Chilean Journal of Statistics
Vol. 5, No. 1, April 2014, 87–101

Parametric Inference
Research Paper

Normal Population Parameters Estimation Using Moving Ranked Set Sampling: Grassland Biodiversity Application

M. Al-Rawwash[1,*], M. T. Alodat[2], and N. Odat[3]

[1]Department of Mathematics, University of Sharjah, Sharjah, UAE,
[2]Department of Mathematics, Statistics and Physics, Qatar University, Doha, Qatar,
[3]Department of Biological Sciences, Al-Hussein Bin Talal University, Ma'an, Jordan

(Received: 15 September 2012 · Accepted in final form: 06 February 2014)

Abstract

In this paper, we discuss estimating the normal population parameters using moving ranked set sampling (MRSS) scheme. We derive new estimators and obtain confidence intervals for these parameters. Also, we conduct a simulation study to asses the performance and efficiency of the proposed estimators compared with their competitors based on simple random samples (SRS). Results indicate that the proposed population mean estimators are more efficient than those obtained using SRS. A grassland biodiversity example in central Europe is used to illustrate the usefulness of the proposed method in the field of Ecology.

Keywords: Ranked set sampling · Moving ranked set sampling · Confidence interval · Grassland biodiversity.

Mathematics Subject Classification: Primary 62F07 · Secondary 62D99.

1. Introduction

Order statistics in general and ranked set sampling introduced by McIntyre (1952) in particular played vital role in several areas of applications including engineering, pharmaceutical, agriculture as well as ecology. As an illustration, ecologists defined species diversity as the number of different species in a given area in ecosystems which is considered as a basic measure of biodiversity or biological diversity. Ecosystems with high species diversity gain greater resilience; consequently they will be able to recover more readily from natural stresses and disasters such as drought and human-induced habitat degradation, see Gaston and Spicer (2004). Calculating the number of species in a given ecological community requires counting all species living in that community which is a time-consuming and hard to achieve process in addition to the considerable effort required. Nonetheless, it would be relatively easy to estimate the number of species in the whole community via ranking small sets of samples (say plots) that represent the community on the basis of visual inspection of species occurrence. Patil (1995) concluded that such approaches provide a great obser-

* Emails: malrawwash@sharjah.ac.ae and alodatmts@yahoo.com

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)
© Chilean Statistical Society – Sociedad Chilena de Estadística
http://www.soche.cl/chjs

vation economy that can be achieved assuming the ability to identify a large number of sample units in the whole community under consideration.

The idea of sampling observations based on ranked data was initially proposed by McIntyre (1952) in his practical experiment to estimate the mean pasture and forage yields. This approach was explored and dubbed as *ranked set sampling* (RSS) by Halls and Dell (1966). In the last five decades, researchers investigated extensively the significant role of RSS and explored several theoretical aspects and application ideas, see Takahasi and Wakimoto (1968), Dell and Clutter (1972), Martin et al. (1980), Patil (1995), Alodat et al. (2009a), Alodat et al. (2009b) and Al-Rawwash et al. (2010). Detailed and comprehensive literature review of the RSS history can be found in Kaur et al. (1996) and Patil et al. (1999).

McIntyre (1952) proposed selecting an RSS of size m by drawing m independent random samples each of size m from the population of interest. Accordingly, the i^{th} order statistic of the i^{th} sample can be detected visually or by any crude method and the selected unit can be chosen for actual quantification. RSS is considered a very useful technique when we rely on ranking units through visual inspection which leads to a negligible cost compared with actual quantification.

In comparison with an SRS of the same size, it is known that RSS provides more representative sample from the target population and can produce more efficient estimators than those obtained using SRS (Chen (2000)). Nevertheless, the efficiency of estimators obtained via the RSS scheme is affected by the set size and the ranking error. In fact, the larger the set size, the greater the efficiency of the estimators and the larger the set size the greater the possibility of making errors through visual ranking, see Alodat and Al-Saleh (2001); Al-Saleh and Al-Omari (2002).

In order to improve the efficiency of RSS and diminish the potential ranking error effect without the need for increasing the set size, several modifications of RSS have been developed based on sampling extreme values. For example, Samawi et al. (1996) investigated the extreme ranked set sampling (ERSS) in which the smallest and largest order statistics are quantified. Another modification of the RSS scheme was proposed by Alodat and Al-Saleh (2001) to estimate the location parameter of the location family as follows:

(1) Select $m-1$ simple random samples of sizes $2,\ 3,\ \ldots, m$, respectively.
(2) For the j^{th} sample, $j = 2,\ 3,\ \ldots,\ m$ and using visual inspection or any other cheap method, identify and then quantify the j^{th} order statistic.
(3) Repeat steps 1, 2 and quantify the 1^{st} order statistic from the j^{th} sample.
(4) Repeat steps (1-3) n times to obtain a larger sample of size $2n(m-1)$ for fixed m.

Boosting the efficiency of the estimator as we increase the set size is considered an advantage of the ranking scheme proposed by Alodat and Al-Saleh (2001). Besides the ranking error possibly made by the experimenter will be minimized since spotting the minimum or the maximum of a random sample is easier than spotting other order statistics. This method was adopted later by Al-Saleh and Al-Hadrami (2003) and was referred to as the *Moving Ranked Set Sampling* (MRSS). They studied the maximum likelihood estimators of the symmetric and location families.

The main goal of this work is divided into three folds. First of all, we intend to perform a statistical inference for the normal population where the random sample is obtained via the MRSS scheme. Secondly, we propose new estimators and confidence intervals for the normal distribution parameters and finally we illustrate our method to make inference regarding grassland biodiversity. The rest of the article is organized as follows. In Section 2, we introduce the general setup and the framework of the MRSS scheme. In Section 3, we discuss the statistical properties of the parameters estimators of the normal population using MRSS. Also, we conduct simulation studies to investigate the efficiency of the proposed estimators and the effect of ranking errors on the estimation process. In section 4, we derive the confidence intervals of the parameters based on MRSS. As an illustration, we apply our estimation approach using the grassland biodiversity data in Section 5. Finally, we present our findings and conclusions in Section 6.

2. General Setup

To put the theoretical setup, let $\left\{X_{j1}^1,\ X_{j2}^1,\ldots,\ X_{jj}^1\right\}$ and $\left\{X_{j1}^2,\ X_{j2}^2,\ldots,\ X_{jj}^2\right\}$, $j=2,\ 3,\ \ldots,\ m$ be a collection of $2(m-1)$ random samples obtained from a normal distribution with mean μ and variance σ^2. Define $Y_{1j}=\min\left\{X_{j1}^1,\ X_{j2}^1,\ldots,\ X_{jj}^1\right\}$ and $Y_{2j}=\max\left\{X_{j1}^2,\ X_{j2}^2,\ldots,\ X_{jj}^2\right\}$ to be the sample observations collected using the MRSS scheme. Note that the distribution functions of Y_{1j} and Y_{2j} are given as

$$F_{1j}\left(y;\ \mu,\ \sigma\right)=1-\left[1-\Phi\left(\tfrac{y-\mu}{\sigma}\right)\right]^j,$$
$$=\quad 1-\Phi^{\mathrm{j}}\left(\tfrac{\mu-y}{\sigma}\right),$$

and

$$F_{2j}\left(y;\ \mu,\ \sigma\right)=\Phi^{\mathrm{j}}\left(\frac{y-\mu}{\sigma}\right),$$

respectively, where $\phi(x)$ denotes the probability density function and $\Phi(x)$ denotes the cumulative distribution function of the standard normal distribution. The corresponding probability density functions are

$$f_{1j}\left(y;\mu,\ \sigma\right)=\frac{j}{\sigma}\Phi^{\mathrm{j}-1}\left(\frac{\mu-y}{\sigma}\right)\phi(\frac{y-\mu}{\sigma}),$$

and

$$f_{2j}\left(y;\mu,\ \sigma\right)=\frac{j}{\sigma}\Phi^{\mathrm{j}-1}\left(\frac{y-\mu}{\sigma}\right)\phi(\frac{y-\mu}{\sigma}),$$

respectively. It is easy to show that

$$E\left(Y_{1j}\right)=\int_{-\infty}^{\infty}yf_{1j}\left(y;\ \mu,\ \sigma\right)dy=\mu+\sigma A_j,$$

$$E\left(Y_{2j}\right)=\int_{-\infty}^{\infty}yf_{2j}\left(y;\ \mu,\ \sigma\right)dy=\mu+\sigma B_j,$$

where $A_j=\int_{-\infty}^{\infty}yf_{1j}\left(y;;0,\ 1\right)dy$ and $B_j=\int_{-\infty}^{\infty}yf_{2j}\left(y;;\ 0,\ 1\right)dy$. The classical order statistics theory clearly shows that $-Y_{1j}$ and Y_{2j} have the same distribution when $\mu=0$. This allows us to conclude that $B_j=-A_j$ for all $j=2,\ 3,\ldots,m$, see Arnold et al. (1992).

3. Estimation of μ and σ based on MRSS

Suppose that $\{Y_{ij},\ i = 1, 2 \text{ and } j = 2,\ 3, \ldots, m\}$ is the selected MRSS from the target population, then the sample mean is an unbiased estimator of μ and it can be written as

$$\widehat{\mu}_{MRSS} = \frac{1}{2(m-1)} \sum_{j=2}^{m} (Y_{1j} + Y_{2j}),$$

while the variance of $\widehat{\mu}_{MRSS}$ is

$$Var\,(\widehat{\mu}_{MRSS}) = \frac{\sigma^2}{2(m-1)^2} \sum_{j=2}^{m} D_j,$$

where

$$D_j = \int_{-\infty}^{\infty} y^2 f_{2j}\,(y;\ 0,\ 1)\, dy - \left(\int_{-\infty}^{\infty} y f_{2j}\,(y; 0,\ 1)\, dy \right)^2.$$

The estimator $\widehat{\mu}_{MRSS}$ of the population mean is proposed by Alodat and Al-Saleh (2001). It can be shown easily that $D_j \leq 1$ (Arnold et al. (1992)) which implies that

$$Var\,(\widehat{\mu}_{MRSS}) = \frac{\sigma^2}{2(m-1)^2} \sum_{j=2}^{m} D_j \leq \frac{\sigma^2}{2(m-1)} \tag{1}$$

On the other hand, an unbiased estimator of μ based on a SRS, say $X_1,\ \ldots,\ X_{2(m-1)}$, is given by

$$\widehat{\mu}_{SRS} = \frac{1}{2(m-1)} \sum_{j=1}^{2(m-1)} X_j,$$

with the following variance

$$Var\,(\widehat{\mu}_{SRS}) = \frac{\sigma^2}{2(m-1)}.$$

Comparing the variance of $\widehat{\mu}_{SRS}$ with inequality (1) allows us to conclude that the estimator $\widehat{\mu}_{MRSS}$ is more efficient than $\widehat{\mu}_{SRS}$. Seeking a potential unbiased estimator of the parameter σ, we note that

$$E\,(Y_{2j} - Y_{1j}) = \sigma\,(B_j - A_j) = 2\sigma B_j, \quad \text{for } j = 2,\ 3, \ldots, m.$$

which implies that

$$E\left(\frac{Y_{2j} - Y_{1j}}{2B_j}\right) = \sigma.$$

Consequently, we introduce the following unbiased estimator of σ

$$\widehat{\sigma}_{MRSS} = \frac{1}{2(m-1)} \sum_{j=2}^{m} \frac{Y_{2j} - Y_{1j}}{B_j}.$$

The variance of $\widehat{\sigma}_{MRSS}$ is obtained as

$$\begin{aligned} Var\left(\widehat{\sigma}_{MRSS}\right) &= \frac{1}{4(m-1)^2} \sum_{j=2}^{m} \frac{Var\left(Y_{2j}\right) + Var\left(Y_{1j}\right)}{B_j^2}, \\ &= \frac{1}{2(m-1)^2} \sum_{j=2}^{m} \frac{Var\left(Y_{2j}\right)}{B_j^2}, \\ &= \frac{\sigma^2}{2(m-1)^2} \sum_{j=2}^{m} \frac{D_j}{B_j^2}. \end{aligned}$$

Lehmann (1983) introduced an unbiased estimator of σ based on a SRS of size $2\,(m-1)$ drawn from $N\left(\mu\ ,\sigma^2\right)$ as follows

$$\widehat{\sigma}_{SRS} = K_m S,$$

where $K_m = \frac{\Gamma\left(m-\frac{3}{2}\right)}{\sqrt{2}\Gamma(m-1)}$ and $S^2 = \sum_{i=1}^{2(m-1)} \left(X_i - \overline{X}\right)^2$. It has been reported that its variance is given by

$$Var\left(\widehat{\sigma}_{SRS}\right) = \sigma^2 \left[\left(m - \frac{3}{2}\right)\left(\frac{\Gamma(m-\frac{3}{2})}{\Gamma(m-1)}\right)^2 - 1\right].$$

Accordingly, the efficiencies of the MRSS estimators, $\widehat{\mu}_{MRSS}$ and $\widehat{\sigma}_{MRSS}$, relative to the corresponding SRS estimators are

$$eff\left(\widehat{\mu}_{SRS},\ \widehat{\mu}_{MRSS}\right) = \frac{m-1}{\sum_{j=2}^{m} D_j}$$

and

$$eff\left(\widehat{\sigma}_{SRS},\ \widehat{\sigma}_{MRSS}\right) = \frac{2(m-1)^2\left(\left(m-\frac{3}{2}\right)\left(\frac{\Gamma(m-\frac{3}{2})}{\Gamma(m-1)}\right)^2 - 1\right)}{\sum_{j=2}^{m} \frac{D_j}{B_j^2}}.$$

THEOREM 3.1 Based on the MRSS scheme, the best linear unbiased estimators of the normal distribution parameters μ and σ as well as their variances are

(1) $\mu^*_{MRSS} = \dfrac{1}{2D^*} \sum_{j=2}^{m} \dfrac{Y_{1j} + Y_{2j}}{D_j}$.

(2) $\sigma^*_{MRSS} = \dfrac{1}{2D^{**}} \sum_{j=2}^{m} \dfrac{B_j\left(Y_{2j} - Y_{1j}\right)}{D_j}$.

(3) $Var\left(\mu^*_{MRSS}\right) = \dfrac{\sigma^2}{2D^*}$.

(4) $Var(\sigma^*_{MRSS}) = \dfrac{\sigma^2}{2D^{**}}$,

where $D^* = \sum_{j=2}^{m} \dfrac{1}{D_j}$ and $D^{**} = \sum_{j=2}^{m} \dfrac{B_j^2}{D_j}$.

Proof. Assuming that the random variables $X_1, X_2, \ldots, X_n$ have a common mean θ and corresponding variances $\sigma_1^2,\ \sigma_2^2, \ldots, \sigma_n^2$, respectively, it is known that the best linear unbiased estimator (BLUE) of θ is $\widehat{\theta} = \sum_{k=1}^{n} w_k X_k$, where $w_k = \frac{1/\sigma_k^2}{\sum_{i=1}^{n} 1/\sigma_i^2},\ k = 1,\ 2, \ldots, n$.

In our setup, we know that $E\left(\left(\frac{1}{2}\left(Y_{1j} + Y_{2j}\right)\right)\right) = \mu$ and $Var\left(\frac{1}{2}\left(Y_{1j} + Y_{2j}\right)\right) = \frac{1}{2}\sigma^2 D_j$ which allows us to propose the following estimator as the BLUE of μ

$$\mu^*_{MRSS} = \frac{1}{2D^*}\sum_{j=2}^{m} \frac{Y_{1j} + Y_{2j}}{D_j},$$

with variance

$$Var\left(\mu^*_{MRSS}\right) = \frac{\sigma^2}{2D^*}.$$

Similarly, the BLUE of σ is

$$\sigma^*_{MRSS} = \frac{1}{2D^{**}}\sum_{j=2}^{m} \frac{B_j(Y_{2j} - Y_{1j})}{D_j},$$

with variance

$$Var(\sigma^*_{MRSS}) = \frac{\sigma^2}{2D^{**}}.$$

This motivates us to obtain and compare the efficiency of the estimators of μ and σ presented in Theorem 1 with some of the well known estimators. The following values are introduced to observe the performance of the new estimators

$$eff\left(\widehat{\mu}_{MRSS},\ \mu^*_{MRSS}\right) = \frac{1}{(m-1)^2}\left(\sum_{j=2}^{m} D_j\right)\left(\sum_{j=2}^{m} \frac{1}{D_j}\right).$$

$$eff\left(\widehat{\sigma}_{SRS}, \sigma^*_{MRSS}\right) = \left(\sum_{j=2}^{m} \frac{B_j^2}{D_j}\right)\left((2m-3)\left(\frac{\Gamma(m-\frac{3}{2})}{\Gamma(m-1)}\right)^2 - 2\right).$$

$$eff\left(\widehat{\sigma}_{MRSS}, \sigma^*_{MRSS}\right) = \frac{1}{(m-1)^2}\left(\sum_{j=2}^{m} \frac{B_j^2}{D_j}\right)\left(\sum_{j=2}^{m} \frac{D_j}{B_j^2}\right).$$

A numerical evaluation of the previous efficiencies are obtained using Mathematica Software and the results are illustrated in Table 1. It is important to point out that we drop the case $m = 2$ in Table 1 because of the invalidity of the variance of $\widehat{\sigma}_{SRS}$ when $m = 2$, see

Lehmann (1983). In order to investigate the performance of the proposed estimators, we evaluate the efficiencies for different values of m as reported in Table 1. The results reveal the following

(1) The MRSS estimators are more efficient than those obtained via SRS scheme.
(2) The efficiencies of the MRSS estimators increase as the set size increases.
(3) The performance of the classical MRSS estimator of μ is comparable to the BLUE values.
(4) The BLUE estimator of σ is better than the classical MRSS estimator when $m > 3$ and better than the SRS counterpart when $m > 4$.

Table 1. Efficiencies of the parameters estimators of μ and σ using different methods

m	$(\widehat{\mu}_{SRS}$ vs $\widehat{\mu}_{MRSS})$	$(\widehat{\mu}_{MRSS}$ vs $\mu^{*}_{MRSS})$	$(\widehat{\sigma}_{MRSS}$ vs $\sigma^{*}_{MRSS})$	$(\widehat{\sigma}_{SRS}$ vs $\sigma^{*}_{MRSS})$
3	1.61144	1.00979	0.81648	0.622311
4	1.73128	1.01817	1.01758	0.815275
5	1.83458	1.02514	1.18223	1.02184
6	1.92586	1.03094	1.32012	1.22971
7	2.00794	1.0358	1.43770	1.43502
8	2.08271	1.03991	1.53948	1.63627
9	2.15151	1.04342	1.62867	1.83285
10	2.21531	1.04644	1.70765	2.02458

To shed more light on the estimators μ^{*}_{MRSS} and σ^{*}_{MRSS}, we compare the performance of these estimators with the corresponding Maximum Likelihood Estimators (MLE). To this end, we write the likelihood function of (μ, σ) as

$$L(\mu,\sigma) = \prod_{j=2}^{m} \left(\frac{j}{\sigma}\right)^2 \Phi^{j-1}\left(\frac{\mu - Y_{1j}}{\sigma}\right) \phi\left(\frac{Y_{1j}-\mu}{\sigma}\right) \times \Phi^{j-1}\left(\frac{Y_{2j}-\mu}{\sigma}\right) \times \phi\left(\frac{Y_{2j}-\mu}{\sigma}\right)$$

and the log likelihood function is

$$L^{*}(\mu,\sigma) = c - 2(m-1)\log\sigma \; + \sum_{j=2}^{m} (j-1)\log\left(\Phi\left(\frac{\mu - Y_{1j}}{\sigma}\right) + \Phi\left(\frac{Y_{2j}-\mu}{\sigma}\right)\right) -$$

$$\frac{1}{2\sigma^2}\sum_{j=2}^{m}\left((Y_{2j}-\mu)^2 + (Y_{1j}-\mu)^2\right),$$

where c is a constant. The MLE of (μ, σ) are the solution of the system $\frac{\partial}{\partial\mu}L^{*}(\mu,\sigma) = 0$ and $\frac{\partial}{\partial\sigma}L^{*}(\mu,\sigma) = 0$, provided that the Hessian matrix at these solutions is negative. It is clear that the MLE of (μ, σ) has no closed form, thus numerical methods such as Newton-Raphson algorithm must be introduced to obtain these MLEs. For comparison purposes, we investigate the performance of the BLUE and the MLE estimators of (μ, σ) using the MRSS scheme where $\widehat{\mu}_{MRSS,\ MLE}$ and $\widehat{\sigma}_{MRSS,\ MLE}$ denote the MLE of μ and σ, respectively.

Table 2 shows the efficiencies of the BLUE estimators of μ and σ compared to their MLE counterparts for different values of μ, σ and m. Results allow us to conclude the following:

(1) The MLE estimators are better than the BLUE in general.

Table 2. Efficiencies of $\widehat{\mu}_{MRSS,\ MLE}$ against μ^*_{MRSS} (line 1) and $\widehat{\sigma}_{MRSS,\ MLE}$ against σ^*_{MRSS} (line 2) for different values of $m,\ \mu$ and σ.

σ	μ	2	3	4	5	6	7	8	9	10
0.5	-3	1.0000	0.9952	0.9781	0.9954	0.9599	0.9677	0.9625	0.9669	0.9441
		0.8906	0.9512	0.9757	0.9848	0.9887	0.9903	0.9937	0.9948	0.9962
	-1	0.9160	0.9850	0.9763	0.9698	0.9751	0.9821	0.9632	0.9453	0.9660
		0.7154	0.8787	0.9380	0.9600	0.9718	0.9803	0.9851	0.9877	0.9890
	0	0.8706	0.9761	0.9891	0.9806	0.9681	0.9642	0.9646	0.9649	0.9585
		0.2902	0.6536	0.8121	0.8696	0.9216	0.9302	0.9542	0.9674	0.9774
	1	0.9336	0.9724	0.9826	0.9688	0.9654	0.9691	0.9680	0.9638	0.9805
		0.9514	1.0863	1.0945	1.0539	1.0518	1.0392	1.0123	1.0283	1.0165
	3	1.0000	0.9959	0.9889	0.9913	0.9796	0.9537	0.9699	0.9608	0.9598
		1.1088	1.0458	1.0341	1.0205	1.0117	1.0111	1.0069	1.0059	1.0054
1.0	-3	0.6565	0.8111	0.9090	0.9327	0.9550	0.9612	0.9537	0.9588	0.9633
		0.8266	0.9253	0.9544	0.9668	0.9832	0.9880	0.9893	0.9939	0.9947
	-1	0.6185	0.8244	0.9181	0.9468	0.9660	0.9580	0.9611	0.9765	0.9729
		0.6235	0.8427	0.9061	0.9410	0.9576	0.9756	0.9779	0.9791	0.9859
	0	1.0000	0.9887	0.9815	0.9667	0.9785	0.9824	0.9700	0.9422	0.9730
		0.2795	0.6456	0.8015	0.8756	0.9122	0.9343	0.9525	0.9603	0.9716
	1	1.0000	0.9891	0.9868	0.9969	0.9733	0.9890	0.9605	0.9600	0.9539
		0.3134	0.3803	0.4977	0.5564	0.6433	0.6795	0.7260	0.7771	0.7719
	3	1.0000	0.9929	0.9842	0.9818	0.9826	0.9722	0.9566	0.9399	0.9589
		1.1526	1.1181	1.0631	1.0453	1.0349	1.0214	1.0152	1.0137	1.0132
3.0	-3	1.0000	0.9959	0.9818	0.9867	0.9609	0.9716	0.9713	0.9962	0.9558
		0.5662	0.8308	0.9089	0.9468	0.9655	0.9702	0.9766	0.9810	0.9874
	-1	1.0000	0.9980	0.9807	0.9850	0.9691	0.9895	0.9659	0.9614	0.9697
		0.3858	0.7299	0.8526	0.9057	0.9267	0.9528	0.9600	0.9778	0.9786
	0	1.0000	0.9915	0.9906	0.9750	0.9857	0.9692	0.9482	0.9572	0.9439
		0.2758	0.6416	0.7974	0.8667	0.9073	0.9359	0.9509	0.9681	0.9726
	1	1.0000	0.9909	0.9992	0.9811	0.9718	0.9672	0.9748	0.9550	0.9538
		0.1784	0.4874	0.6818	0.7906	0.8573	0.9006	0.9303	0.9468	0.9549
	3	1.0000	0.9977	0.9850	0.9904	0.9672	0.9759	0.9640	0.9624	0.9603
		0.3077	0.4226	0.4827	0.5587	0.6372	0.7275	0.7033	0.7670	0.7927

(2) The performance of MLE and BLUE of μ are comparable for different initial values.
(3) The MLE of μ is slightly better than the BLUE as m gets large.
(4) The MLE and BLUE of σ are comparable for small initial values of σ while the MLE gets better for large values of σ.

3.1 Effect of ranking errors

Ranking errors play vital role in the RSS procedure and may affect the obtained results assuming that the personal judgment error is not absent. However, we cannot ignore the ranking errors in an RSS sample where the judgment ordering of the sample units does not match the true order. Several articles discussed the effect of ranking errors on the efficiency of the RSS estimation approach. For example, Dell and Clutter (1972), Stokes (1977) and Nahhas et al. (2004) proposed different models for visual ranking errors and discussed the consequences on the estimation using RSS. In this section, we consider the model proposed by Dell and Clutter (1972) assuming that the i^{th} visual score for the i^{th} observation in RSS set is defined as $V_i = X_i + \tau_i$, where $\tau_1,\ \tau_2, ...,\ \tau_n$ are independent and identically

distributed as $N\left(0, \sigma_\tau^2\right)$ independent of the X_i's. To obtain an RSS sample with ranking errors, according to the proposed additive model, we adopt the following steps:
1. Obtain $V_i = X_i + \tau_i$, where $\tau_1,\ \tau_2, ...,\ \tau_n$ are independent and identically $N\left(0, \sigma_\tau^2\right)$.
2. Rank the V_i's in an ascending order so that we may obtain $V_{(1)} \leq V_{(2)} \leq \cdots \leq V_{(n)}$. Also, let $X_{[i]}$ denote the value of X associated with $V_{(i)}$.
3. The values $X_{[1]},\ X_{[2]},\ ...,\ X_{[n]}$ represent an RSS sample with raking errors.

To investigate the effect of ranking errors on the estimates of μ and σ obtained using MRSS, we conduct simulation studies to compare the performance of the MLE and BLUE of these parameters. We set the values of σ_τ^2 to introduce the ranking errors and allow the values of m to vary between 2 to 10. The results given in Tables 3 and 4 show the efficiencies of $\widehat{\mu}_{MRSS,\ MLE}$ with respect to μ^*_{MRSS} and $\widehat{\sigma}_{MRSS,\ MLE}$ with respect to σ^*_{MRSS}, which allows us to conclude the following:

(1) The results are decreasing in m which means that the MLE is slightly better than the BLUE for large set sizes.
(2) The efficiency gets slightly larger when σ_τ^2 gets larger. It means that μ^*_{MRSS} provides slightly better estimates in the presence of ranking errors.

Table 3. Efficiencies of $\widehat{\mu}_{MRSS,\ MLE}$ w.r.t μ^*_{MRSS} for different values of m, μ and σ, under ranking errors when $\sigma_\tau = 0.01$ (line 1) and $\sigma_\tau = 0.50$ (line 2).

m (μ, σ)	2	3	4	5	6	7	8	9	10
(0, 1)	1	0.9879	0.9842	0.9812	0.9745	0.9812	0.9647	0.9436	0.9425
	1	1.0029	1.0019	1.0058	0.9966	0.9990	1.0152	1.0227	1.0151
(-5, 2)	1	0.9978	0.9737	0.9694	0.9548	0.9716	0.9642	0.9644	0.9377
	1	0.9949	0.9839	0.9835	0.9867	0.9923	0.9943	0.9726	0.9636
(5, 2)	1	0.9964	1.0045	0.9662	0.9720	0.9776	0.9876	0.9648	0.9253
	1	0.9947	0.9770	0.9888	0.9906	0.9801	0.9738	0.9923	0.9817

Table 4. Efficiencies of $\widehat{\sigma}_{MRSS,\ MLE}$ w.r.t σ^*_{MRSS} for different values of m, μ and σ, under ranking errors when $\sigma_\tau = 0.01$ (line 1) and $\sigma_\tau = 0.50$ (line 2).

m (μ, σ)	2	3	4	5	6	7	8	9	10
(0, 1)	0.2785	0.6423	0.8143	0.8683	0.9044	0.9305	0.9510	0.9750	0.9730
	0.3051	0.7150	0.9813	1.0365	1.0984	1.1113	1.1357	1.1181	1.1257
(-5, 2)	1.0806	1.1127	1.0787	1.0521	1.0425	1.0278	1.0262	1.0204	1.0143
	0.7506	0.9245	0.9520	0.9810	0.9874	0.9985	0.9993	1.0022	1.0021
(5, 2)	1.1266	1.1381	1.0612	1.0376	1.0344	1.0344	1.0257	1.0177	1.0132
	1.0143	1.0466	1.0272	1.0167	1.0023	0.9956	0.9885	0.9912	0.9849

4. Confidence Intervals for μ and σ

In order to construct confidence intervals for μ and σ as part of our estimation strategy of the normal distribution parameters, we plan to introduce a pivotal quantity for each parameter based on classical MRSS estimator as well as the BLUE approach. These ideas will be discussed in the following two theorems

THEOREM 4.1 Assuming that we have an MRSS scheme, the $100\,(1-\alpha)\,\%$ confidence intervals for μ and σ based on pivotal quantities are given by

(1) $\frac{\widehat{\sigma}_{MRSS}}{q_{1-\frac{\alpha}{2}}} \leq \sigma \leq \frac{\widehat{\sigma}_{MRSS}}{q_{\frac{\alpha}{2}}}$
(2) $\widehat{\mu}_{MRSS} \mp \widehat{\sigma}_{MRSS} u_{1-\frac{\alpha}{2}}$,

where q_α denotes the 100α quantile of the random variable $Q = \frac{\widehat{\sigma}_{MRSS}}{\sigma}$ and U_α is the 100α quantile of the random variable $U = \frac{\widehat{\mu}_{MRSS}-\mu}{\widehat{\sigma}_{MRSS}}$.

PROOF. (1) The pivotal quantity for σ is introduced by rewriting $\widehat{\sigma}_{MRSS}$ as follows

$$\begin{aligned}
\widehat{\sigma}_{MRSS} &= \frac{1}{2\,(m-1)} \sum_{j=2}^{m} \frac{Y_{2j}-Y_{1j}}{B_j} \\
&= \frac{1}{2\,(m-1)} \sum_{j=2}^{m} \frac{\left(\mu+\sigma Z_{2j}\right)-\left(\mu+\sigma Z_{1j}\right)}{B_j} \\
&= \sigma,
\end{aligned}$$

where Z_{1j} and Z_{2j} have the pdf's $f_{1j}\,(z;;\ 0,\ 1)$ and $f_{2j}\,(z;;\ 0,\ 1)$, respectively and

$$Q = \frac{1}{2(m-1)} \sum_{j=2}^{m} \left(\frac{Z_{2j}-Z_{1j}}{B_j}\right).$$

Note that Q has a parameter-free distribution which means that the random quantity $\frac{\widehat{\sigma}_{MRSS}}{\sigma} = \mathrm{Q}$ represents a pivot for σ. Hence, the $100\,(1-\alpha)\,\%$ confidence interval for σ is

$$\frac{\widehat{\sigma}_{MRSS}}{q_{1-\frac{\alpha}{2}}} \leq \sigma \leq \frac{\widehat{\sigma}_{MRSS}}{q_{\frac{\alpha}{2}}},$$

The quantiles of Q are obtained via simulation and the previously obtained interval will be compared with the classical $100\,(1-\alpha)\,\%$ confidence interval for σ.

(2) Similarly, to obtain a pivotal quantity for μ, we assume that

$$\widehat{\mu}_{MRSS} = \frac{1}{2\,(m-1)} \sum_{j=2}^{m} (Y_{1j}+Y_{2j}) = \mu + \sigma\nabla,$$

where $\nabla = \frac{1}{2(m-1)} \sum_{j=2}^{m} (Z_{1j}+Z_{2j})$. Eventually, we get

$$\widehat{\mu}_{MRSS} - \mu = \sigma\nabla \tag{2}$$

and

$$\frac{\widehat{\sigma}_{MRSS}}{\sigma} = Q \tag{3}$$

Having in mind that both ∇ and Q have parameter-free distributions, we use (2) and

(3) to define the following pivotal quantity for μ

$$\mathrm{U} = \frac{\nabla}{\mathrm{Q}} = \frac{\widehat{\mu}_{MRSS} - \mu}{\widehat{\sigma}_{MRSS}}.$$

Therefore, the $100\,(1-\alpha)\,\%$ confidence interval for μ is

$$\widehat{\mu}_{MRSS} \mp \widehat{\sigma}_{MRSS} u_{1-\frac{\alpha}{2}},$$

where $u_{1-\frac{\alpha}{2}}$ is the $100\left(1-\frac{\alpha}{2}\right)$ quantile of the random variable U. Also, simulation will be used here to obtain the quantiles of U.

Theorem 4.2 Based on the BLUE of μ and σ using the MRSS scheme, the $100\,(1-\alpha)\,\%$ confidence intervals for μ and σ are

(1) $\mu^*_{MRSS} \pm \sigma^*_{MRSS} \tilde{u}_{1-\alpha/2}$,
(2) $\frac{\sigma^*_{MRSS}}{\tilde{q}_{1-\alpha/2}} \leq \sigma \leq \frac{\sigma^*_{MRSS}}{\tilde{q}_{\alpha/2}}$

where $\tilde{q}_{1-\alpha/2}$ and $\tilde{u}_{1-\alpha/2}$ are the $100\left(1-\frac{\alpha}{2}\right)\%$ quantiles of $\tilde{Q}$ and $\tilde{U}$, respectively such that $\tilde{Q} = \frac{\sigma^*_{MRSS}}{\sigma}$ and $\tilde{U} = \frac{D^{**}}{D^*} \frac{\sum_{j=2}^m (Z_{1j}+Z_{2j})/D_j}{\sum_{j=2}^m B_j(Z_{2j}-Z_{1j})/D_j}$.

Proof. To outline the proof, we consider the following two expressions of μ^*_{MRSS} and σ^*_{MRSS}

$$\mu^*_{MRSS} - \mu = \frac{\sigma}{2D^*} \sum_{j=2}^{m} \frac{Z_{1j} + Z_{2j}}{D_j} \tag{4}$$

and

$$\tilde{Q} = \frac{\sigma^*_{MRSS}}{\sigma} = \frac{1}{2D^{**}} \sum_{j=2}^{m} \frac{B_j(Z_{2j} - Z_{1j})}{D_j}. \tag{5}$$

Hence, the $100\,(1-\alpha)\,\%$ confidence intervals for μ and σ are

$$\mu^*_{MRSS} \pm \sigma^*_{MRSS} \tilde{u}_{1-\alpha/2},$$

and

$$\frac{\sigma^*_{MRSS}}{\tilde{q}_{1-\alpha/2}} \leq \sigma \leq \frac{\sigma^*_{MRSS}}{\tilde{q}_{\alpha/2}},$$

which completes the proof.

On the other hand, we may use the quantity $R = \frac{\widehat{\mu}_{MRSS} - \mu_0}{\widehat{\sigma}_{MRSS}}$ to test the hypothesis $H_0 :\ \mu = \mu_0$. Knowing that the distribution of R is free of μ and σ allows us to reject H_0 if $|R| > u_{1-\frac{\alpha}{2}}$. A similar test can be developed for testing σ based on the distribution of Q.

5. Application to grassland biodiversity data

The idea of RSS has been credited to McIntyre (1952) and since that time several researchers followed his footsteps and rigorous ideas were proposed in disciplines such as ecology due to the tremendous applications of RSS in that field, see Patil (1995) and Patil et al. (1999). For example, Halls and Dell (1966) utilized the RSS technique to estimate the weights of browse and herbage in a pine-hardwood forest of east Texas. They concluded that RSS is more efficient than SRS. Similarly, RSS was reported to be more robust when applied for estimating the shrub phytomass in forest stands. Further applications of RSS can be found in Evan (1967) and Cobby et al. (1985).

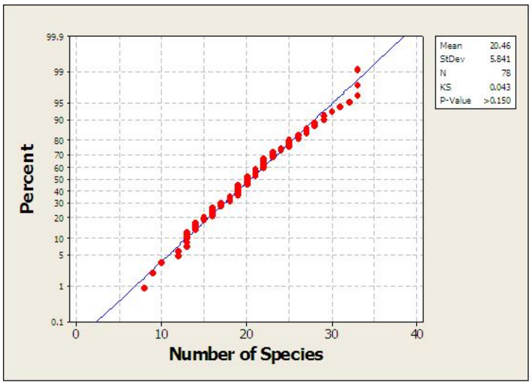

Figure 1. Histogram and normal curve for the number of species data.

Table 5. Illustration of MRSS for the Grassland Biodiversity

Sets	Minimum	Sets	Maximum
{9, 21}	**9**	{22, 13}	**22**
{15, 22, 17}	**15**	{16, 22, 31}	**31**
{18, 17, 19, 19}	**17**	{ 23, 14, 13, 19}	**23**
{8, 33, 22, 33, 23}	**8**	{16, 22, 20, 18, 21}	**22**
23, 19, 20, 19, 17, 12	**12**	{20, 25, 30, 28, 21, 27}	**30**
{18, 23, 26, 24, 13, 10, 15}	**10**	{25, 14, 28, 33, 20, 22, 21}	**33**
{23, 13, 22, 13, 19, 16, 21, 24}	**13**	{27, 25, 27, 29, 26, 14, 25, 32}	**32**

In this article, we consider the data set on grassland biodiversity in central Europe to illustrate the usefulness of our statistical method. This data set was collected based on a biodiversity project carried out in the Thueringer Schiefergebirge/Frankenwald, a plateau-like mountain range at the Thuringian/Bavarian border in central Germany with a maximum height of 870 m. Average annual temperature in the area varies between 68F and 78F and average annual precipitation varies between 950 and 1099 mm, see Perner et al. (2005). The selected plant communities located between 11.018^o and 11.638^o eastern longitudes and between 50.358^o and 50.578^o northern latitudes and were about one hectare. The collected data set represents grassland biodiversity in 78 different sites. The histogram in Figure 1 suggests a normal distribution while the normal probability plot in Figure 2 shows a Kolmogrov-Simirnov test of p-value larger than 0.15. As a result, the data provides us with no evidence to reject the normality assumption. In this section, we apply the MRSS procedure to the 78 collected observations and we restrict our application to the SRS estimators and the optimal estimators μ^*_{MRSS} and σ^*_{MRSS}. Having in mind that the population is homogeneous in all sites, we divide the first 35 observations into 7 sets of sizes 2, 3,. . . ,8, and the last 35 observations are divided in the same fashion. From each set in the first group, we select the minimum while we select the maximum from the second group (see Table 5).

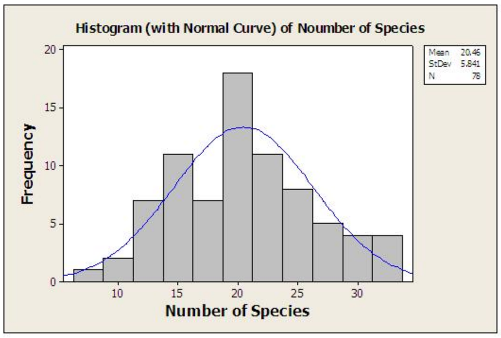

Figure 2. Normal probability plot for the number of species data.

Eventually, we get the set of minima {9, 15, 17, 8, 12, 10, 13} and the set of maxima {22, 31, 23, 22, 30, 33, 32}. We use these data to obtain estimates and confidence intervals for μ and σ. On the other hand, we obtain a simple random sample of size 14 to produce estimates and confidence intervals for μ and σ. The elements of the SRS are {9, 21, 15, 22, 17, 18, 17, 19, 19, 8, 33, 22, 33, 23}. The qauntiles of U, Q, $\tilde{U}$ and $\tilde{Q}$ are obtained via simulation based on 5000 independent random samples obtained from their distributions. The empirical quantiles of these samples are used as an approximation for the true quantiles. The results presented in this article are in agreement with previous published results. For example, Martin et al. (1980) found that RSS is more robust when applied for estimating the shrub phytomass in forest stands.

Table 6. Parameters estimates and confidence intervals for Grassland Biodiversity data using SRS and MRSS.

Method		μ	σ
SRS	Estimate	19.7143	2.028
	Variance	3.67504	2.0154
	95% confidence interval	(15.9569, 23.4717)	(5.20, 11.5559)
MRSS	Estimate	20.0996	6.17573
	Variance	1.25784	0.91305
	95% confidence interval	(17.4177, 22.7742)	(4.7565, 8.9171)

6. Conclusion

In this paper, we introduced new ideas concerning the estimation strategies of the normal distribution parameters using the MRSS scheme and we compared the results to those obtained using SRS scheme. We explored the MRSS scheme and showed how it produces efficient estimators for the normal population parameters. The simulation and grassland biodiversity example reveal the effectiveness of the MRSS scheme compared to the SRS plan. The merits of the MRSS versus SRS motivates us to use the ranking approach especially when the variable of interest is easier to rank rather than quatified. These findings are proved to be useful in the field of ecology which is illustrated in grassland biodiversity estimation example. Comparing these findings with the results reported in Alodat and

Al-Saleh (2001), we conclude that the new estimators are more efficient. Alodat and Al-Saleh (2001) limited their work to the location family which prevents us from conducting a pairwise comparison with our proposed estimators of the standard deviation. The present inference can be extended further to other families' distributions and we may also obtain prediction intervals concerning future characteristics from the normal distribution. Moreover, the present method can be used to collect data for the simple linear regression model proposed in Alodat et al. (2010).

Acknowledgements

The authors thank the editor and the reviewers for their helpful comments that helped us make substantial improvements.

References

Alodat, M. T., Alodat, N.A., Al-Rawwash, M., Alodat, T.T., 2009a. Moving extreme ranked set sampling for simple linear regression. Statistica & Applicazioni, VII(2), 135-147.

Alodat, M. T., AL-Rawwash, M., Nawajah, I. M., 2009b. Analysis of simple linear regression via median ranked set sampling. METRON - International Journal of Statistics, 67(1), 1-18.

Alodat, M. T., Al-Rawwash, M., Nawajah, I., 2010. Inference about the regression parameters using median ranked set sampling. Communications in Statistics - Theory and Methods, 39(14), 2604-2616.

Alodat, M. T., Al-Saleh, M. F., 2001. Variation of Ranked Set Sampling. Journal of Applied Statistical Sciences, 10, 137-146.

Al-Rawwash, M., Alodat, M.T., Aludaat, K. M., Odat, N., Muhaidat, R., 2010. Prediction intervals for characteristics of future normal sample under moving ranked set sampling. Statistica, 70(2), 137-152.

Al-Saleh, M. F., Al-Hadrami, S. A., 2003. Parametric estimation for the location parameter for symmetric distributions using moving extremes ranked set sampling with application to trees data. Environmetrics, 14(7), 651-664.

Al-Saleh, M. F., Al-Omari, A. I., 2002. Multistage ranked set sampling. Journal of Statistical planning and Inference, 102, 273-286.

Arnold, C. B., Balakrishnan, N., Nagaraja, H. N., 1992. A first Course in Order Statistic. John Wiley and Sons Inc, NewYork.

Chen, Z., 2000. The efficiency of ranked-set sampling relative to simple random sampling under multi-parameter families. Statistics Sinica, 10, 247-263.

Cobby, J. M., Ridout, M. S., Bassett, P. J., Large, R. V., 1985. An investigation into the use of ranked set sampling on grass-clover swards. Grass and Forage Science, 40, 257-263.

Dell, T. R., Clutter, J. L., 1972. Ranked set sampling theory with order statistics background. Biometrics, 28, 545-555.

Evan, M. J., 1967. Application of ranked set sampling to regeneration serveys in areas direct-seeded to longleaf pine. Master's Thesis, School of Forestry and Wilflife Management, Louisiana State University, Baton Rouge.

Gaston, K. J., Spicer, J. I., 2004. Biodiversity: An Introduction. Wiley-Blackwell

Halls, K. L., Dell, T. R., 1966. Trial of ranked set sampling for forage yields. Forest Science, 12, 22-32.

Kaur, A., Patil, G.P., Sinha, A.K., Taillie, C. 1996. Ranked set sampling: an annotated bibliography. Environ. Ecol. Statist., 2, 25-54.

Lehmann, E. L., 1983. Theory of point estimation. Chapman and Hall, London.

Martin, W.L., Sharik, T.L., Oderwald, R.G., Smith, D.W., 1980. Evaluation of ranked set sampling for estimating shrub phytomass in Appalachian oak forests. Publication Number FWS-4-80, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, VA.

McIntyre, G. A., 1952. A method of unbiased selective sampling using ranked sets. Australian Journal of Agricultural Research, 3, 385-390.

Nahhas, R.W., Wolfe, D.A., Chen, H. 2004. Ranked set sampling: Ranking error models and estimation of visual judgement error variance. Biometrical Journal, 46(2), 255-263.

Patil, G. P., 1995. Statistical ecology and related ecological statistics-25 years. Environmental and Ecological Statsitics, 2, 81-89.

Patil, G. P., Sinha, A.K., Taillie, C. 1999. Ranked set sampling: a bibliography. Environ. Ecol. Statist., 6, 91-98.

Perner, J., Wytrykush, C., Kahmen, A., Buchmann, N., Egerer, I., Creutzburg, S., Odat, N., Audorff, V., Weisser, W., 2005. Effects of plant diversity, plant productivity and habitat parameters on arthropod abundance in montane European grasslands. Ecography, 28, 429-442.

Stokes, S. L., 1977. Ranked set sampling with concomitant variables. Communications in Statistics - Theory and Methods, 6, 1207-1211.

Takahasi, K., 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.

Samawi, H. M., Ahmed, M. S., Abu-Dayyeh, W., 1996. Estimating the population mean using extreme ranked set sampling. The Biometrical Journal, 38(5), 577-586.