

# Quantile estimation using near optimal unbalanced ranked set sampling

Raman Nautiyal<sup>a</sup>, Neeraj Tiwari<sup>a</sup>, Girish Chandra<sup>1,b</sup>

<sup>a</sup>Department of Statistics, Kumaun University, India;

<sup>b</sup>Division of Forestry Statistics, Indian Council of Forestry Research and Education, India

---

## Abstract

Few studies are found in literature on estimation of population quantiles using the method of ranked set sampling (RSS). The optimal RSS strategy is to select observations with at most two fixed rank order statistics from different ranked sets. In this paper, a near optimal unbalanced RSS model for estimating  $p^{\text{th}}$  ( $0 < p < 1$ ) population quantile is proposed. Main advantage of this model is to use each rank order statistics and is distribution-free. The asymptotic relative efficiency (ARE) for balanced RSS, unbalanced optimal and proposed near-optimal methods are computed for different values of  $p$ . We also compared these AREs with respect to simple random sampling. The results show that proposed unbalanced RSS performs uniformly better than balanced RSS for all set sizes and is very close to the optimal RSS for large set sizes. For the practical utility, the near optimal unbalanced RSS is recommended for estimating the quantiles.

**Keywords:** asymptotic relative efficiency, Neyman's allocation, order statistics, quantiles, ranked set sampling

---

## 1. Introduction

Ranked set sampling (RSS), introduced by McIntyre (1952), is a sampling scheme that can be utilized to potentially increase precision and reduce costs when actual measurements of the observations are costly or/and time-consuming but the ranking of the set of items can be easily done without actual measurements. Such situations normally arise in environmental monitoring and assessment that require observational data. Since the inception of the concept of RSS by McIntyre (1952) and development of its mathematical foundation by Takahasi and Wakimoto (1968), various researchers have investigated the utility of RSS and the conditions under which it may be useful and cost-effective. RSS has been satisfactorily used to estimate pasture yield by McIntyre (1952), forage yields by Halls and Dell (1966), mass herbage in a paddock by Cobby *et al.* (1985), shrub phytomass by Martin *et al.* (1980), tree volume in forest by Stokes and Sager (1988), root weight of *Arabidopsis thaliana* by Barnett and Moore (1997) and bone mineral density in a human population by Nahhas *et al.* (2002). RSS is a nonparametric procedure (Stokes and Sagar, 1988; Bohn and Wolfe, 1992; Hettmansperger, 1995) but also has been used in the parametric setting (Bhoj (1997); Lam *et al.* (1996); Stokes (1995)). Most of the distributions considered by these investigators belong to the family of random variables with cumulative distribution function (CDF) of the form  $F((x - \mu)/\sigma)$ , where  $\mu$  and  $\sigma$  are the location and scale parameters, respectively.

---

<sup>1</sup> Corresponding author: Division of Forestry Statistics, Indian Council of Forestry Research and Education, PO New Forest, Dehradun 248006, India. E-mail: [gchandra23@yahoo.com](mailto:gchandra23@yahoo.com)

The selection of a ranked set sample of size  $k$  involves drawing  $k$  random samples with  $k$  units in each sample from a population for which an estimate of the mean is required. The units in each sample are ranked by using judgment or other inexpensive methods. The unit with lowest rank is measured from the first sample, the unit with second lowest rank is measured from the second sample, and this procedure is continued until the unit with the highest rank is measured from the last sample. The  $k^2$  ordered observations in  $k$  samples can be displayed as,

$$\begin{array}{cccc} Y_{(11)} & Y_{(12)} & \cdots & Y_{(1k)} \\ Y_{(21)} & Y_{(22)} & \cdots & Y_{(2k)} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{(k1)} & Y_{(k2)} & \cdots & Y_{(kk)} \end{array}$$

We measure only  $k$  ( $Y_{(ii)}$ ,  $i = 1, 2, \dots, k$ ) diagonal observations, and they constitute the RSS. It is noted that these  $k$  observations are independently distributed. The above procedure is repeated  $m \geq 2$  times to get a sample of size  $n = mk$ . Since each order statistic is measured equal number of times, therefore, this procedure is called the balanced RSS. For convenience, we take  $m = 1$  for the computation of asymptotic relative efficiency (ARE).

For estimating the population mean, it is known that balanced RSS is a more precise method than the simple random sampling (SRS) (Tiwari and Chandra, 2011). For the case of balanced RSS designs, it has been shown by Takahasi and Wakimoto (1968) that the relative precision (RP) of RSS with respect to SRS lies between 1 and  $(k + 1)/2$ , where  $k$  is the set size. However, the gain in the performance of the RSS can be further improved when an appropriate unequal allocation for the order statistics is made. There are some unbalanced RSS procedures for skewed distributions suggested by investigators (see ‘ $t$ ’ and ‘ $s, t$ ’ allocation model (Kaur *et al.*, 1997); systematic allocation model (Tiwari and Chandra, 2011) and simple allocation model (Chandra *et al.*, 2018; Bhoj and Chandra, 2019)). For skewed distributions, Neyman’s allocation provides the optimal allocation model with  $\text{Var}(\bar{y}_{\text{Ney}}) < \text{Var}(\bar{y}_{\text{bal}}) < \text{Var}(\bar{y}_{\text{SRS}})$ . However, for symmetric distributions, Neyman’s optimal allocation model is marginal. The allocation models for symmetric distributions are given by Kaur *et al.* (2000) and Chandra and Tiwari (2015). With the use of unequal allocation, Takahasi and Wakimoto (1968) showed that the RP of RSS relative to SRS lies between 0 and  $k$ . This shows that an appropriate use of unequal allocation increases the performance of RSS beyond that achievable with balanced RSS.

In this paper, a practical unbalanced RSS design for estimating the population quantiles has been developed. In the next Section, the method of quantile estimation using RSS is discussed. Section 3 provides the estimator of quantiles based on the proposed near-optimal unbalanced RSS model and its asymptotic variance. The AREs under balanced RSS, optimal and proposed near optimal models are given in Section 4. Section 5 discussed about the imperfect ranking cases. The conclusion of the study is given in Section 6.

## 2. Quantile estimation using ranked set sampling

Suppose the unknown CDF and corresponding probability density function (pdf) are respectively denoted by  $F(x)$  and  $f(x)$  and a random sample of size  $n$  is drawn from ‘ $F$ ’ using unbalanced RSS such that  $m_i$  observations are taken for measurements for  $i^{\text{th}}$  rank order statistic,  $i = 1, 2, \dots, k$  with  $\sum_{i=1}^k m_i = n$ . Let  $X_{(i:k)j}$ ,  $i = 1, 2, \dots, k$ ,  $j = 1, 2, \dots, m_i$  denotes the  $j^{\text{th}}$  observation of  $i^{\text{th}}$  order statistic. Note that the observations corresponding to the  $i^{\text{th}}$  rank order statistic are i.i.d. with mean  $\mu_{(i:k)}$  and variance  $\sigma_{(i:k)}^2$ . Letting  $F_{(i)}$  and  $f_{(i)}$  be the CDF and pdf of the  $i^{\text{th}}$  rank order statistic. We

consider perfect ranking in this paper, however, some description about the imperfect ranking is also discussed in the later section. Our aim is to estimate the  $p^{th}$  quantile,  $\xi_p$ ,

$$\xi_p = \inf\{x : F(x) \geq p\}.$$

It is known that,

$$f_{(i)}(x) = \frac{k!}{(i-1)!(k-i)!} F^{i-1}(x) (1-F(x))^{k-i} f(x)$$

The simplified form of  $f_{(i)}(x)$  can be rewritten as,

$$f_{(i)}(x) = b(F(x); i, k + 1 - i),$$

where,  $b(y; a, b) = y^{a-1}(1-y)^{b-1}/\beta(a, b)$  denotes the pdf of the random variable which follows beta distribution with parameters  $a$  and  $b$ , where,  $\beta(a, b) = \int_0^1 y^{a-1}(1-y)^{b-1} dy = \Gamma(a)\Gamma(b)/\Gamma(a+b)$  and  $\Gamma(a) = (a-1)!$ . The corresponding CDF is denoted by  $B(y; a, b)$ . That is,

$$B(y; a, b) = \int_0^y b(x; a, b) dx.$$

Using these notations, we can write the CDF of  $i^{th}$  rank order statistic,  $F_{(i)}(x)$ , as

$$F_{(i)}(x) = B(F(x); i, k + 1 - i).$$

Since  $F(\xi_p) = P(X \leq \xi_p) = p$ , therefore,

$$F_{(i)}(\xi_p) = B(p; i, k + 1 - i).$$

The quantile  $\xi_p$  can be estimated using quantile of  $F_{(i)}$  corresponding to the probability of  $B(p; i, k + 1 - i)$ , which is known for any  $p$ . Let  $\hat{F}_{(i)}(x) = 1/(m_i) \sum_{j=1}^{m_i} I(X_{(i)j} \leq x)$ , denotes the empirical distribution of  $F_{(i)}(x)$  based on  $i^{th}$  order statistic.

$$\hat{\xi}_i = \inf\{x : \hat{F}_{(i)}(x) \geq B(p; i, k + 1 - i)\}.$$

The problem of quantile estimation using the method of balanced RSS was first considered by Chen (2000) and found that the RSS substantially improves the efficiency of quantile estimation. This can be regarded as a generalization of sign test given by Hettmansperger (1995). The problem of quantile estimation for any distribution function using unbalanced RSS has already been discussed by Chen (2001) and Zhu and Wang (2005). Chen (2001) proved that the method of unbalanced RSS (optimal choice of unbalanced RSS) outperforms the methods of balanced RSS and SRS in terms of ARE. He used a probability vector  $\mathbf{q} = (q_1, q_2, \dots, q_k)$  with  $\sum_{i=1}^k q_i = 1$  and  $0 \leq q_i \leq 1$ , in constructing a new CDF  $F_{\mathbf{q}}(x) = \sum_{i=1}^k q_i F_{(i)}(x)$  and corresponding  $f_{\mathbf{q}}(x) = \sum_{i=1}^k q_i f_{(i)}(x)$  which is useful to find the asymptotical unbiased estimator of  $\xi_p$ . The quantile estimators in balanced RSS (Chen, 2000) and unbalanced RSS (Chen, 2001) were based on the empirical distribution of the pooled data and considered the balanced cases by Stokes and Sagar (1988).

Chen (2001) used at most two rank orders for the optimal unbalanced RSS design. Since the estimator of  $\xi_p$  under SRS scheme has the asymptotic variance  $p(1-p)/(nf^2(\xi_p))$ , therefore, the derived ARE of unbalanced RSS scheme with respect to SRS scheme using Chen (2001) is given by,

$$\text{ARE}(\hat{\xi}_p)_{\text{URSS:SRS}} = \frac{p(1-p) \left\{ \sum_{i=1}^k q_i b_i \right\}^2}{\sum_{i=1}^k q_i B_i (1 - B_i)} \tag{2.1}$$

Table 1: Proposed near optimal allocation model for the quantile estimation

Class of $p$	Condition for $k$	Allocation model
$0 < p \leq 0.17$	all $k$	$m_i = 1, i = 2, 3, \dots, k$ $m_1 = n - \sum_{i=2}^k m_i$
	$k = 2, 3$	$m_i, i = 2, 3, \dots, k;$ $m_1 = n - \sum_{i=2}^k m_i$
$0.17 < p \leq 0.33$	$k = 3(1)7$	$m_i = 1, \text{ for all } i \neq 2;$ $m_2 = n - m_1 - \sum_{i=3}^k m_i$
	$k > 7$	$m_i = 1, \text{ for all } i \neq 3;$ $m_3 = n - m_1 - m_2 - \sum_{i=4}^k m_i$
$0.33 < p \leq 0.50$	odd $k$	$m_i = 1, \text{ for all } i \neq \frac{k+1}{2};$ $m_{\frac{k+1}{2}} = n - \sum_{i=1}^{(k-1)/2} m_i - \sum_{i=(k+3)/2}^k m_i$
	even $k$ and $k = 2$	$m_1 = 1, m_2 = n - m_1$
	even $k$ and $k > 2$	$m_i = 1, \text{ for all } i \neq \frac{k}{2} \text{ and } \frac{k+2}{2};$ $m_{\frac{k}{2}} = m_{\frac{k+2}{2}} = \frac{n - \sum_{i=1}^{(k-2)/2} m_i - \sum_{i=(k+4)/2}^k m_i}{2}$

where  $q_i = 1/k,$

The equation (2.1) gives the expression for ARE of balanced RSS with respect to SRS and is given by,

$$\text{ARE} \left( \hat{\xi}_p \right)_{\text{BRSS:SRS}} = \frac{p(1-p)}{k \sum_{i=1}^k B_i(1-B_i)}. \tag{2.2}$$

For the optimal RSS design, Chen (2001) given the values of ARE for  $k = 2(1)10$  which has also been computed in Table 4.

Zhu and Wang (2005) suggested a new weighted estimator of  $\xi_p$ . They suggested that as each  $\hat{\xi}_i, i = 1, 2, \dots, k$  is a consistent estimator of  $\xi_p$ , therefore, a weighted estimator can be constructed by combining them and assigning some weights to each  $\hat{\xi}_i$ , where,  $\hat{\xi}_i(x) = \inf\{x : \hat{F}_i(x) \geq p\}$  and  $\hat{F}_{(i)}(x) = 1/m_i \sum_{j=1}^{m_i} I(X_{(ij)} \leq x)$  is the empirical CDF of  $F_{(i)}(x)$ .

The optimal strategy in the estimator of Zhu and Wang (2005) is to select observations with one fixed rank from different ranked sets. Chen (2001) and Zhu and Wang (2005) also proved that the optimal rank and the gain in relative efficiency are distribution free and depend on the set size and given probability only. This is also seen in the near optimal RSS model discussed in the next Section.

### 3. Near optimum probability vector for quantile estimation

In this Section, we deal with the problem of quantile estimation based on the unbalanced RSS which depends on each rank order. Motivated from the various allocation models of population mean (Tiwari and Chandra, 2011; Kaur *et al.*, 1997; Kaur *et al.* 2000; Bhoj and Chandra, 2019), the following allocation models are proposed for the purpose of estimating  $p^{th}$  quantile  $\xi_p$  in the range  $0 < p \leq 0.50$ . This model can be generalized for the range  $0.50 < p < 1$ . The range  $0 < p \leq 0.50$  is divided into three non-overlapping classes with equal width i.e.  $0 < p \leq 0.17, 0.17 < p \leq 0.33, 0.33 < p \leq 0.50$ . For each class, there is different allocation  $m_i, i = 1, 2, \dots, k$  depending upon the nature of  $k$ , for finding the probability  $q_i = m_i/n$ . The proposed allocation models are presented in Table 1. In this paper we are using the sample size  $n = k(k + 1)/2$  as taken in Tiwari and Chandra (2011) for the purpose of calculations. Using Table 1, the number of allocations corresponding to each rank order for  $p = 0.01, 0.05, 0.1, 0.25$  and  $0.5$  for set size  $k = 2(1)8$  is shown in Table 2. The basic drawback

Table 2: Proposed allocation model for the quantile estimation at  $k = 2(1)8$  and different  $p$

$p$	$k$	Rank order ( $i$ )								$n$
		1	2	3	4	5	6	7	8	
0.01, 0.05, 0.10	2	2	1							3
	3	4	1	1						6
	4	7	1	1	1					10
	5	11	1	1	1	1				15
	6	16	1	1	1	1	1			21
	7	22	1	1	1	1	1	1		28
	8	29	1	1	1	1	1	1	1	36
	0.25	2	2	1						
3		4	1	1						6
4		1	7	1	1					10
5		1	11	1	1	1				15
6		1	16	1	1	1	1			21
7		1	22	1	1	1	1	1		28
8		1	1	29	1	1	1	1	1	36
0.50		2	1	2						
	3	1	4	1						6
	4	1	4	4	1					10
	5	1	1	11	1	1				15
	6	1	1	9	9	1	1			22
	7	1	1	1	21	1	1	1		28
	8	1	1	1	15	15	1	1	1	36

in the optimal unbalanced RSS strategy of Chen (2001) and Zhu and Wang (2005) is that mostly the probability of 1 is given to one rank orders and in few cases probability 0.5 each to two rank orders so that their asymptotic variance is minimum. This results in use of at most two rank orders in estimating the population quantiles and ignoring the others. This results in the respective estimators losing the property of sufficiency. In our model, we propose to use each rank order at least once. By doing so, the proposed estimator contains the information about all order statistics. With this in mind, the estimator of  $\xi_p$  is proposed to be a linear combination of  $\hat{\xi}_i$ ,  $i = 1, 2, \dots, k$  as,

$$\hat{\xi}_N = \sum_{i=1}^k q_i \hat{\xi}_i, \tag{3.1}$$

where  $q_i = m_i/n$ ,  $i = 1, 2, \dots, k$ , for each  $k$ . For  $k = 2(1)8$ , the values of  $m_i$  are given in Table 2. Since this model performs close to optimal model of Chen (2001), we named it as near optimal model and used ‘ $N$ ’ in the subscript of the estimator. In RSS, it is to be known that the cost of ranking of the units is negligible. For the fixed sample size, the actual measurement cost under the optimal and near optimal allocation models will be almost same since each order statistic taken for measurement (even the same order statistic) has to be measured.

In our case, the probability vector  $\mathbf{q} = (q_1, q_2, \dots, q_k)$  has non-zero components. The following theorem (Rohatgi and Saleh, 2000) for finding the asymptotic variance of  $\hat{\xi}_N$ ,  $AVAR(\hat{\xi}_N)$ , is useful.

**Theorem 1.** *Let  $X_1, X_2, \dots, X_n$  are  $n$  independent random variables with a common mean  $\mu$  and with variances  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ . The linear combination,  $A_1X_1 + A_2X_2 + \dots + A_nX_n$ , with  $A_1 + A_2 + \dots + A_n = 1$ , that has the smallest variance and is obtained by taking  $A_i$  inversely proportional to  $\sigma_i^2$ . The resulting*

minimum variance is,

$$\frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_n^2}} \quad (3.2)$$

Let us denote  $B_i$  and  $b_i$  as  $B_i = B(p; i, k + 1 - i)$  and  $b_i = b(p; i, k + 1 - i)$ . Since, the asymptotic variance of  $\hat{\xi}_i$  (Chen, 2001) is given by,

$$AVAR(\hat{\xi}_i) = \frac{B_i(1 - B_i)}{q_i b_i^2},$$

with  $\sum_{i=1}^k q_i = 1$ . Therefore, using (3.2),  $AVAR(\hat{\xi}_N)$  is given by,

$$AVAR(\hat{\xi}_N) = \left\{ \sum_{i=1}^k \frac{q_i [b_i]^2}{B_i(1 - B_i)} \right\}^{-1},$$

where  $n = \sum_{i=1}^k m_i$ , the sample size. We now find the  $AVAR(\hat{\xi}_N)$  as,

$$AVAR(\hat{\xi}_N) = \frac{1}{AVAR(\hat{\xi}_N)} = \sum_{i=1}^k \frac{q_i [b_i]^2}{B_i(1 - B_i)} \quad (3.3)$$

$AVAR(\hat{\xi}_N)$  is different for different allocation procedures. In the next Section, we compute AREs for three RSS procedures.

#### 4. Computations of AREs

In this Section, the performance of the estimators of population quantiles based on three procedure of balanced, optimum unbalanced RSS (estimator  $\hat{\xi}_c$ ) by Chen (2001) and the near optimal unbalanced RSS (estimator  $\hat{\xi}_N$ ) are compared. In the first case we compared these models using equation (3.3) for different probability values  $p = 0.01, 0.05, 0.10, 0.25$  and  $0.50$  and are given in Table 3. For determining the values of  $B_i$  and  $b_i$ , the online software <https://keisan.casio.com/exec/system/1180573225>, after duly verified, is used.

We now calculated the AREs of three different methods with respect to SRS using equation (2.1) for the same values of  $p$  and  $k$  as used in Table 3. These are presented in Table 4. These AREs helps to understand more, since the values of ARE of the estimator  $\hat{\xi}_c$  are same as calculated by Chen (2001). As known, the performance of balanced RSS is better with respect to SRS for each  $p$  and  $k$ , however it is marginal (Table 4).

From Tables 3 and 4, it is seen that ARE of proposed method is substantially more than that of balanced RSS for each  $p$  and  $k$ . All three AREs are same for  $k = 2$  and  $p = 0.05$ . The difference among these increases with the decrease of  $p$ . We note that each of the ARE increases with  $k$ . Interestingly, the rate of increase of ARE (over the set size  $k$ ) of the proposed method is higher than the optimum method for all  $p$  except  $p = 0.5$  (Table 5). For  $p = 0.5$ , this rate is close (for few  $k$  it is higher and for few  $k$  it is lower) to the optimum method. This implies that as the set size increases the ARE of the proposed method tends to closer to the optimum method. This is one of the important advantages of the proposed method over optimal method.

Now we wish to see the performance of the proposed methods with the increasing values of sample sizes. For this purpose, we take  $k = 4$  and three values of  $p$ , one from each category of Table 1. These

Table 3: Comparison of three AREs (using Equation (3.3)) at different probability values for  $k = 2(1)8$

Method		Set size $k$							
		2	3	4	5	6	7	8	
$p = 0.01$	Balanced	102.50	103.95	105.43	106.86	108.27	109.67	111.05	
	Unbalanced	$\hat{\xi}_C$	201.01	299.93	397.97	494.95	590.93	685.93	779.94
		$\hat{\xi}_N$	135.34	201.94	280.95	365.59	453.03	541.87	631.30
$p = 0.05$	Balanced	22.52	23.89	25.20	26.44	27.63	28.77	29.87	
	Unbalanced	$\hat{\xi}_C$	41.03	59.95	77.85	94.75	110.69	125.69	139.78
		$\hat{\xi}_N$	28.69	41.92	56.79	71.98	86.96	101.46	115.35
$p = 0.10$	Balanced	12.55	13.83	15.02	16.12	17.16	18.14	19.08	
	Unbalanced	$\hat{\xi}_C$	21.05	29.89	37.69	44.50	50.41	55.46	59.72
		$\hat{\xi}_N$	15.38	21.86	28.62	35.04	40.91	46.13	50.69
$p = 0.25$	Balanced	6.70	7.85	8.87	9.79	10.63	11.41	12.15	
	Unbalanced	$\hat{\xi}_C$	9.14	11.68	14.74	19.15	22.63	25.14	28.46
		$\hat{\xi}_N$	7.52	9.76	12.39	16.03	19.20	21.71	24.84
$p = 0.50$	Balanced	5.33	6.43	7.37	8.21	8.96	9.66	10.32	
	Unbalanced	$\hat{\xi}_C$	5.33	9.00	10.47	14.06	15.58	19.14	20.69
		$\hat{\xi}_N$	5.33	7.71	9.23	12.11	13.78	16.68	18.38

Table 4: Comparison of three AREs with respect to SRS (using equation 2.1) at different probability values for  $k = 2(1)8$

Method		Set size $k$							
		2	3	4	5	6	7	8	
$p = 0.01$	Balanced	1.0100	1.0200	1.0300	1.0400	1.0500	1.0599	1.0699	
	Unbalanced	$\hat{\xi}_C$	1.9899	2.9699	3.9399	4.9000	5.8503	6.7907	7.7214
		$\hat{\xi}_N$	1.3366	1.9949	2.7759	3.6132	4.4786	5.3578	6.2431
$p = 0.05$	Balanced	1.0499	1.0995	1.1488	1.1976	1.2459	1.2937	1.3408	
	Unbalanced	$\hat{\xi}_C$	1.9487	2.8475	3.6977	4.5006	5.2577	5.9702	6.6394
		$\hat{\xi}_N$	1.3493	1.9727	2.6767	3.3975	4.1087	4.7978	5.4582
$p = 0.10$	Balanced	1.0989	1.1959	1.2904	1.3821	1.4708	1.5565	1.6392	
	Unbalanced	$\hat{\xi}_C$	1.8947	2.6900	3.3917	4.0054	4.5368	4.9915	5.3748
		$\hat{\xi}_N$	1.3636	1.9406	2.5470	3.1254	3.6544	4.1258	4.5376
$p = 0.25$	Balanced	1.2308	1.4382	1.6248	1.7942	1.9500	2.0947	2.2303	
	Unbalanced	$\hat{\xi}_C$	1.7143	2.1892	2.7633	3.5904	4.2431	4.7136	5.3369
		$\hat{\xi}_N$	1.3901	1.8092	2.3079	2.9904	3.5846	4.0545	4.6459
$p = 0.50$	Balanced	1.3333	1.6000	1.8286	2.0317	2.2165	2.3869	2.5461	
	Unbalanced	$\hat{\xi}_C$	1.3333	2.2500	2.6182	3.5156	3.8961	4.7852	5.1718
		$\hat{\xi}_N$	1.3333	1.9231	2.3013	3.0178	3.4360	4.1596	4.5858

three chosen values of  $p$  are  $p = 0.01, 0.25$  and  $0.50$ . The performance of ARE of proposed method with respect to SRS for increasing sample size is shown in Figure 1.

From Figure 1, we see that as sample size increases, the performance of proposed estimator is increases for each  $p$  and after attaining a peak, its increase remains almost the same. We see that at this peak, the value of ARE is very close to the optimum ARE of Chen (2001). This also helps us to choose the suitable sample size for each value of  $p$  so that the performance of the proposed method could be very near to the optimum method and the other discussed advantage of our proposed method remains the same.

Table 5: Increasing rate of ARE over  $k$  for different methods

Method		Set size $k$						
		3	4	5	6	7	8	
$p = 0.01$	Balanced		1.0099	1.0098	1.0097	1.0096	1.0094	1.0094
	Unbalanced	$\hat{\xi}_C$	1.4925	1.3266	1.2437	1.1939	1.1607	1.1371
		$\hat{\xi}_N$	1.4925	1.3915	1.3016	1.2395	1.1963	1.1652
$p = 0.05$	Balanced		1.0472	1.0448	1.0425	1.0403	1.0384	1.0364
	Unbalanced	$\hat{\xi}_C$	1.4612	1.2986	1.2171	1.1682	1.1355	1.1121
		$\hat{\xi}_N$	1.4620	1.3569	1.2693	1.2093	1.1677	1.1376
$p = 0.10$	Balanced		1.0883	1.0790	1.0711	1.0642	1.0583	1.0531
	Unbalanced	$\hat{\xi}_C$	1.4197	1.2609	1.1809	1.1327	1.1002	1.0768
		$\hat{\xi}_N$	1.4231	1.3125	1.2271	1.1693	1.1290	1.0998
$p = 0.25$	Balanced		1.1685	1.1297	1.1043	1.0868	1.0742	1.0647
	Unbalanced	$\hat{\xi}_C$	1.2770	1.2622	1.2913	1.1818	1.1109	1.1322
		$\hat{\xi}_N$	1.3015	1.2756	1.2957	1.1987	1.1311	1.1459
$p = 0.50$	Balanced		1.2000	1.1429	1.1111	1.0910	1.0769	1.0667
	Unbalanced	$\hat{\xi}_C$	1.6875	1.1636	1.3428	1.1082	1.2282	1.0808
		$\hat{\xi}_N$	1.4424	1.1967	1.3113	1.1386	1.2106	1.1025

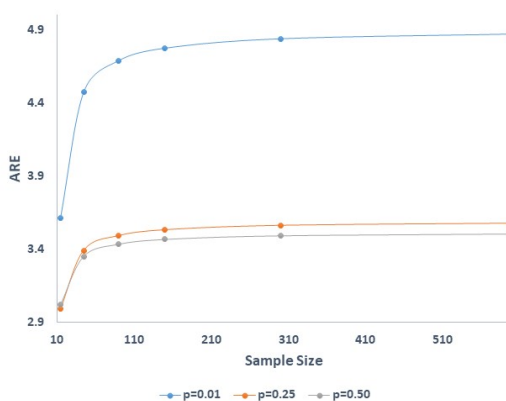


Figure 1: Performance of ARE of proposed unbalanced RSS with SRS over increasing sample size.

### 5. Case of imperfect ranking

Since the RSS procedures perform better in the absence of ranking errors of units, however, this situation is rarely seen in practical situations particularly for the large values of  $k$ . Minimal uncertainty in the rankings may not cause an excessive increase in the variance, but if the ranking process is not very reliable, the precision of RSS estimators (particularly unbalanced ones) may be reduced. The Mann-Whitney-Wilcoxon procedures under imperfect ranking have been discussed by Bohn and Wolf (1992). They discussed proposing a model for the probabilities of imperfect judgment rankings



based on the concept of expected spacing and have used this model to study the properties of tests based on the ranked set analogue of the Mann-Whitney-Wilcoxon statistic. Greater precision in the RSS estimator requires more accurate rankings in each set (David and Levine, 1972; Barnett and Moore, 1997; Stark and Wolfe, 2002). Al-Omari and Bouza (2014) discussed the impact of perfect and imperfect ranking in details. Chen (2001) showed that their model is not optimal when there is a presence of ranking errors. It is difficult to find the exact loss of efficiency due to imperfect ranking, but it is obvious that the probability vector  $\mathbf{q} = (q_1, q_2, \dots, q_k)/$  gets affected by the presence of ranking error. Since the estimators are entirely dependent on this vector, there will be a loss in the AREs. This loss is more for large set sizes. Therefore, it might be expected that the estimator based on optimal RSS is more affected than near optimal and least affected to the balanced RSS methods. The model due to Dell and Clutter (1972) is given by,

$$y_{[ii]} = y_{(ii)} + \varepsilon_{(ii)},$$

where,  $y_{[ii]}$  and  $y_{(ii)}$  denotes the “judgment order statistic” and the “true order statistic” respectively, and  $\varepsilon_{ii} \sim N(0, \sigma_e^2)$ . Dell and Clutter (1972) have shown that in judgment are likely to be influenced by the set size  $k$ .

## 6. Conclusion

In the first study of quantile estimation by unbalanced RSS (Chen, 2001), the optimal  $\mathbf{q}$  for estimating population quantiles have at most two non-zero components. While in the weighted estimator given by Zhu and Wang (2005) only one rank order is used for finding the optimal unbalanced design. In this paper, we have focused on quantile estimation and proposed a new near optimal allocation model for estimating the  $p^{th}$  ( $0 < p < 1$ ) population quantile. The probability vector  $\mathbf{q}$  consists of each rank orders in which each component becomes the ratio of number of allocations corresponding to that component to the total number of observations. For balanced RSS, each component of  $\mathbf{q}$  is taken as  $1/k$  for the set size  $k$ , however, the same of unbalanced RSS is  $m_i/n$  ( $m_i$  being the number of allocations corresponding to  $i^{th}$  order statistic).

The proposed method is shown to be more efficient asymptotically with balanced RSS method and close to optimal allocation design for large sample sizes. As seen in Figure 1, for each set size and desired  $p$ , the sample size may be chosen so that the performance of the proposed method is close to the optimum allocation model of Chen (2001). There is a major disadvantage of the RSS scheme for quantile estimation in that the performance decreases with the increase of ranking errors. This case might arise for any of the estimators and depends upon the surveyor with their expertise and tools used for the ranking purpose. It is also seen that the ARE of the proposed method tends to that of the optimum method with the increase of set size  $k$ . Our near optimal allocation model has a practical importance, and it provides an applied allocation sequence for the purpose of quantile estimation.

## References

- Al-Omari AI and Bouza CN (2014). Review of ranked set sampling: modifications and applications, *Revista Investigación Operacional*, **35**, 215–240.
- Barnett V and Moore K (1997). Best linear unbiased estimates in ranked-set sampling with particular reference to imperfect ordering, *Journal of Applied Statistics*, **24**, 697–710.
- Bhoj DS (1997). New parametric ranked set sampling, *Journal of Applied Statistical Sciences*, **6**, 275–289.

- Bhoj DS and Chandra G (2019). Simple unequal allocation procedure for ranked set sampling with skew distributions, *Journal of Modern Applied Statistical Methods*, **18**, eP2811.
- Bohn LL and Wolfe DA (1992). Nonparametric two-sample procedures for ranked set samples data, *Journal of the American Statistical Association*, **87**, 552–561.
- Bohn LL and Wolfe DA (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann Whitney-Wilcoxon statistic, *Journal of the American Statistical Association*, **89**, 168–176.
- Chandra G, Bhoj DS, and Pandey R (2018). Simple unbalanced ranked set sampling for mean estimation of response variable of developmental programs, *Journal of Modern Applied Statistical Methods*, **17**.
- Chandra G, Tiwari N, and Nautiyal R (2015). Near optimal allocation models for symmetric distributions in ranked set sampling, *Statistics in Forestry: Methods and Applications*, 85–90.
- Chen Z (2000). On ranked set sample quantiles and their applications, *Journal of Statistical Planning and Inference*, **83**, 125–135.
- Chen Z (2001). The optimal ranked set sampling scheme for inference on population quantiles, *Statistica Sinica*, **11**, 23–37.
- Cobby JM, Ridout MS, Bassett PJ, and Large RV (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards, *Grass and Forage Science*, **40**, 257–263.
- David HA and Levine DN (1972). Ranked set sampling in the presence of judgment error, *Biometrics*, **28**, 553–555.
- Dell TR and Clutter JL (1972). Ranked set sampling theory with order statistics background, *Biometrics*, **28**, 545–555.
- Halls LK and Dell TR (1966). Trials of ranked set sampling for forage yields, *Forest Science*, **12**, 22–26.
- Hettmansperger TP (1995). The ranked-set sampling sign test, *Nonparametric Statistics*, **4**, 263–70.
- Kaur A, Patil GP, and Taillie C (1997). Unequal allocation models for ranked set sampling with skew distributions, *Biometrics*, **53**, 123–130.
- Kaur A, Patil GP, and Taillie C (2000). Optimal allocation for symmetric distributions in ranked sampling, *Annals of the Institute of Statistical Mathematics*, **52**, 239–254.
- Lam K, Sinha BK, and Wu Z (1996). Estimation of location and scale parameters of a logistic distribution using a ranked set sample, *Statistical Theory and Applications*, Springer, New York.
- Latpate R, Kshirsagar J, Gupta VK, and Chandra G (2021). *Advanced Sampling Methods*, Springer.
- Martin WL, Sharik TL, Oderwald RG, and Smith DW (1980). 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*, 4–80.
- McIntyre GA (1952). A method for unbiased selective sampling using ranked sets, *Australian Journal of Agricultural Research*, **3**, 385–390.
- Nahhas RW, Wolfe DA, and Chen H (2002). Ranked set sampling: cost and optimal set size, *Biometrics*, **58**, 964–971.
- Rohatgi VK and Saleh AK (2000). *An introduction to probability and statistics (2nd Ed)*, John Wiley and Sons.
- Stark GV and Wolfe DA (2002). Evaluating ranked-set sampling estimators with imperfect rankings, *Journal of Statistical Studies*, 77–103.
- Stokes SL (1995). Parametric ranked set sampling, *Annals of the Institute of Statistical Mathematics*, **47**, 465–482.
- Stokes SL and Sager TW (1988). Characterization of a ranked set sample with application to estimat-

- ing distribution functions, *Journal of the American Statistical Association*, **83**, 374–381.
- Takahasi K and Wakimoto K (1968). 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.
- Tiwari N and Chandra G (2011). A systematic approach for unequal allocation for skewed distributions in ranked set sampling, *Journal of the Indian Society of Agricultural Statistics*, **65**, 331–338.
- Zhu M and Wang Y (2005). Quantile estimation from ranked set sampling data, *Sankhya*, **67**, 295–304.

*Received June 04, 2021; Revised September 24, 2021; Accepted September 28, 2021*