} \sum_{i=1}^k \sum_{j=1}^m Y_{(i:k)_j} \tag{1}$$

with its variance using equation (2)

$$Var(\bar{Y}_{(k)RSS}) = \frac{1}{k^2 m} \sum_{i=1}^k \sigma_{(i:k)}^2 \quad (2)$$

We know, in SRS a simple random sample  $(X_1, X_2, \dots, X_n)$  from the population with cumulative distribution function  $F(x)$  having mean  $\mu$  and finite variance  $\sigma^2$  has the standard unbiased estimator of  $\mu$  as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (3)$$

with variance

$$Var(\bar{X}) = \sigma^2 / n \quad (4)$$

**2.2.2 Preference of RSS over SRS.** To compare RSS and SRS, one can find the relative precision (RP) of RSS procedure with respect to the SRS as

$$RP = \frac{Var \bar{Y}_{(k)SRS}}{Var \bar{Y}_{(k)RSS}} = \frac{\frac{\sigma^2}{km}}{\frac{1}{k^2 m} \sum_{i=1}^k \sigma_{(i:k)}^2} = \frac{\sigma^2}{\bar{\sigma}^2} \quad (5)$$

Where  $\bar{\sigma}^2 = \frac{1}{k} \sum_{i=1}^k \sigma_{(i:k)}^2$  represents the average variances within the rank order statistics.

Takahasi and Wakimoto [15] proved that under the perfect ranking, the RP lies between 1 and  $(k+1)/2$ . The RSS performs better than SRS even in presence of the ranking error [16]. When the actual measurement of the units requires more resources and the reasonable ranking is cheap, the RSS is most effective.

The RP can be easily calculated, in case of completely known nature of the population (type of distribution, population mean, population variance and variances of order statistics, etc.). However, on several occasions, the nature of the population is not known in advance and therefore, it is to be estimated from the sampled values. The present study is based upon the latter case. Three types of distributions *viz.* skewed, symmetric and moderately skewed are possible. For skewed distributions, the inequality  $\sigma_{(1:k)}^2 \geq \sigma_{(2:k)}^2 \geq \dots \geq \sigma_{(k:k)}^2$  holds; however, for the symmetric distributions there is

---

$$\sigma_{(1:k)}^2 = \sigma_{(k:k)}^2, \sigma_{(2:k)}^2 = \sigma_{(k-1:k)}^2, \dots \text{ with either } \sigma_{(1:k)}^2 \geq \sigma_{(2:k)}^2 \geq \dots \text{ or } \sigma_{(1:k)}^2 \leq \sigma_{(2:k)}^2 \leq \dots$$

The moderately skewed distribution does not follow any of the above systematic patterns of variances.

Various authors also improved balanced RSS procedure by using the appropriate allocation for each rank order statistics, termed as unbalanced RSS. Kaur et al. [21] and Chandra et al. [22] provided allocation procedures for skewed populations by making allocations in directly proportional to the rank order statistics. For symmetric distributions, Kaur et al. [18] and Chandra et al. [23] provided allocations inversely proportional to the rank order statistics for improving the estimates.

Some recent developments (for mean estimation using RSS [24, 25]; for variance estimation [26]; for unequal allocation model in RSS for response estimation of developmental programs [27]) can also be seen in the literature.

**2.2.3. Application of RSS in estimating *Indarbela quadrinotata* in *Populus deltoides* tree.** In the present study, the concomitant variable, number of ribbons made by *Indarbela quadrinotata* was chosen for the ranking purpose.

Four small set sizes  $k = 3, 5, 7$  and  $10$  were used to reduce the possible ranking errors while selecting samples. For illustration of selection procedure, let us consider the set size of  $k = 5$ . Five sets of poplar trees each having size of five was selected from the randomly selected sites by the use of simple random sampling without replacement. Such sets of five trees were ranked based upon the number of ribbons made by the insect by visual inspection. Once ranking is done, from the first set, the smallest ranked tree was taken for the actual measurement and other trees were discarded. Similarly, the second smallest ranked tree from the second set was selected for the actual measurement. This process was continued until the largest ranked tree from the last set of five trees was selected for the actual measurement. In this process, 25 trees were randomly selected from five different sets of trees and five were taken for the actual measurement. Now, from these selected five trees, the ribbons from the trees were removed and the numbers of active caterpillars were counted. This process is termed as one cycle. In the present case study, five cycles were taken so as to get 25 trees for actual measurement out of 125 trees. The actual numbers of active caterpillar found for different set sizes are shown in Table 2 to Table 5.

### 3. Results and Discussion

The population values were calculated based upon the 125 measured ordered observations (Tables 2 to Table 5). The population mean and population variance (using equation (3) and (4)) was found to be 6.5760 and 7.7462, respectively. The variances ( $\sigma_{(i;k)}^2$ ) corresponding to the 10 rank order statistics were computed (Table 6) to enable to calculate the RP of RSS over SRS. The values of  $\sigma_{(i;k)}^2$  were not found in a systematic way, i.e., in increasing or decreasing order with the set size  $k$ . Therefore, based on the values of  $\sigma_{(i;k)}^2$ , the nature of the distribution of insect infestation on poplar tree was found to follow the moderately positively skewed distribution (For detail about the distribution, one may refer to [18, 21, 22, 23]).

Table 2. Ranked set sample with  $m = 5$  and  $k = 3$

Cycle	Set		
	I	II	III
I	6	7	10
II	6	5	9
III	7	7	11
IV	5	4	6
V	5	4	7

Table 3. Ranked set sample with  $m = 5$  and  $k = 5$

Cycle	Set				
	I	II	III	IV	V
I	4	5	9	8	9
II	5	4	4	8	7
III	6	9	6	10	10
IV	4	6	5	9	11
V	3	5	5	6	10

Table 4. Ranked set sample with  $m = 5$  and  $k = 7$

Cycle	Set						
	I	II	III	IV	V	VI	VII
I	3	7	8	9	8	7	10
II	3	3	9	6	7	8	9
III	4	2	6	7	6	6	5
IV	3	4	6	5	7	7	8
V	5	5	4	6	7	10	9

Table 5. Ranked set sample with  $m = 5$  and  $k = 10$

Cycle	Set									
	I	II	III	IV	V	VI	VII	VIII	IX	X
I	6	4	3	7	9	11	11	9	9	13
II	3	7	4	5	8	12	10	8	13	13
III	2	5	2	9	3	8	7	12	12	13
IV	1	2	4	4	4	6	6	6	7	10
V	1	2	3	4	5	5	5	6	8	9

Table 6. Estimates of  $\sigma_{(i,k)}^2$  for  $k = 3, 5, 7$  and  $10$

$k$	$i$									
	1	2	3	4	5	6	7	8	9	10
3	2.9368	3.6079	6.5763							
5	2.9368	3.6079	6.5763	3.6952	5.1143					
7	2.9368	3.6079	6.5763	3.6952	5.1143	5.3333	4.6667			
10	2.9368	3.6079	6.5763	3.6952	5.1143	5.3333	4.6667	6.2000	6.7000	3.8000

The RSS estimator for the population mean using equation (1) and values of RP using equation (5) for different values of  $k$  were calculated (Table 7). The calculation revealed that the RSS estimator is very close to the population mean 6.5760 and the



variance is substantially reduced when the set size  $k$  increases. The gain in RP of RSS over SRS was found significantly higher and RP increases as the set size  $k$  increases.

Table 7. RSS estimator and RP over SRS for  $k = 3, 5, 7$  and  $10$

$k$	$\bar{Y}_{(k)RSS}$	$Var(\bar{Y}_{(k)RSS})$	$RP$
3	6.6000	0.2916	26.5664
5	6.7200	0.1754	44.1518
7	6.2571	0.1303	59.4358
10	6.7200	0.0973	79.6433

Fortunately, for some skewed and symmetric distributions, the values of  $\sigma_{(i,k)}^2$  are available [28]. The calculated values of RP for the five distributions viz. Uniform (0, 1), Normal (0, 1), Lognormal (0, 1), Weibull (0, 1) and Gamma (1) shown in the Figure 1 for set sizes  $k = 2$  to  $10$ .

Empirically, for all the distributions, the RP increases as the set size  $k$  increases. The present study also suggested the same (Table 7). The rate of increase in RP becomes higher as skewness of the distribution increases.

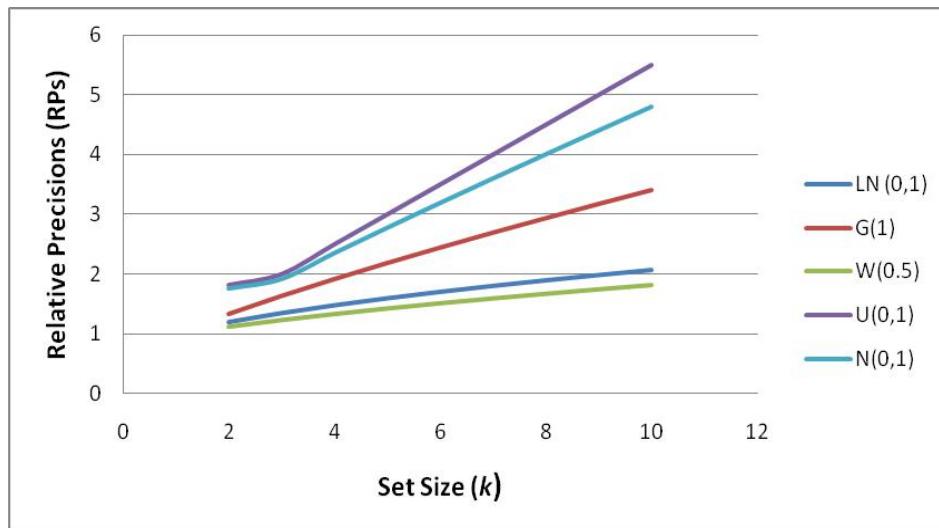


Figure 1. RP of RSS over SRS for some skewed and symmetric distributions

#### 4. Conclusions

The bark eating caterpillar, *Indarbela quadrinotata*, is a serious pest that damages variety of horticulture and forestry trees in India and is a pest of national significance. It eats the bark of the plant and the heavily infested plants ultimately lead to mortality. To

confirm the active insects in the tree, it was compulsory to remove the ribbon made by *Indarbela quadrinotata* which requires more resources in terms of effort, cost and time. The assessment of this caterpillar was quite important for taking up remedial measures like control of pest at the egg or larval stages. The conventional sampling technique was not suitable here as the measurement on each and every unit in neither feasible nor desirable. An easier sampling procedure, RSS, which require only ranking of tree after visual inspection, estimated the number of caterpillar with greater precision and at a substantially lower cost. The technique of the RSS can be applied to other similar estimation problems of entomology cases.

### References

- [1] CROPSAP. *Pests of Fruits (Banana, Mango and Pomegranate) : 'E' Pest Surveillance and Pest Management Advisory*. In: Ahuja DB (ed) Published by National Centre for Integrated Pest Management, New Delhi and State Department of Horticulture, Commissionerate of Agriculture, New Delhi. pp. 67. 2013.
- [2] AESA. *AESA based IPM Package: AESA based IPM - Amla*. In. technical report of department of agriculture and cooperation, ministry of agriculture, govt. of India. 2014. <http://niphm.gov.in/IPMPackages/Amla.pdf>, assessed on March 20, 2018.
- [3] Mathew, G. *Management of the bark caterpillar Indarbela quadrinotata in forest plantations of Paraserianthesfalcataria*. KFRI Research Report No. 122. Kerala Forest Research Institute, Peechi, India. 1997.
- [4] Bharpoda, T. M., Koshiya, D. J., and Korat, D. M. Seasonal Occurrence of insect-pests on Aonla (*Embllica officinalis* Gaertn) and Their Natural Enemies. *Karnataka J. Agric. Sci.* 22(2): 314-318. 2009.
- [5] Mann, G. S. and Bindra, O. S. Field screening of different jujube cultivars against the bark eating caterpillar. *Ind. J. Hortic.* 34: 313-315. 1977.
- [6] Pathak, R. K. *Status Report on Genetic Resources of Indian Gooseberry-Aonla (Embllica officinalis Gaertn.) in South and Southeast Asia*. IPGRI, NASC, New Delhi. 2003.
- [7] Sasidharan, K. R. and Varma, R. V. Seasonal population variations of the bark eating caterpillar (*Indarbela quadrinotata*) in Casuarina plantations of Tamil Nadu. *Tropical Ecology.* 49(1): 79-83. 2008.
- [8] Sasidharan, K. R., Varma, R. V., and Sivaram, M. Impact of *Indarbela quadrinotata* on the Growth of *Casuarina equisetifolia*. *The Indian Forester.* 136(2): 182-186. 2010.
- [9] Verma, A. N. and Khurana, A. D. Bark eating borers (*Indarbelaspp.*). The pest of fruit, avenue and forest trees in India and South East Asia. *Haryana Journal of Horticultural Science.* 7(1-2): 40-46. 1978.
- [10] Sengupta, J. M., Chakarvarti, I. M. and Sarkar, D. Experimental surveys for the estimation of cinchona yield. *Bulletin of the International Statistical Institute.* 33(2): 313-331. 1951.
- [11] Halls, L. K. and Dell, T. R. Trials of Ranked Set Sampling for Forage Yields. *Forest Science.* 12: 22-26. 1966.
- [12] Martin, W. L., Sharik, T. L., Oderwald, R. G. and Smith, D. W. *Evaluation of Ranked Set Sampling for Estimating Shrub Phytomass in Appalachian Oak Forests.*

- Blacksburg, Virginia: School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, FWS, pp. 4-80. 1980.
- [13] Platt, W. J., Evans, G. M., and Rathbun, S. L. The Population Dynamics of a Long-Lived Conifer (*Pinus palustris*). *American Naturalist*. 131: 491-525. 1988.
- [14] McIntyre, G. A. A Method for Unbiased Selective Sampling Using Ranked Sets. *Australian Journal of Agricultural Research*. 3: 385-390. 1952.
- [15] Takahasi, K. and Wakimoto, K. On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering. *Annals of the Institute of Statistical Mathematics*. 20: 1-31. 1968.
- [16] Dell, T. R. and Clutter, J. L. Ranked Set Sampling Theory with Order Statistics Background. *Biometrics*. 28: 545-553. 1972.
- [17] Patil, G. P., Sinha, A. K., and Taillie, C. *Ranked Set Sampling*. In: Patil, G. P., and Rao, C. R., editors, Handbook of Statistics. 12 ed: N. Holland/Elsevier Publishers, New York-Amsterdam, pp. 167-200. 1994.
- [18] Kaur, A., Patil, G. P., and Taillie, C. *Optimal Allocation Models for Symmetric Distributions in Ranked Set Sampling*. Technical Report TR 95-0201, Center for Statistical Ecology and Environmental Statistics. Penn. State University, University Park PA. 1995.
- [19] Chen, Z., Bai, Z. D., and Sinha, B. K. *Ranked Set Sampling: Theory and Applications*. New York: Springer-Verlag. 2004.
- [20] Tiwari, N. and Chandra, G. A systematic procedure for unequal allocation for skewed distributions in ranked set sampling. *Journal of the Indian Society of Agricultural Statistics*. 65(3): 331-338. 2011.
- [21] Kaur, A., Patil, G. P., and Taillie, C. Unequal Allocation Models for Ranked Set Sampling with Skew Distributions. *Biometrics*. 53: 123-130. 1997.
- [22] Chandra, G., Tiwari, N., and Singh, S. K. Near optimal allocation model for skewed distributions in ranked set sampling. *International Journal of Statistics and Systems*. 6(1): 57-65. 2011.
- [23] Chandra, G., Tiwari, N., and Nautiyal, R. Near optimal allocation models for symmetric distributions in ranked set sampling. In: Chandra, G., Nautiyal, R., Chandra, H., Roychoudury, N., Mohammad, N., editors, *Statistics in Forestry: Methods and Applications*: Bonfring Publication, Coimbatore. pp. 85-90. 2015.
- [24] Chandra, G., Pandey, R., Bhoj, D., Nautiyal, R., Ashraf, J., and Verma, M. R. Ranked Set Sampling Approach for Estimating Response of Developmental Programs with Linear Impacts Under Successive Phases. *Pakistan Journal of Statistics*. 34(2): 163-176. 2018.
- [25] Zamanzade, E. and Mohammadi, M. Some Modified Mean Estimator in Ranked Set Sampling Using A Covariate. *Journal of Statistical Theory and Applications*. 15(2): 142-152. 2016.
- [26] Zamanzade, E. and Vock, M. Variance Estimation in Ranked Set Sampling Using A Concomitant Variable. *Statistics and Probability Letters*. 105: 1-5. 2015.
- [27] Chandra, G., Bhoj, D., and Pandey, R. Simple unbalanced ranked set sampling for mean estimation of response variable of developmental programs. *Journal of Modern Applied Statistical Methods*. 17(1). doi: 10.22237/jmasm/1543856083. 2018.
- [28] Harter, H. L. and Balakrishnan, N. *CRC Handbook of Tables for the Use of Order Statistics in Estimation*: CRC Press, Boca Raton. 1996.