Hacettepe Journal of Mathematics and Statistics
Volume 39 (1) (2010), 89 – 95

# GENERALIZATION OF INCLUSION PROBABILITIES IN RANKED SET SAMPLING

Fikri Gokpinar*† and Yaprak Arzu Ozdemir‡

Received 02 : 02 : 2009 : Accepted 30 : 10 : 2009

**Abstract**

In a finite population setting, Ozdemir and Gokpinar (*A Generalized formula for inclusion probabilities in ranked set sampling*, Hacettepe J. Math. Stat **36** (1), 89–99, 2007) obtained a generalized formula for inclusion probabilities in Ranked Set Sampling for all set sizes when the cycle size is 1. This paper extends the generalized formula for inclusion probabilities to all set and cycle sizes. We also support this formula with a numerical example.

**Keywords:** Finite population, Inclusion probability, Ranked set sampling, Simple random sampling.

*2000 AMS Classification:* Primary 62 D 05; Secondary 65 C 60.

## 1. Introduction

Ranked Set Sampling(RSS) is a common sampling technique that has been used recently in various areas such as the environment, ecology and agriculture [6]. In these areas, measurement of the units according to the variables of interest can be quite difficult in some cases, in terms of cost, time and other factors. In such conditions, by using the RSS, the sample selection process is done with less cost and time, than the Simple Random Sampling(SRS) technique.

The sample mean of a ranked set sample is an unbiased estimator for the population mean, and the variance of this estimator is smaller than the variance of the sample mean obtained from a simple random sample of the same sample size (Takahasi and Wakimoto [15]), Dell and Clutter [3]). RSS can also be used for the estimation of the parameters

*Gazi University, Faculty of Arts and Science, Department of Statistics, Teknikokullar 06500, Ankara, Turkey; TOBB, University of Economy and Technology, Engineering Faculty, Department of Industrial Engineering, Söğütözü, Ankara, Turkey. E-mail: fikri@gazi.edu.tr

†Corresponding Author.

‡Gazi University, Faculty of Arts and Science, Department of Statistics, Teknikokullar 06500, Ankara, Turkey. E-mail: yaprak@gazi.edu.tr

of various distributions. The efficiency of these estimators depends on the main feature of the underlying distribution. In order to increase the efficiency, some modified RSS designs were suggested by Samawi *et al.* [12], Muttlak [7], Hossain [5], and Al-Saleh and Al-Omari [1] for different distribution types. These researches were developed for infinite populations.

Finite population theory in RSS was introduced by Takahasi and Futatsuya [13, 14]. Patil *et al.* [11] generalized the results of Takahasi and Futatsuya [13, 14] to a larger set size. A practical use of the RSS relative to the SRS for the estimation of the population mean and variance in a finite population setting was demonstrated by Ozturk *et al.* [10]. In these studies, they used the same selection procedure as proposed for an infinite population setting to obtain the ranked set samples. However, this procedure may cause some problems in a finite population setting. Deshpande *et al.* [4] described three different sample selection procedures called Level 0, Level 1 and Level 2 sampling, where the higher levels correspond to a greater insistence on sampling without replacement.

In finite population settings, because of the equal inclusion probability, there is no control on which element enters the sample in SRS. On the other hand, in RSS the inclusion probabilities of the population elements are different from each other, and it is difficult to determine the inclusion probabilities for all sample size cases.

Al-Saleh and Samawi [2] gave an adjusted selection procedure to obtain the inclusion probabilities in RSS based on the assumption of a finite population only when the set size was 2 and 3, and the cycle size 1. Ozdemir and Gokpinar [8] obtained the inclusion probabilities in RSS for all set sizes when the cycle size is 1. In RSS the set size is usually kept small to make the ranking error minimal. To obtain the desired sample size, the cycle size can be increased. So, taking the cycle size to be greater than 1 is more realistic in practice. Ozdemir and Gokpınar [9] adapted this procedure to Median RSS (MRSS), and made a generalization to all set and cycle sizes. In this paper, we have developed a new formula to calculate the inclusion probabilities of the population elements in RSS for all set and cycle sizes.

This paper is organized as follows. The required definitions and formulas for the inclusion probabilities for RSS in a finite population setting are given in Section 2. In section 3, a numerical example is given to illustrate the formulas of calculating the inclusion probabilities.

## 2. Computation of inclusion probabilities

Let $u_1 < u_2 < \cdots < u_N$ be the distinct ordered population elements and let $n = mr$ elements be chosen from this population by RSS, where $m$ is the set size and $r$ is the number of cycles. In this paper, the selection procedure for RSS described below is equivalent to the adjusted selection procedure of Al-Saleh and Samawi [2] and the "Level 1 sampling" selection procedure given by Deshpande *et al.* [4]. This selection procedure consists of the following steps, which are repeated for $j = 1, 2, \dots, m$ to obtain a ranked set sample of size $m$;

1. A SRS of size $m$ is selected without replacement from the population.
2. The sampled elements are ranked with respect to the variable of interest and the $j^{\text{th}}$ order statistic selected for measurement.
3. The remaining $m - 1$ elements are returned to the population.

The entire cycle is repeated $r$ times to produce a ranked set sample of size $n = mr$.

The following notation is used when calculating the inclusion probabilities based on the RSS selection procedure for any set size $m$, cycle size $r$ and population size $N$;

$A_{c,j}$: The event of choosing $u_k$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle,

$y_{c,j}$: The element selected in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle,

$$l_{c,j} = \begin{cases} 0 & y_{c,j} > u_k \\ 1 & y_{c,j} < u_k, \end{cases}$$

$B_{c,j}^{l_{c,j}} = B_j^0$: The event of $\{y_{c,j} > u_k\}$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle,

$B_{c,j}^{l_{c,j}} = B_{c,j}^1$: The event of $\{y_{c,j} < u_k\}$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle.

The number of elements smaller than $u_k$ that can be chosen previous to the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle is given by

$$a = \begin{cases} 0 & c = 1,\ j = 1 \\ \sum\limits_{z=1}^{j-1} l_{1,z} & c = 1,\ j > 1 \\ \sum\limits_{v=1}^{c-1}\sum\limits_{z=1}^{m} l_{v,z} & c > 1,\ j = 1 \\ \sum\limits_{v=1}^{c-1}\sum\limits_{z=1}^{m} l_{v,z} + \sum\limits_{z=1}^{j-1} l_{c,z} & c > 1,\ j > 1, \end{cases}$$

and the number of elements greater than $u_k$ that can be chosen previous to the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle is given by

$$b = \begin{cases} 0 & c = 1,\ j = 1 \\ j - 1 - \sum\limits_{z=1}^{j-1} l_{1,z} & c = 1,\ j > 1 \\ (c-1)m - \sum\limits_{v=1}^{c-1}\sum\limits_{z=1}^{m} l_{v,z} & c > 1,\ j = 1 \\ (c-1)m + j - 1 - (\sum\limits_{v=1}^{c-1}\sum\limits_{z=1}^{m} l_{v,z} + \sum\limits_{z=1}^{j-1} l_{c,z}) & c > 1,\ j > 1. \end{cases}$$

The inclusion probability $\pi_N(k)$ of the $k^{\text{th}}$ element $u_k$, for $k = 1, 2, \ldots, N$, are defined by

$$\pi_N(k) = \sum_{c=1}^{r} \pi_N^{(c)}(k) = \sum_{c=1}^{r}\sum_{j=1}^{m} \pi_N^{(c,j)}(k), \tag{2.1}$$

where $\pi_N^{(c)}(k)$ is the inclusion probability of $u_k$ in the $c^{\text{th}}$ cycle ($c = 1, 2, \ldots, r$) and $\pi_N^{(c,j)}(k)$ is the inclusion probability of $u_k$ in the $j^{\text{th}}$ selection ($j = 1, 2, \ldots, m$) and $c^{\text{th}}$ cycle.

Using the notation above, $\pi_N^{(c,j)}(k)$ is defined as follows,

$$\begin{aligned} \pi_N^{(c,j)}(k) = \sum P\big(A_{c,j} \mid B_{c,j-1}^{l_{c,j-1}} \cap B_{c,j-2}^{l_{c,j-2}} \cap \cdots \cap B_{1,1}^{l_{1,1}}\big) \\ \times P\big(B_{c,j-1}^{l_{c,j-1}} \mid B_{c,j-2}^{l_{c,j-2}} \cap \cdots \cap B_{1,1}^{l_{1,1}}\big) \cdots P\big(B_{1,1}^{l_{1,1}}\big), \end{aligned} \tag{2.2}$$

where the summation is over all the $2^{(c-1)m+j-1}$ possible permutations of $(l_{c,j}, \ldots, l_{c,1}, \ldots, l_{1,m}, \ldots, l_{1,1})$ (Ozdemir and Gokpınar, [8]).

In the population concerned, $a + b = (c-1)m + j - 1$ elements are chosen before the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle, according to the selection procedure for RSS given in Section 1. So, in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle, the total number of remaining elements in the population is $N - [(c-1)m + j] + 1 = N - (a + b)$. In the remaining population, $k - 1 - a$ elements are smaller than $u_k$, $N - k - b$ elements are greater than $u_k$, and there is only one $u_k$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle. The value $l_{c,j} = 0$ indicates that the chosen element $y_{c,j}$ in the $j^{\text{th}}$ selection and $c^{th}$ cycle is greater than $u_k$. One of the possible cases that give $l_{c,j} = 0$ is to choose $m$ elements from $N-k-b$ elements, another possible case is to choose $m-1$ elements from $N-k-b$ elements and to choose $u_k$, and so on. The other cases can be obtained in a similar way. So the probability of choosing an element greater than $u_k$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle under the condition that $u_k$ is not chosen in the previous selections is given by,

$$P\big(B_{c,j}^{0} \mid B_{c,j-1}^{l_{c,j-1}} \cap B_{c,j-2}^{l_{c,j-2}} \cap \cdots \cap B_{c,1}^{l_{c,1}} \cap \cdots \cap B_{1,1}^{l_{1,1}}\big) = \frac{\sum_{i=0}^{j-1} \binom{k-a}{i}\binom{N-k-b}{m-i}}{\binom{N-(a+b)}{m}}. \tag{2.3}$$

When $l_{c,j} = 1$, the probability of choosing an element smaller than $u_k$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle, under the condition that $u_k$ is not included at the previous selections, is computed as

$$P\big(B_{c,j}^{1} \mid B_{c,j-1}^{l_{c,j-1}} \cap B_{c,j-2}^{l_{c,j-2}} \cap \cdots \cap B_{c,1}^{l_{c,1}} \cap \cdots \cap B_{1,1}^{l_{1,1}}\big) = \frac{\sum_{i=0}^{m-j} \binom{k-a-1}{m-i}\binom{N-k-b+1}{i}}{\binom{N-(a+b)}{m}}. \tag{2.4}$$

Finally, for choosing $u_k$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle, $j-1$ elements must be chosen from elements that are smaller than $u_k$, and $m-j$ elements must be chosen from elements greater than $u_k$. So the probability of choosing $u_k$ in the $j^{\text{th}}$ selection and $c^{\text{th}}$ cycle is given by

$$P\big(A_{c,j} \mid B_{c,j-1}^{l_{c,j-1}} \cap B_{c,j-2}^{l_{c,j-2}} \cap \cdots \cap B_{c,1}^{l_{c,1}} \cap \cdots \cap B_{1,1}^{l_{1,1}}\big) = \frac{\binom{k-a-1}{j-1}\binom{N-k-b}{m-j}}{\binom{N-(a+b)}{m}}. \tag{2.5}$$

Using these formulas, the inclusion probabilities for all the elements in the population can be easily derived.

## 3. A numerical example

In this section, we give a numerical example to illustrate the formulas for calculating the inclusion probabilities. In this example, we take the population, set and cycle sizes $N = 10$, $m = 3$ and $r = 2$ respectively. So the sample size is $n = mr = 6$. Using Equations (2.1)-(2.5), we illustrate the calculation of the inclusion probability of the $5^{\text{th}}$ element $u_5$ in the population, in detail.

Using Equation (2.1), the inclusion probability of $u_5$ can be written as

$$\begin{aligned} \pi_{10}(5) &= \sum_{c=1}^{2} \pi_{10}^{(c)}(5) \\ &= \pi_{10}^{(1)}(5) + \pi_{10}^{(2)}(5) \\ &= \pi_{10}^{(1,1)}(5) + \pi_{10}^{(1,2)}(5) + \pi_{10}^{(1,3)}(5) + \pi_{10}^{(2,1)}(5) + \pi_{10}^{(2,2)}(5) + \pi_{10}^{(2,3)}(5). \end{aligned} \tag{3.1}$$

In Equation (3.1), we should calculate $\pi_{10}^{(c,j)}(5)$ for $c = 1,2$; $j = 1,2,3$. Using Equation (2.2), $\pi_{10}^{(1,1)}(5)$ can be calculated as follows,

$$\pi_{10}^{(1,1)}(5) = P(A_{1,1}) = \frac{\binom{4}{0}\binom{5}{2}}{\binom{10}{3}} = 0.083.$$

$\pi_{10}^{(1,2)}(5)$ can be calculated based on previous selections as follows:

$$\pi_{10}^{(1,2)}(5) = P(A_{1,2}/B_{1,1}^{0})P(B_{1,1}^{0}) + P(A_{1,2}/B_{1,1}^{1})P(B_{1,1}^{1}). \tag{3.2}$$

In Equation (3.2), $P(B^0_{1,1})$ and $P(B^1_{1,1})$ can be found from Equations (2.3) and (2.4), as below:

$$P(B^0_{1,1}) = \frac{\binom{5}{0}\binom{5}{3}}{\binom{10}{3}} = 0.083, \tag{3.3}$$

$$P(B^1_{1,1}) = \frac{\binom{4}{3}\binom{6}{0} + \binom{4}{2}\binom{6}{1} + \binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = 0.833, \tag{3.4}$$

and $P(A_{1,2}/B^0_{1,1})$, $P(A_{1,2}/B^1_{1,1})$, can be calculated from Equation (2.5):

$$P(A_{1,2}/B^0_{1,1}) = \frac{\binom{4}{1}\binom{4}{1}}{\binom{9}{3}} = 0.190, \tag{3.5}$$

$$P(A_{1,2}/B^1_{1,1}) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{9}{3}} = 0.179. \tag{3.6}$$

When we substitute the values from Equations (3.3), (3.4), (3.5) and (3.6) in Equation (3.2), $\pi^{(1,2)}_{10}(5)$ is obtained as 0.165.

In the same manner $\pi^{(1,3)}_{10}(5)$ can be calculated as follows,

$$\begin{aligned}
\pi^{(1,3)}_{10}(5) = {} & P(A_{1,3}/B^0_{1,2} \cap B^0_{1,1})P(B^0_{1,2}/B^0_{1,1})P(B^0_{1,1}) \\
& + P(A_{1,3}/B^1_{1,2} \cap B^0_{1,1})P(B^1_{1,2}/B^0_{1,1})P(B^0_{1,1}) \\
& + P(A_{1,3}/B^0_{1,2} \cap B^1_{1,1})P(B^0_{1,2}/B^1_{1,1})P(B^1_{1,1}) \\
& + P(A_{1,3}/B^1_{1,2} \cap B^1_{1,1})P(B^1_{1,2}/B^1_{1,1})P(B^1_{1,1}).
\end{aligned} \tag{3.7}$$

In Equation (3.7), the required probabilities can easily be found using equations (2.3), (2.4), (2.5), as follows.

$$P(B^0_{1,2}/B^0_{1,1}) = \frac{\binom{5}{0}\binom{4}{3} + \binom{5}{1}\binom{4}{2}}{\binom{9}{3}} = 0.405, \tag{3.8}$$

$$P(B^1_{1,2}/B^0_{1,1}) = \frac{\binom{4}{3}\binom{5}{0} + \binom{4}{2}\binom{5}{1}}{\binom{9}{3}} = 0.405, \tag{3.9}$$

$$P(B^0_{1,2}/B^1_{1,1}) = \frac{\binom{4}{0}\binom{5}{3} + \binom{4}{1}\binom{5}{2}}{\binom{9}{3}} = 0.595, \tag{3.10}$$

$$P(B^1_{1,2}/B^0_{1,1}) = \frac{\binom{3}{3}\binom{6}{0} + \binom{3}{2}\binom{6}{1}}{\binom{9}{3}} = 0.226, \tag{3.11}$$

$$P(A_{1,3}/B^0_{1,2} \cap B^0_{1,1}) = \frac{\binom{4}{2}\binom{3}{0}}{\binom{8}{3}} = 0.107, \tag{3.12}$$

$$P(A_{1,3}/B^0_{1,2} \cap B^1_{1,1}) = P(A_{1,3}/B^1_{1,2} \cap B^0_{1,1}) = \frac{\binom{3}{2}\binom{4}{0}}{\binom{8}{3}} = 0.054, \tag{3.13}$$

$$P(A_{1,3}/B^1_{1,2} \cap B^1_{1,1}) = \frac{\binom{2}{2}\binom{5}{0}}{\binom{8}{3}} = 0.018. \tag{3.14}$$

When we substitute the values from Equations (3.8)-(3.14) in Equation (3.7), $\pi^{(1,3)}_{10}(5)$ is obtained as 0.035.

In the same manner, $\pi^{(2,1)}_{10}(5)$, $\pi^{(2,2)}_{10}(5)$ and $\pi^{(2,3)}_{10}(5)$ can be found using Equations (2.2)–(2.5) as 0.076, 0.178 and 0.027, respectively. So, the inclusion probabilities of $u_5$ are obtained from Equation (3.1) as $\pi_{10}(5) = 0.565$.

For the other elements of the population, the inclusion probabilities are given in Table 1.

**Table 1 The inclusion probabilities of the $k^{th}$ element $u_k$ for $k = 1, 2, \ldots, 10$, when the set size is $m = 3$ and cycle size $r = 2$**

| | $u_1$ | $u_2$ | $u_3$ | $u_4$ | $u_5$ | $u_6$ | $u_7$ | $u_8$ | $u_9$ | $u_{10}$ | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\pi_{10}^{(1,1)}(k)$ | 0.300 | 0.234 | 0.175 | 0.125 | 0.083 | 0.050 | 0.025 | 0.008 | 0.000 | 0.000 | 1 |
| $\pi_{10}^{(1,2)}(k)$ | 0.000 | 0.039 | 0.086 | 0.131 | 0.165 | 0.180 | 0.174 | 0.142 | 0.083 | 0.000 | 1 |
| $\pi_{10}^{(1,3)}(k)$ | 0.000 | 0.000 | 0.004 | 0.016 | 0.035 | 0.063 | 0.102 | 0.159 | 0.246 | 0.375 | 1 |
| $\pi_{10}^{(1)}(k)$ | 0.300 | 0.273 | 0.265 | 0.272 | 0.284 | 0.293 | 0.301 | 0.309 | 0.329 | 0.375 | 3 |
| $\pi_{10}^{(2,1)}(k)$ | 0.300 | 0.251 | 0.185 | 0.124 | 0.076 | 0.042 | 0.018 | 0.004 | 0.000 | 0.000 | 1 |
| $\pi_{10}^{(2,2)}(k)$ | 0.000 | 0.024 | 0.069 | 0.131 | 0.178 | 0.198 | 0.190 | 0.145 | 0.065 | 0.000 | 1 |
| $\pi_{10}^{(2,3)}(k)$ | 0.000 | 0.000 | 0.001 | 0.008 | 0.027 | 0.053 | 0.092 | 0.159 | 0.285 | 0.375 | 1 |
| $\pi_{10}^{(2)}(k)$ | 0.300 | 0.275 | 0.255 | 0.263 | 0.281 | 0.294 | 0.300 | 0.308 | 0.350 | 0.375 | 3 |
| $\pi_{10}(k)$ | 0.600 | 0.548 | 0.520 | 0.534 | 0.565 | 0.587 | 0.600 | 0.617 | 0.679 | 0.750 | 6 |

From Table 1, we point out that in the first selection of both the first and second cycles, $u_9$ and $u_{10}$ have zero inclusion probabilities, since there is no chance of these elements being selected in the first order in a set of size 3. Similarly, $u_1$ and $u_{10}$ have zero inclusion probabilities in the second selection, and $u_1$ and $u_2$ have zero inclusion probabilities in the third selection for both first and second cycles. However, extreme values in the population have greater inclusion probabilities than the others in the first and second cycles. So in general, the inclusion probabilities of extreme values are larger than the others.

## 4. Conclusion

In this paper we have derived a new formula for calculating inclusion probabilities of the population elements in an RSS design. Using these inclusion probabilities it is possible to determine the probability distribution of any statistics with a ranked set sample.

## References


[1] Al-Saleh, M. F. and Al-Omari, A. *Multistage ranked set sampling*, Statistical Planning and Inference **102**, 273–286, 2002.

[2] Al-Saleh, M. F. and Samawi H. M. *A note on inclusion probability in ranked set sampling and some of its variations* Test **16** (1), 198–209, 2007.

[3] Dell, D. R. and Clutter, J. L. *Ranked set sampling theory with order statistics background*, Biometrics **28**, 545–555, 1972.

[4] Deshpande, J. V., Frey, J. and Oztürk, O. *Nonparametric ranked-set sampling confidence intervals for quantiles of a finite population*, Environmental and Ecological Statistics **13**, 25–40, 2006.

[5] Hossain, S. S. *Non-parametric selected ranked set sampling*, Biometrical Journal **43**, 97–105, 2001.

[6] McIntyre, G. A. *A method of unbiased selective sampling, using ranked sets*, Australian Journal of Agricaltural Research **3**, 385–390, 1952.

[7] Muttlak, H. A. *Median ranked set sampling*, Applied Statistical Science **6**, 245–255, 1997.

[8] Ozdemir, Y. A. and Gokpinar, F. *A Generalized formula for inclusion probabilities in ranked set sampling*, Hacettepe J. Math. Stat. **36** (1), 89–99, 2007.

[9] Ozdemir, Y. A. and Gokpinar, F. *A new formula for inclusion probabilities in median ranked set sampling*, Communications in Statistics: Theory and Methods **37**, 2022–2033, 2008.

[10] Oztürk, O., Bilgin, O. C. and Wolfe, D. A. *Estimation of population mean and variance in flock management: A ranked set sampling approach in a finite population setting*, Journal of Statistical Computation and Simulation **75**, 905–919, 2005.

[11] Patil, G. P., Sinha, A. K. and Taillie, C. *Finite population correction for ranked set sampling*, Annals of The Institute of Statistical Mathematics **47**, 621–636, 1995.

[12] Samawi, H. M., Ahmed, M. S. and Abu-Dayyeh, W. *Estimation the population mean using extreme ranked set sampling*, Biometrical Journal **38**, 577–586, 1996.

[13] Takahasi, K. and Futatsuya, M. *Ranked set sampling from a finite population*, Proceeding of the Institute Statistical Mathematics **36**, 55–68, 1988.

[14] Takahasi, K. and Futatsuya, M. *Dependence between order statistics in samples from finite population and its application to ranked set sampling*, Proceeding of the Institute Statistical Mathematics **50**, 49–70, 1998.

[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 Institude of Statistical Mathematics **21**, 249–255, 1968.