

Homepage: <https://esju.journals.ekb.eg/>

The Egyptian Statistical Journal

Print ISSN 0542-1748– Online ISSN 2786-0086



Parameter Estimation based on Neoteric Ranked Set Samples with Applications to Weibull Distribution

Mohamed A. Sabry¹, Essam A. Amin¹, Abdeltawab A. Gira^{1*}

Received 21 May 2024; revised 30 September 2024; accepted 27 October 2024

Keywords

Simple Random Sampling; Ranked Set Sampling; Neoteric Ranked Set Sampling; Estimation Parameter; Maximum Likelihood Estimation.

Abstract

This paper investigates the parameter estimation in a Weibull distribution using Neoteric Ranked Set Sampling (NRSS) designs, thus extending the earlier study by Sabry et al. (2020). Maximum likelihood estimators (MLEs) using NRSS, comparing their performance against estimators from Simple Random Sampling (SRS) and traditional Ranked Set Sampling (RSS) were derived and evaluated. A Monte Carlo simulation was used to evaluate their performances with respect to metrics that include absolute relative bias, mean squared error, and overall efficiency. The results showed that the NRSS estimators performed better than SRS and RSS estimators for a Weibull distribution. It demonstrates that NRSS-based estimators consistently outperform their SRS and RSS counterparts, exhibiting lower bias and higher efficiency. This is important because in many real-world situations, collecting data can be expensive or time-consuming. The study's findings strongly suggest that NRSS designs offer a compelling alternative for achieving more accurate parameter estimates in relevant applications.

1. Introduction

The quest for efficient sampling methods in statistical analysis has led to the development of various techniques that aim to improve upon traditional Simple Random Sampling (SRS). Among these, Ranked Set Sampling (RSS) has emerged as a powerful alternative, particularly in scenarios where the variable of interest is costly or challenging to measure directly. The concept of RSS was introduced by McIntyre in 1952 as an innovative approach to estimating mean pasture yields. Since its inception, RSS has proven to be more efficient than SRS in estimating various population parameters across a wide range of applications (Zamanzade & Al-Omari, 2016). The method's strength lies in its ability to leverage easily observable or inexpensive ordering criteria to inform the sampling process, thereby extracting more information from the available data. Building upon the RSS framework, researchers have developed several extensions to further enhance sampling efficiency. One such extension is Neoteric Ranked Set Sampling (NRSS), proposed by Al-Saleh and Al-Kadiri (2000). NRSS differentiates itself from traditional RSS and its variants by employing a streamlined single-stage rating and selection process. This approach is designed to maximize the information extracted from the ranking process while maintaining a relatively simple selection procedure.

The Weibull distribution, widely used in reliability engineering and lifetime data analysis, serves as an excellent candidate for exploring the efficacy of NRSS in parameter estimation. Its versatility in modeling various types of data make it a crucial tool in many scientific and

✉ Corresponding author*: Abdeltawab.gira@cu.edu.eg.

¹ Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt.

© 2024 by the author(s). EJS published by Faculty of Graduate Studies for Statistical Research on behalf of Cairo University. This is an open access article under the terms of the Creative Commons Attribution (CC BY) License.



engineering applications. Neoteric Ranked Set Sampling on the other hand is relatively newer in the field of ranked set sampling. In the case of NRSS, it is established to differ from DRSS and the normal RSS since it adopts a more straightforward single-stage rating and selection. In NRSS case, the initial means of forming units is by sets, and then there is a single ranking within the set of units. Then in an orderly manner one unit is selected from each set in accordance with some plan that has been adopted beforehand. This scheme is designed to the finest detail to extract most of the information possible from the ranking process while keeping the selection process simple, an aspect that could lead to even better efficiency improvements than what is provided by DRSS (Sabry & Shaaban, 2020).

Due to its rather simple structure, NRSS is most effective in situations where it is impossible or too complicated to apply multiple stages of ranking.

This paper explores the application of NRSS in estimating the parameters of the Weibull distribution, a versatile lifetime distribution commonly used in reliability engineering. We derive maximum likelihood estimators for NRSS and compare their performance with MLEs obtained from SRS and RSS through a comprehensive Monte Carlo simulation study. Our aim is to quantify the potential improvements in estimation accuracy and efficiency that NRSS can offer over traditional sampling methods when applied to the Weibull distribution.

2. Ranked Set Sampling Design

Ranked set sampling (RSS) is a cost-effective sampling technique that enhances the efficacy of population parameter estimation when measuring the variable of interest is costly or time-consuming. RSS employs auxiliary data or expert judgment to order sampling units, allowing for the selection of more informative samples without requiring exact measurements of the target variable. In comparison to simple random sampling, this ranking procedure enables the selection of a more informative sample. The RSS design can be described as follows (see Wolfe, 2004):

Step 1: Randomly select m^2 units from a target population with a cumulative distribution function (cdf) and probability density function (pdf), $F(x; \theta)$ and $f(x; \theta)$, respectively.

Step 2: Allocate the m^2 selected units as randomly as possible into m sets, each of size m .

Step 3: Arrange the units within each group based on the variable of interest. This ordering can be achieved using expert judgment, visual assessment, or a related variable that is correlated with the variable of interest. It's important to emphasize that actual measurements of the variable of interest are not necessary at this stage.

Step 4: To select a sample for measurement, include the lowest-ranked unit from the first group and the second lowest-ranked unit from the second group. Units in each group are ranked according to the variable of interest, using expert judgment, visual observation, or a related variable. At this stage, actual measurements of the variable of interest remain unnecessary, and the process continues until the highest-ranked unit from the final group is chosen.

Step 5: Repeat steps 1 to 4 for r cycles to create a sample consisting of $m \times r$ units.

In RSS uses only one observation is used from each cycle. Specifically, in the j^{th} cycle, the lowest-ranked unit $X_{(11)j}$ is selected from the set. Next, the second -lowest unit $X_{(22)j}$ is chosen from another independent set of m observations. Finally, the largest ranked unit $X_{(mm)j}$



is selected from last set of m observations. This entire process is illustrated in Figure 1.

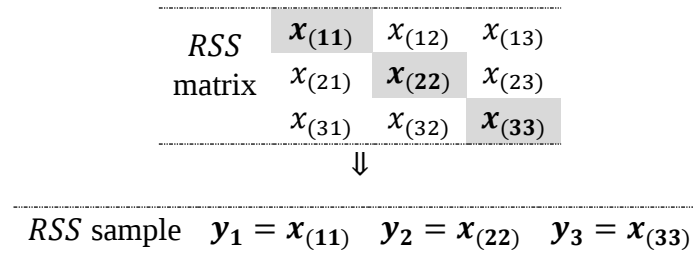


Figure 1. RSS sample display of m^2 observations in one cycle and the selected RSS sample of size $m = 3$

Let $\{X_{(ij)}, i = 1, 2, \dots, m; j = 1, 2, \dots, r\}$ represent the RSS where m denotes the set size and r represent the number of cycles. For simplicity, throughout this paper, $X_{(i)j}$ will be used instead of $X_{(ij)}$. The cumulative distribution function (cdf) and probability density function (pdf) are given by

$$F_{i:m}(x_{(i)j}; \theta) = \sum_{t=i}^m \binom{m}{t} [F(x_{(i)j}; \theta)]^t [1 - F(x_{(i)j}; \theta)]^{m-t}, \quad (1)$$

and

$$f_{i:m}(x_{(i)j}; \theta) = \frac{m!}{(i-1)!(m-i)!} f(x_{(i)j}; \theta) [F(x_{i;j}; \theta)]^{i-1} [1 - F(x_{(i)j}; \theta)]^{m-i}. \quad (2)$$

Respectively, where, $-\infty < x_{(i)j} < \infty$. The joint pdf of $x_{(i)j}, i = 1, 2, \dots, m, j = 1, 2, \dots, r$ is then given by $L(\theta; \mathbf{X}_R) = \prod_{i=1}^m f_{i:m}(x_{(i)j}; \theta)$

3. Neoteric Ranked Set Sampling Design

Neoteric Ranked Set Sampling, introduced by Zamanzade and Al-Omari (2016), presents a refined approach to ranked set sampling. NRSS sets itself apart by using a specialized selection process designed to better represent the population distribution in the final sample. The steps involved in the NRSS sampling plan are as follows:

Step 1: Randomly select m^2 units from the target population.

Step 2: Rank the m^2 sample units based on some pre-established ordering criterion;

Step 3: Select the sample unit ranked in position $[(i-1)m + l]$ th for the final sample for $i = 1, \dots, m$. If m is odd, $l = \frac{m+1}{2}$; if m is even, $l = \frac{m+2}{2}$ when i is odd and $l = m/2$ when i is even.

Step 4: Steps 1–3 can be repeated r times to obtain a final sample size $n = mr$.

Figure 2 displays the step for establishing a NRSS sample in one cycle when $m = 3$.

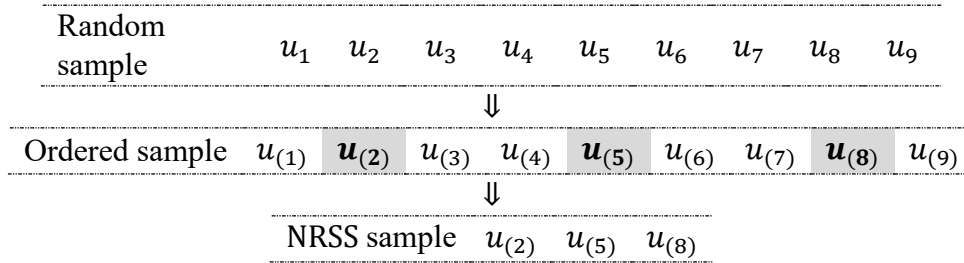


Figure 2. NRSS sample display of m^2 observations in one cycle and the selected NRSS sample of size $m = 3$

Let $\{u_i, i = 1, 2, \dots, n\}$ be a random sample of size n from a continuous population and let $\{u_{(k_i)s}, i = 1, 2, \dots, m, s = 1, 2, \dots, r\}$ be a neoteric ranked set sample drawn from a distribution with pdf $h(u; \theta)$ and cdf $H(u; \theta)$, where m is the set size, r is the number of cycles, θ is the parameter space and $n = mr$. Then, the likelihood function of NRSS samples is then given by

$$L(\theta) = \frac{m^2!}{\prod_{i=1}^{m+1} (k_i - k_{i-1} - 1)!} \prod_{s=1}^r \prod_{i=1}^m h(u_{(k_i)s}) \prod_{i=1}^{m+1} [H(u_{(k_i)s}) - H(u_{(k_{i-1})s})]^{(k_i - k_{i-1} - 1)}, \quad (3)$$

where

$$k_i = \begin{cases} \frac{m+1}{2} + (i-1)m, & m \text{ odd} \\ \frac{m}{2} + (i-1)m, & m \text{ even}, i \text{ even} \\ \frac{m+2}{2} + (i-1)m, & m \text{ even}, i \text{ odd} \end{cases} \quad (4)$$

and $k_0 = 0, k_{m+1} = m^2 + 1$ and $u_{(k_0)} = -\infty, u_{(k_{m+1})} = \infty$. For more details see Sabry and Shaaban (2020).

4. Estimation of the Weibull Distribution Parameters

In the field of reliability engineering, the Weibull distribution is a widely used model for analyzing lifetime data. It is highly adaptable and can emulate the features of other distributions, depending on its parameters. Specifically, the distribution is characterized by a shape parameter and a scale parameter, with the latter depending on the value of the shape parameter. The cumulative distribution function (CDF), probability density function (PDF), and quantile function of the Weibull distribution are defined as follow

$$F(x; \lambda, \beta) = 1 - e^{-\lambda x^\beta}, \quad (5)$$

$$f(x; \lambda, \beta) = \lambda \beta x^{\beta-1} e^{-\lambda x^\beta}, \quad (6)$$

and

$$Q(u) = \left[\frac{-\ln(1-u)}{\lambda} \right]^{\frac{1}{\beta}}, \quad (7)$$

respectively, where $x > 0, \lambda > 0, \beta > 0$ and $0 < u < 1$.

4.1. Estimation Based on SRS

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables that follow the Weibull distribution, with their probability density function (PDF) defined in Equation (6). The

likelihood function for and is expressed as

$$L(\lambda, \beta; x) = \prod_{i=1}^n \lambda \beta x_i^{\beta-1} e^{-\lambda x_i^\beta},$$

and the log likelihood function is then derived as

$$\ell(\lambda, \beta) = n \log \lambda + n \log \beta + (\beta - 1) \sum_{i=1}^n \log x_i - \sum_{i=1}^n \lambda x_i^\beta,$$

now, the likelihood equations are

$$\frac{n}{\hat{\lambda}} - \sum_{i=1}^n x_i^{\hat{\beta}} = 0, \quad (8)$$

and

$$\frac{n}{\hat{\beta}} + \sum_{i=1}^n \log x_i - \sum_{i=1}^n \hat{\lambda} x_i^{\hat{\beta}} \log x_i = 0, \quad (9)$$

It is evident that solving Equations (8) and (9) in closed form is challenging. Thus, iterative techniques are employed to calculate the maximum likelihood estimators (MLEs) for the parameters.

4.2. Estimation Based on RSS

Let $\{X_i^j, i = 1, 2, \dots, n; j = 1, 2, \dots, r\}$ be a ranked set sample with cdf and pdf given in Equations (1) and (2), where n is the set size, r is the number of cycles and $m = n r$. The Likelihood function of the RSS sample for Weibull data is given by,

$$\begin{aligned} L_r(\lambda, \beta; x) &= \prod_{j=1}^r \prod_{i=1}^n C_i f(x_{(i)j}; \lambda, \beta) [F(x_{(i)j}; \lambda, \beta)]^{i-1} [1 - F(x_{(i)j}; \lambda, \beta)]^{n-i} \\ &= \prod_{j=1}^r \prod_{i=1}^n C_i \left(\lambda \beta (x_{(i)j})^{\beta-1} e^{-\lambda (x_{(i)j})^\beta} \right) \left(1 - e^{-\lambda (x_{(i)j})^\beta} \right)^{i-1} \left(e^{-\lambda (x_{(i)j})^\beta} \right)^{n-i} \end{aligned} \quad (10)$$

where $C_i = \frac{n!}{(i-1)!(n-i)!}$. The log likelihood function can be derived directly as follows

$$\begin{aligned} \ell_r(\lambda, \beta) &\propto n r \log \lambda + n r \log \beta + (\beta - 1) \sum_{j=1}^r \sum_{i=1}^n \log x_{(i)j} - \sum_{j=1}^r \sum_{i=1}^n (n - i + 1) \lambda (x_{(i)j})^\beta \\ &\quad + \sum_{j=1}^r \sum_{i=1}^n (i - 1) \log \left(1 - e^{-\lambda (x_{(i)j})^\beta} \right), \end{aligned}$$

The likelihood equations becomes

$$\frac{n r}{\hat{\lambda}} - (n - i + 1) \sum_{j=1}^r \sum_{i=1}^n (x_{(i)j})^{\hat{\beta}} + (i - 1) \sum_{j=1}^r \sum_{i=1}^n \frac{(x_{(i)j})^{\hat{\beta}} e^{-\lambda (x_{(i)j})^{\hat{\beta}}}}{1 - e^{-\lambda (x_{(i)j})^{\hat{\beta}}}} = 0 \quad (11)$$

and

$$\begin{aligned} \frac{n r}{\hat{\beta}} + \sum_{j=1}^r \sum_{i=1}^n \log x_{(i)j} - (n - i + 1) \sum_{j=1}^r \sum_{i=1}^n \hat{\lambda} (x_{(i)j})^{\hat{\beta}} \log x_{(i)j} \\ + (i - 1) \sum_{j=1}^r \sum_{i=1}^n \frac{\lambda (x_{(i)j})^{\hat{\beta}} e^{-\lambda (x_{(i)j})^{\hat{\beta}}}}{1 - e^{-\lambda (x_{(i)j})^{\hat{\beta}}}} \log x_{(i)j} = 0, \end{aligned} \quad (12)$$

These two nonlinear equations (11) and (12) can't be solved analytically and will be solved numerically.

4.3. Estimation Based on NRSS

Let $\{u_{(k_i)}, i = 1, 2, \dots, m$ and k_i is defined as in equation (4)} be a NRSS sample where m is the set size. According to equation (3), the likelihood function of NRSS samples drawn from Weibull(λ, β) for one cycle is given by

$$L_N(\lambda, \beta; u) \propto \prod_{i=1}^m \lambda \beta (u_{(k_i)})^{\beta-1} e^{-\lambda(u_{(k_i)})^\beta} \prod_{i=1}^{m+1} \left[e^{-\lambda(u_{(k_i)})^\beta} - e^{-\lambda(u_{(k_{i-1})})^\beta} \right]^{(k_i - k_{i-1} - 1)},$$

$$\propto \lambda^m \beta^m \prod_{i=1}^m (u_{(k_i)})^{\beta-1} \prod_{i=1}^{m+1} \left[e^{-\lambda(u_{(k_i)})^\beta} - e^{-\lambda(u_{(k_{i-1})})^\beta} \right]^{(k_i - k_{i-1} - 1)},$$

The log-likelihood associated with this design is then given by

$$\ell_N(\lambda, \beta) \propto m \log \lambda + m \log \beta + (\beta - 1) \sum_{i=1}^m \log(u_{(k_i)}) + \sum_{i=1}^{m+1} (k_i - k_{i-1} - 1) \log \left[e^{-\lambda(u_{(k_i)})^\beta} - e^{-\lambda(u_{(k_{i-1})})^\beta} \right],$$

where $k_0 = 0, k_{m+1} = m^2 + 1, u_{(k_0)} = -\infty, u_{(k_{m+1})} = \infty$ The associated normal equations are directly derived as

$$\frac{\partial \ell_N}{\partial \lambda} = \frac{m}{\lambda} + \sum_{i=1}^{m+1} (k_i - k_{i-1} - 1) \frac{(u_{(k_{i-1})})^\beta e^{-\lambda(u_{(k_{i-1})})^\beta} - (u_{(k_i)})^\beta e^{-\lambda(u_{(k_i)})^\beta}}{e^{-\lambda(u_{(k_i)})^\beta} - e^{-\lambda(u_{(k_{i-1})})^\beta}}. \quad (13)$$

and

$$\frac{\partial \ell_N}{\partial \beta} = \frac{m}{\beta} + \sum_{i=1}^m \log(u_{(k_i)}) + \sum_{i=1}^{m+1} (k_i - k_{i-1} - 1) \frac{\lambda(u_{(k_{i-1})})^\beta \log u_{(k_{i-1})} e^{-\lambda(u_{(k_{i-1})})^\beta}}{e^{-\lambda(u_{(k_i)})^\beta} - e^{-\lambda(u_{(k_{i-1})})^\beta}} - \sum_{i=1}^{m+1} (k_i - k_{i-1} - 1) \frac{\lambda(u_{(k_i)})^\beta \log u_{(k_i)} e^{-\lambda(u_{(k_i)})^\beta}}{e^{-\lambda(u_{(k_i)})^\beta} - e^{-\lambda(u_{(k_{i-1})})^\beta}}. \quad (14)$$

The two nonlinear equations (13) and (14) can't be solved using analytical methods and will be solved using numerically.

5. Simulation Study

This section presents a Monte Carlo simulation aimed at evaluating the performance of maximum likelihood estimation (MLE) methods for complete sample, ranked sample, and Neoteric ranked sample designs. Data were generated from the Weibull distribution using various values of λ and β . The simulation was carried out using the R programming environment (version R4.4.1). The algorithm for the simulation is outlined as follows:

For complete samples

- Generate m random samples from Weibull distribution using the quantile function defined in equation (7) with 100,000 iterations
- Utilize different sample sizes ($m = 6, 9, 10, 15, 20, 25$ and 30), and different parameter values for λ and β are, ($\lambda = 0.5, 1.5$ and 3 ; $\beta = 0.5, 1.5$ and 3). Obtain the MLE.
- Compute the bias and mean square errors (MSE) of the estimates using equations (8) and (9).

For RSS samples:

- Simulate ranked set samples as described in Section 2.



- Repeat steps 1 and 2 for r cycles (number of cycles), ensuring that $m = nr$
- The MLE by solving Equations (11) and (12) simultaneously, then calculate the bias, MSE, and the relative efficiency of the RSS estimators compared to the SRS estimators. The relative efficiency of $\hat{\theta}_2$ compared with $\hat{\theta}_1$ is defined as

$$Eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$$

- Use the NRSS approach to simulate NRSS samples, as outlined in Section 2, for both even and odd sample sizes.
- Compute the MLE using Equations (13) and (14). Subsequently, calculate the bias, MSE, and relative efficiency of the NRSS estimators in comparison to the RSS estimators.

The results of the simulation study are reported in tables 1-3 and in the following figures (3-4)

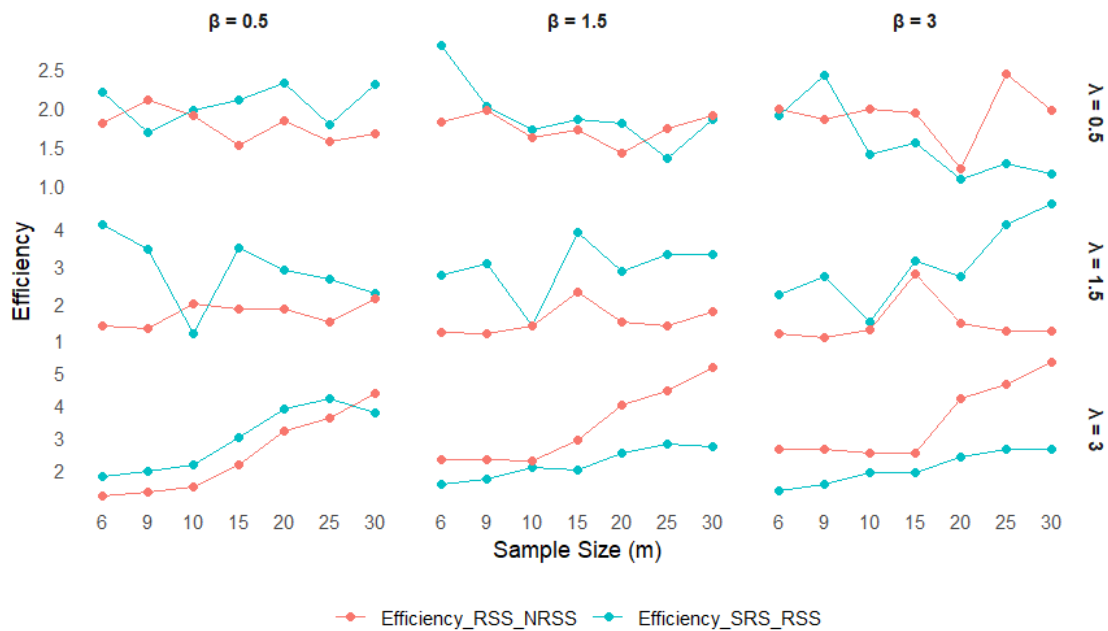


Figure 3. Efficiency Comparison for SRS/RSS and RSS/NRSS for the estimators for λ

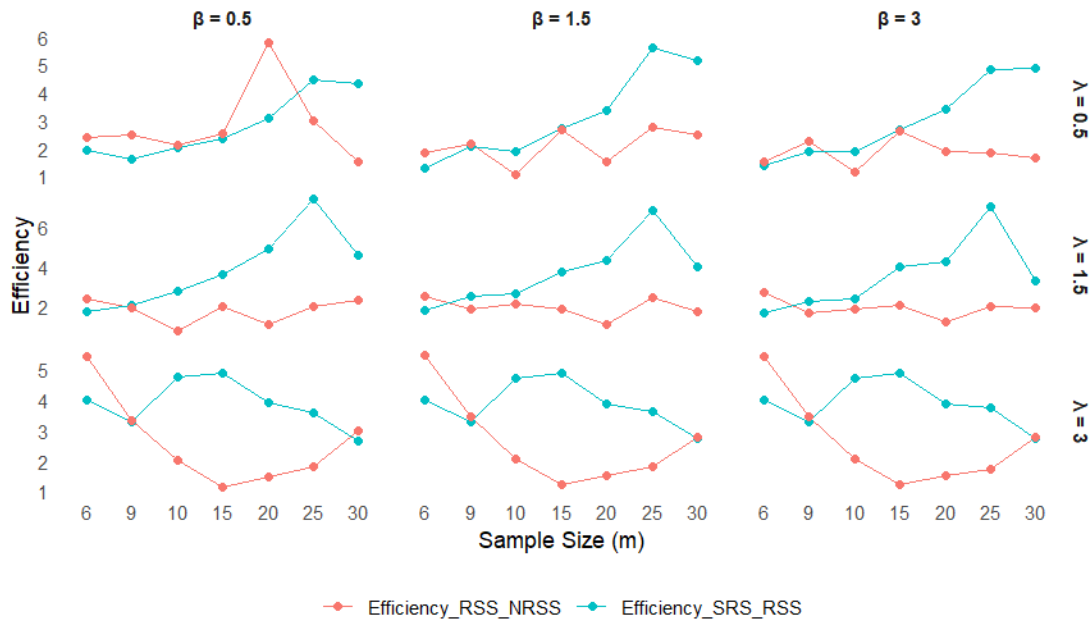


Figure 4. Efficiency Comparison for SRS/RSS and RSS/NRSS of the estimators for β

6. Discussion

For both λ and β parameters, the MSE decreases as the sample size increases across all sampling designs. For instance, the MSE for β at $\alpha=1.5$ decreases from 0.045 (when $m=10$) to 0.020 (when $m=30$) in the NRSS design. This trend is consistent across SRS and RSS designs as well.

As β increases, the MSE for both λ and β parameters also increases for a fixed sample size and λ . For example, at $\lambda=1.5$ and $m=20$, the MSE for β increases from 0.018 (when $\beta=0.5$) to 0.035 (when $\beta=3$) in the NRSS design. This indicates that higher values of β introduce more variability in the estimates.

Similarly, as λ increases, the MSE for both α and β parameters also increases. For instance, at $\beta=1.5$ and $m=20$, the MSE for λ increases from 0.015 (when $\lambda=0.5$) to 0.032 (when $\lambda=3$) in the NRSS design. This implies that higher values of λ introduce more complexity in accurately estimating the parameters.

The relative efficiencies for NRSS estimators are consistently greater than 1, indicating superior performance compared to RSS and SRS estimators. For example, the relative efficiency of NRSS compared to RSS for β is 1.25 at $\lambda=1.5$ and $m=20$. This efficiency gain can be attributed to the improved ranking and selection mechanism in NRSS, which leverages additional information from the sample units, thus providing more accurate parameter estimates.

Overall, the simulation results underscore the advantages of NRSS over traditional SRS and RSS designs, particularly in scenarios requiring precise parameter estimation for the Weibull distribution.

7. Conclusion

In this study, we investigated the effectiveness of NRSS, RSS, and SRS designs for parameter estimation in the Weibull distribution. The results from the Monte Carlo simulation clearly indicate that the NRSS design provides relatively more efficient estimates compared to the RSS and SRS designs. This is evident from the consistently higher relative efficiencies of NRSS estimators.

The NRSS design's ability to achieve lower MSE for the Weibull distribution parameters highlights its potential as a robust sampling method in reliability engineering and other fields where accurate parameter estimation is crucial. Given these findings, we recommend the use of NRSS designs for parameter estimation in Weibull distributions, especially in applications where precision is critical. Future research should continue refining NRSS methodology and exploring its use with other statistical distributions and practical scenarios. By leveraging the enhanced ranking and selection mechanisms of NRSS, practitioners can achieve more accurate and reliable parameter estimates, improving decision-making processes across various fields.

Declaration of interests

The authors declare that they have no conflict of interest.

References

- Al-Omari, A. I. (2012). Ratio estimation of population mean using auxiliary information in simple random sampling and median ranked set sampling. *Statistics & Probability Letters*, 82(11): 1883-1990. doi:[10.1016/j.spl.2012.07.001](https://doi.org/10.1016/j.spl.2012.07.001)
- Al-Saleh, M., and Al-Kadiri, M. (2000). Neoteric ranked set sampling. *Statistics & Probability Letters*, 48: 205–212.
- McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, 3(3): 385-390. <http://dx.doi.org/10.1071/AR9520385>
- Sabry, M. A., and Shaaban, M. (2020). Dependent ranked set sampling designs for parametric estimation with applications. *Annals of Data Science*, 7(2): 357–371. <https://doi.org/10.1007/s40745-020-00247-3>
- Wolfe, D. A. (2004). Ranked set sampling: An approach to more efficient data collection. *Statistical Science*, 19(4): 636–643. <http://dx.doi.org/10.1214/0883423040000000369>
- Zamanzade, E., and Al-Omari, A. I. (2016). New ranked set sampling for estimating the population mean and variance. *Hacettepe Journal of Mathematics and Statistics*, 45(6): 891–1905. doi: [10.15672/HJMS.20159213166](https://doi.org/10.15672/HJMS.20159213166)