Estimation for Ranked Set Sampling Under Complete and Censored Data

Egemen, Didem[1]
Middle East Technical University
Department of Statistics
Ankara 06800, Turkey
E-mail: degemen@metu.edu.tr

Balci, Sibel[2]
Middle East Technical University
Department of Statistics
Ankara 06800, Turkey
E-mail: sbalci@metu.edu.tr

Surucu, Baris[3]
Middle East Technical University
Department of Statistics
Ankara 06800, Turkey
E-mail: sbaris@metu.edu.tr

Introduction

Ranked set sampling (RSS) is a data collection technique which has been shown to lead to more precise estimators when compared with simple random sampling. The technique was introduced by McIntyre (1952) and numerous works have been done on this subject; see Stokes and Sager (1988), Patil et al. (1994), Ozturk (1999), Yu et al (1999) and Akkaya and Balci (2006). Patil et al. (1999) and Patil (2002) give an extensive bibliography about the subject. The method has also been extended to its ordered version by Balakrishnan and Li (2006, 2008). According to the classical ranked set sampling method, n random samples of size n are drawn from a population with the probability density function (pdf) $f(x)$ and the observations are ordered visually within each sample. Then, the ith ordered observation of the ith sample is measured exactly; i=1,2,…,n. In this way, a sample of size n is obtained from the ordered observations. The process is shown by the following scheme:

RSS Scheme

$$\begin{bmatrix} X_{1(1)} & X_{1(2)} & X_{1(3)} & \cdots & X_{1(n)} \\ X_{2(1)} & X_{2(2)} & X_{2(3)} & \cdots & X_{2(n)} \\ X_{3(1)} & X_{3(2)} & X_{3(3)} & \cdots & X_{3(n)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n(1)} & X_{n(2)} & X_{n(3)} & \cdots & X_{n(n)} \end{bmatrix} \Rightarrow \begin{bmatrix} X_{1(1)} \\ X_{2(2)} \\ X_{3(3)} \\ \vdots \\ X_{n(n)} \end{bmatrix}$$

Modified Maximum Likelihood Estimation

Estimation of unknown parameters in ranked set sampling has been discussed by many authors; see for example Ozturk (2011). In this work, we will discuss the typical problems arising from the estimation of unknown parameters of a location-scale distribution and will use the modified maximum likelihood approach which was firstly proposed by Tiku (1967) to overcome these problems.

Let us firstly assume that the random variable X is coming from a location-scale distribution with pdf $\frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$. Because of the fact that observations of a ranked set sample are independent of each other, the likelihood function can be written as

$$L \propto \frac{1}{\sigma^n} \prod_{r=1}^{n} f(z_{(r)}) \prod_{r=1}^{n} [F(z_{(r)})]^{r-1} \prod_{r=1}^{n} [1-F(z_{(r)})]^{n-r} \quad (1)$$

where $z_{(r)} = (x_{(r)} - \mu)/\sigma$ and

$$f_{r:n}(x_{(r)}) = \frac{n!}{(n-r)!(r-1)!} \frac{1}{\sigma} f\left(\frac{x_{(r)}-\mu}{\sigma}\right)\left[F\left(\frac{x_{(r)}-\mu}{\sigma}\right)\right]^{r-1}\left[1-F\left(\frac{x_{(r)}-\mu}{\sigma}\right)\right]^{n-r}, -\infty < x_{(r)} < \infty; \quad (2)$$

$F(x)$ being the cumulative distribution function of the random variable X. Then, partial derivatives of the log-likelihood function to obtain the maximum likelihood estimators are given by

$$\frac{\partial \ln L}{\partial \mu} = \sum_{r=1}^{n} \frac{f'(z_{(r)})}{f(z_{(r)})} + \sum_{r=1}^{n} (r-1) \frac{f(z_{(r)})}{F(z_{(r)})} - \sum_{r=1}^{n} (n-r) \frac{f(z_{(r)})}{1-F(z_{(r)})} = 0 \quad (3)$$

$$\frac{\partial \ln L}{\partial \sigma} = n + \sum_{r=1}^{n} \left[\frac{f'(z_{(r)})}{f(z_{(r)})} z_{(r)}\right] + \sum_{r=1}^{n} \left[(r-1) \frac{f(z_{(r)})}{F(z_{(r)})} z_{(r)}\right] - \sum_{r=1}^{n} \left[(n-r) \frac{f(z_{(r)})}{1-F(z_{(r)})} z_{(r)}\right] = 0 \quad (4)$$

The equations in (3)-(4), however, do not admit explicit solutions due to the nonlinear functions $g_1(z_{(r)}) = \frac{f(z_{(r)})}{F(z_{(r)})}$ and $g_2(z_{(r)}) = \frac{f(z_{(r)})}{1-F(z_{(r)})}$ and therefore, closed form solutions of the maximum likelihood estimators cannot be obtained. This problem was also discussed by Zheng and Al-Saleh (2002). Tiku (1967) proposed the modified maximum likelihood method which uses Taylor series approximation by expanding nonlinear functions around the expected values of the standardized order statistics; see also Tiku and Suresh (1992) and Akkaya and Balci (2006). By applying Taylor series expansion to these nonlinear functions around the expected value of $z_{(r)}$, one can easily obtain the closed form solutions for the estimators which are known to be modified maximum likelihood estimators (MMLE's).

MMLE's for Normal Distribution

Suppose that the random variable X has a normal distribution with mean μ and variance σ^2. Then, the partial derivatives of the likelihood function L with respect to the location parameter μ and the scale parameter σ can be written as

$$\frac{\partial \ln L}{\partial \mu} = -\sum_{r=1}^{n} z_{(r)} + \sum_{r=1}^{n} (r-1) \frac{f(z_{(r)})}{F(z_{(r)})} - \sum_{r=1}^{n} (n-r) \frac{f(z_{(r)})}{1-F(z_{(r)})} = 0 \quad (5)$$

$$\frac{\partial \ln L}{\partial \sigma} = n - \sum_{r=1}^{n} \left(z_{(r)}\right)^2 + \sum_{r=1}^{n} \left[(r-1) \frac{f(z_{(r)})}{F(z_{(r)})} z_{(r)} \right] - \sum_{r=1}^{n} \left[(n-r) \frac{f(z_{(r)})}{1-F(z_{(r)})} z_{(r)} \right] = 0 \tag{6}$$

where

$$f(z_{(r)}) = \frac{1}{\sqrt{2\pi}} \exp\left(-z_{(r)}^2/2\right).$$

The nonlinear functions in the likelihood for a ranked set sample of a normal $N(\mu, \sigma^2)$ can be expanded as

$$g_1\left(z_{(r)}\right) = \frac{f(z_{(r)})}{F(z_{(r)})} = \alpha_{1r} + \beta_{1r} z_{(r)} \tag{7}$$

where

$$\alpha_{1r} = \left[\frac{f(t_{(r)})}{F(t_{(r)})} + \frac{t_{(r)}{}^2 f(t_{(r)})}{F(t_{(r)})} + \frac{t_{(r)} f(t_{(r)})^2}{F(t_{(r)})^2} \right],$$

$$\beta_{1r} = \left[\frac{-t_{(r)} f(t_{(r)})}{F(t_{(r)})} - \frac{f(t_{(r)})^2}{F(t_{(r)})^2} \right]$$

and

$$g_2\left(z_{(r)}\right) = \frac{f(z_{(r)})}{1-F(z_{(r)})} = \alpha_{2r} + \beta_{2r} z_{(r)} \tag{8}$$

where

$$\alpha_{2r} = \left[\frac{f(t_{(r)})}{1-F(t_{(r)})} + \frac{t_{(r)}{}^2 f(t_{(r)})}{1-F(t_{(r)})} - \frac{t_{(r)} f(t_{(r)})^2}{\left(1-F(t_{(r)})\right)^2} \right],$$

$$\beta_{2r} = \left[\frac{-t_{(r)} f(t_{(r)})}{1-F(t_{(r)})} + \frac{f(t_{(r)})^2}{\left(1-F(t_{(r)})\right)^2} \right];$$

$t_{(r)}$ being the expected value of the *r*th order statistic for the standard normal distribution. Here also note that $\alpha_{1i} = \alpha_{2(n-i+1)}$ and $\beta_{1i} = -\beta_{2(n-i+1)}$. After solving the equations (5) and (6), the resulting MML estimators are obtained as follows:

$$\hat{\mu} = \frac{-\sum_{r=1}^{n} x_{(r)} + \sum_{r=1}^{n} (r-1)\beta_{1r} x_{(r)} - \sum_{r=1}^{n} (n-r)\beta_{2r} x_{(r)}}{2\sum_{r=1}^{n} (r-1)\beta_{1r} - n} \tag{9}$$

$$\hat{\sigma} = \frac{-b + \sqrt{b^2 - 4nc}}{2n} \tag{10}$$

where

$$b=\sum_{r=1}^{n}(r-1)\alpha_{1r}x_{(r)}-\sum_{r=1}^{n}(n-r)\alpha_{2r}x_{(r)}$$

and

$$c=\hat{\mu}^2\left[2\sum_{r=1}^{n}(r-1)\beta_{1r}-n\right]+2\hat{\mu}\left[\sum_{r=1}^{n}(n-r)\beta_{2r}x_{(r)}-\sum_{r=1}^{n}(r-1)\beta_{1r}x_{(r)}+\sum_{r=1}^{n}x_{(r)}\right]$$
$$+\left[\sum_{r=1}^{n}(r-1)\beta_{1r}x_{(r)}{}^2-\sum_{r=1}^{n}(n-r)\beta_{2r}x_{(r)}{}^2-\sum_{r=1}^{n}x_{(r)}{}^2\right].$$

Unlike the classical maximum likelihood approach, MMLE's have closed form solutions and one does not need to use iterative methods to solve the likelihood equations. The MML estimator $\hat{\mu}$ is unbiased due to the symmetry. MMLE's are known to be highly efficient and robust; see Tiku and Suresh (1992). Moreover, MMLEs are asymptotically equivalent to maximum likelihood estimators under some general regularity conditions. For the proof, readers are referred to Bhattacharyya (1985) and Vaughan and Tiku (2000) for both complete and censored samples.

Fisher Information Matrix

The elements of the Fisher Information matrix can be obtained as follows:

$$I_{11}=E\left\{-\frac{\partial^2\ln L}{\partial\mu^2}\right\}$$
$$=E\left\{\frac{1}{\sigma^2}\left[n+\sum_{r=1}^{n}\left[(r-1)\frac{f(z_{(r)})}{F(z_{(r)})}\left(z_{(r)}+\frac{f(z_{(r)})}{F(z_{(r)})}\right)\right]-\sum_{r=1}^{n}\left[(n-r)\frac{f(z_{(r)})}{1-F(z_{(r)})}\left(z_{(r)}-\frac{f(z_{(r)})}{1-F(z_{(r)})}\right)\right]\right]\right\} \tag{11}$$

$$I_{12}=E\left\{-\frac{\partial^2\ln L}{\partial\mu\partial\sigma}\right\}$$
$$=E\left\{\frac{1}{\sigma^2}\left[2\sum_{r=1}^{n}z_{(r)}-\sum_{r=1}^{n}\left[(r-1)\frac{f(z_{(r)})}{F(z_{(r)})}\left(1-z_{(r)}^2-z_{(r)}\frac{f(z_{(r)})}{F(z_{(r)})}\right)\right]\right.\right.$$
$$\left.\left.+\sum_{r=1}^{n}\left[(n-r)\frac{f(z_{(r)})}{1-F(z_{(r)})}\left(1-z_{(r)}^2+z_{(r)}\frac{f(z_{(r)})}{1-F(z_{(r)})}\right)\right]\right]\right\}, \tag{12}$$

$$I_{22}=E\left\{-\frac{\partial^2\ln L}{\partial\sigma^2}\right\}$$
$$=E\left\{\frac{1}{\sigma^2}\left[-n+3\sum_{r=1}^{n}z_{(r)}^2-\sum_{r=1}^{n}\left[(r-1)z_{(r)}^2\frac{f(z_{(r)})}{F(z_{(r)})}\left(2-z_{(r)}-\frac{f(z_{(r)})}{F(z_{(r)})}\right)\right]\right.\right.$$
$$\left.\left.+\sum_{r=1}^{n}\left[(n-r)z_{(r)}^2\frac{f(z_{(r)})}{1-F(z_{(r)})}\left(2-z_{(r)}+\frac{f(z_{(r)})}{1-F(z_{(r)})}\right)\right]\right]\right\} \tag{13}$$

Due to the complicated structures in the expectation operators, one cannot obtain the explicit expressions for the elements of the exact Fisher information matrix. However, a similar approach that was used in the MMLE can be utilized to obtain approximate values of these elements.

Simulation Study

We show the high efficiency of the MML estimators obtained from the ranked set sampling by simply comparing them with those of the classical simple random sampling (SRS) estimators. For this, a Monte Carlo simulation study with 10,000 runs has been conducted. The results in following table show that MMLE's are highly efficient which, in fact, is straightforward due to the definition of the RSS.

Mean Square Error Comparisons for RSS and SRS

n	RSS		SRS	
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$
4	0.1055	0.0999	0.2496	0.6577
5	0.0692	0.0722	0.2015	0.5082
6	0.0489	0.0527	0.1679	0.3973
7	0.0375	0.0427	0.1423	0.3446
8	0.0288	0.0351	0.1272	0.2855
9	0.0234	0.0285	0.1102	0.2492
10	0.0191	0.0238	0.0999	0.2219
20	0.0050	0.0081	0.0502	0.1069
50	0.0008	0.0016	0.0195	0.0408
100	0.0002	0.0004	0.0098	0.0201

Censored Samples

Many works so far have been done on complete samples; however, a limited number of works have noted the usage of ranked set sampling for censored samples; see for example Yu and Tam (2002). Some of the samples (shown in rows in the RSS scheme) may be subject to censoring. Especially in environmental problems, some of the selected observations may not be measured as they are below a detection limit. Because of this reason, a censoring indicator should be used in the likelihood function as Yu and Tam (2002) has considered. The MML method can also be employed for the purpose of obtaining explicit solutions of estimators. However, we do not give the details here for conciseness.

REFERENCES

Akkaya, A. D. and Balci, S. (2006). Ranked set sampling, BAP-08-11-DPT2002K120510, METU BAP Project Report.

Balakrishnan, N. and Li, T. (2006). Fisher information and Tukey's linear sensitivity measure based on ordered ranked set samples, *Advances in Distribution Theory, Order Statistics, and Inference (ed N. Balakrishnan, J. M. Sarabia and E. Castillo)*, Birkhäuser Boston, 187-206.

Balakrishnan, N. and Li, T. (2008). Ordered ranked set samples and applications to inference, *Journal of Statistical*

Planning and Inference, **138**, 3512-3524.

Bhattacharyya, G.K. (1985). The asymptotics of maximum likelihood and related estimators based on type II censored data, *J. Amer. Stat. Assoc.*, **80** (1985), 398–404.

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

Ozturk, O. (1999). Two sample inference based on one sample ranked set sample sign statistics. J. *Nonparametr. Statist.*, **10**, 197-212.

Ozturk, O. (2011). Parametric estimation of location and scale parameters in ranked set sampling, *Journal of Statistical Planning and Inference*, **141**, 1616-1622.

Patil, G. P. (2002). Ranked set sampling, *Encyclopedia of Environmetrics (ed. A. H. El-Shaarawi and W. W. Piegorsch)*, John Wiley and Sons, **Vol 3**, pp 1684-1690.

Patil, G.P., Sinha, A.K. & Taillie, C. (1994). Ranked set sampling. *Handbook of Statistics*, Environmental Statistics, **Vol. 12**, G.P. Patil & C.R. Rao, eds, North-Holland, Amsterdam.

Patil, G. P., Sinha, A. K. and Taillie, C. (1999). Ranked set sampling: A bibliography, *Environmental and Ecological Statistics*, **6**, 91-98.

Stokes, S. L. and Sager, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions, *Journal of the American Statistical Association*, **83**, 374-381.

Tiku, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, **54**, 155-165.

Tiku, M. L., Suresh, R. P. (1992). A new method of estimation for location and scale parameters, *Journal of Statistical Planning and Inference*, **30**, 281-292.

Vaughan, D.C. and Tiku M. L. (2000). Estimation and hypothesis testing for a non-normal bivariate distribution, *J. Math. Comp. Modelling*, **32**, 53–67.

Yu, P.L.H., Lam, K. & Sinha, B.K. (1999). Estimation of normal variance based on balanced and unbalanced ranked set samples. *Environmental and Ecological Statistics*, **6**, 23–46.

Yu, P.L.H. and Tam, C.Y.C. (2002). Ranked set sampling in the presence of censored data. *Environmetrics*, **13(4)**, 379-396.

Zheng, G. and Al-Saleh, M. F. (2002). Modified maximum likelihood estimators based on ranked set samples, *The Institute of Statistical Mathematics*, **54**, 641-658.