
Contents

1	Introduction	1
1.1	What is ranked set sampling?	1
1.2	A historical note	7
1.3	Scope of the rest of the monograph	8
2	Balanced Ranked Set Sampling I: Nonparametric	11
2.1	Ranking mechanisms	12
2.2	Estimation of means using ranked set sample	14
2.3	Estimation of smooth-function-of-means using ranked set sample	18
2.4	Estimation of variance using an RSS sample	20
2.4.1	Naive moment estimates	20
2.4.2	Minimum variance unbiased non-negative estimate	22
2.5	Tests and confidence intervals for population mean	24
2.5.1	Asymptotic pivotal method	25
2.5.2	Choice of consistent estimates of $\sigma_{\bar{X}_{RSS}}$	25
2.5.3	Comparison of power between tests based on RSS and SRS	29
2.6	Estimation of quantiles	29
2.6.1	Ranked set sample quantiles and their properties	29
2.6.2	Inference procedures for population quantiles based on ranked set sample	32
2.6.3	The relative efficiency of RSS quantile estimate with respect to SRS quantile estimate	35
2.7	Estimation of density function with ranked set sample	35
2.7.1	RSS density estimate and its properties	36
2.7.2	The relative efficiency of the RSS density estimate with respect to its SRS counterpart	40
2.8	M-estimates with RSS data	44
2.8.1	The RSS M-estimates and their asymptotic properties	44
2.8.2	The relative efficiency of the RSS M-estimates	46

2.9	Appendix: Technical details for Section 2.4*	46
2.10	Bibliographic notes	50
3	Balanced Ranked Set Sampling II: Parametric	53
3.1	The Fisher information of ranked set samples	54
3.1.1	The Fisher information of ranked set samples when ranking is perfect	54
3.1.2	The Fisher information of ranked set samples when ranking is imperfect*	58
3.2	The maximum likelihood estimate and its asymptotic relative efficiency	63
3.3	The best linear unbiased estimation for location-scale families	66
3.4	The regularity conditions and the existence of Fisher information matrix of ranked set samples*	68
3.5	Bibliographic notes	71
4	Unbalanced Ranked Set Sampling and Optimal Designs	73
4.1	General structure of unbalanced RSS	74
4.2	Nonparametric analysis of unbalanced ranked-set data	75
4.2.1	Asymptotic properties of unbalanced ranked-set sample quantiles	75
4.2.2	Nonparametric estimation of quantiles and distribution function	76
4.2.3	Confidence interval and hypothesis testing	78
4.2.4	Estimation of statistical functionals*	79
4.2.5	Inference by bootstrapping	82
4.2.6	Simulation results	82
4.3	Parametric analysis of unbalanced ranked-set data	84
4.4	Optimal design for location-scale families	86
4.4.1	Optimal RSS for MLE	86
4.4.2	Optimal RSS for BLUE	91
4.5	Optimal design for estimation of quantiles	91
4.5.1	Optimal RSS scheme for the estimation of a single quantile	91
4.5.2	Optimal RSS scheme for the simultaneous estimation of several quantiles	94
4.6	The computation of optimal designs	95
4.7	Asymptotic relative efficiency of the optimal schemes	96
4.7.1	Asymptotic relative efficiency of the optimal schemes for parametric location-scale families	96
4.7.2	Relative efficiency of the optimal schemes for the estimation of quantiles	97
4.8	Other design methodologies	98
4.8.1	Bayesian design	99
4.8.2	Adaptive design	99

4.9	Bibliographic notes	100
5	Distribution-Free Tests with Ranked Set Sampling	103
5.1	Sign test with RSS	103
5.1.1	Distributional properties of S_{RSS}^{+*}	104
5.1.2	Decision rules of the sign test and confidence intervals for median	110
5.1.3	Effect of imperfect ranking on RSS sign test	110
5.1.4	Comparison of efficiency and power with respect to S_{SRS}^{+}	113
5.2	Mann-Whitney-Wilcoxon test with RSS	115
5.2.1	Distributional properties of U_{RSS}^{+*}	117
5.2.2	Decision rules of the RSS Mann-Whitney-Wilcoxon test	122
5.2.3	The estimate and confidence interval of Δ	123
5.2.4	Effect of imperfect ranking on RSS Mann-Whitney- Wilcoxon test	124
5.3	Wilcoxon signed rank test with RSS	124
5.3.1	Distributional properties of W_{RSS}^{+*}	125
5.3.2	Decision rules of the Wilcoxon signed rank test	130
5.3.3	Estimation of θ	131
5.3.4	Effect of imperfect ranking on RSS Wilcoxon signed rank test	133
5.4	Optimal design for distribution-free tests	133
5.4.1	Optimal design for sign test	133
5.4.2	Optimal design for Wilcoxon signed rank tests	135
5.5	Bibliographic notes	142
6	Ranked Set Sampling with Concomitant Variables	143
6.1	Multi-layer ranked set sampling	144
6.1.1	Motivation and definition	144
6.1.2	Consistency of multi-layer ranked set sampling	146
6.1.3	Comparison between multi-layer RSS and marginal RSS by simulation studies	147
6.1.4	Issues on the choice of concomitant variables	150
6.2	Adaptive ranked set sampling	151
6.2.1	The best ranking mechanism	151
6.2.2	Adaptive ranked set sampling procedure	152
6.2.3	A simulation study	153
6.3	Regression analysis based on RSS with concomitant variables	155
6.3.1	Estimation of regression coefficients with RSS	156
6.3.2	Regression estimate of the mean of Y with RSS	157
6.3.3	Comparison of the RSS regression estimate and the ordinary RSS estimate	158
6.3.4	How does a specific ranking mechanism affect the efficiency of the regression estimate?	160
6.4	Optimal RSS schemes for regression analysis	161

XII Contents

6.4.1	Asymptotically optimal criteria	162
6.4.2	Asymptotically optimal schemes for simple linear regression	164
6.4.3	Applications	166
6.5	Technical details*	170
6.6	Bibliographic notes	174
7	Ranked Set Sampling as Data Reduction Tools	175
7.1	Remedian: a motivation	175
7.2	Repeated ranked-set procedure for a single quantile	177
7.3	Repeated ranked-set procedure for multiple quantiles	180
7.4	Balanced repeated ranked-set procedure	184
7.5	Procedure for multivariate populations	186
8	Case Studies	191
8.1	Plantations of cinchona	191
8.2	Cost effective gasoline sampling	194
8.3	Estimation of weights of browse and herbage	195
8.4	Estimation of shrub phytomass in Appalachian oak forests	198
8.5	Single family homes sales price data	201
8.6	Tree data	202
	References	213
	Index	223

Introduction

Cost-effective sampling methods are of major concern in statistics, especially when the measurement of the characteristic of interest is costly and/or time-consuming. In the early 1950's, in seeking to effectively estimate the yield of pasture in Australia, McIntyre [96] proposed a sampling method which later became known as ranked set sampling (RSS). The notion of RSS provides an effective way to achieve observational economy under certain particular conditions. Although the method remained dormant for a long time, its value was rediscovered in the last 20 years or so because of its cost-effective nature. There is a surge of research interest of statisticians on RSS in recent years. There have been many new developments from the original idea of McIntyre, which made the method applicable in a much wider range of fields than originally intended. The theoretical foundation of RSS has been laid. Many variations and ramifications of the original idea have been proposed and studied. More and more applications of RSS have been cited in the literature. It is the purpose of this monograph to give a systematic account of the theory and application of RSS. In this introduction, we give a brief discussion on the notion of RSS and its applicability, a historical note on the development of RSS since it was proposed by McIntyre and, finally, an outline of the contents of the monograph.

1.1 What is ranked set sampling?

The basic premise for RSS is an infinite population under study and the assumption that a set of sampling units drawn from the population can be ranked by certain means rather cheaply *without* the actual measurement of the variable of interest which is costly and/or time-consuming. This assumption may look rather restrictive at first sight, but it turns out that there are plenty of situations in practice where this is satisfied. We shall give some examples at the end of this section.

$$\begin{array}{c}
\text{Cycle 1} \\
\mathbf{X}_{(1)11} \leq X_{(2)11} \leq X_{(3)11} \Rightarrow \mathbf{X}_{(1)1} \\
X_{(1)21} \leq \mathbf{X}_{(2)21} \leq X_{(3)21} \Rightarrow \mathbf{X}_{(2)1} \\
X_{(1)31} \leq X_{(2)31} \leq \mathbf{X}_{(3)31} \Rightarrow \mathbf{X}_{(3)1} \\
\\
\text{Cycle 2} \\
\mathbf{X}_{(1)12} \leq X_{(2)12} \leq X_{(3)12} \Rightarrow \mathbf{X}_{(1)2} \\
X_{(1)22} \leq \mathbf{X}_{(2)22} \leq X_{(3)22} \Rightarrow \mathbf{X}_{(2)2} \\
X_{(1)32} \leq X_{(2)32} \leq \mathbf{X}_{(3)32} \Rightarrow \mathbf{X}_{(3)2} \\
\\
\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
\\
\text{Cycle } m \\
\mathbf{X}_{(1)1m} \leq X_{(2)1m} \leq X_{(3)1m} \Rightarrow \mathbf{X}_{(1)m} \\
X_{(1)2m} \leq \mathbf{X}_{(2)2m} \leq X_{(3)2m} \Rightarrow \mathbf{X}_{(2)m} \\
X_{(1)3m} \leq X_{(2)3m} \leq \mathbf{X}_{(3)3m} \Rightarrow \mathbf{X}_{(3)m}
\end{array}$$

Fig. 1.1. Demonstration of the ranked set sampling procedure

The original form of RSS conceived by McIntyre can be described as follows. First, a simple random sample of size k is drawn from the population and the k sampling units are ranked with respect to the variable of interest, say X , by judgment *without* actual measurement. Then the unit with rank 1 is identified and taken for the measurement of X . The remaining units of the sample are discarded. Next, another simple random sample of size k is drawn and the units of the sample are ranked by judgment, the unit with rank 2 is taken for the measurement of X and the remaining units are discarded. This process is continued until a simple random sample of size k is taken and ranked and the unit with rank k is taken for the measurement of X . This whole process is referred to as a cycle. The cycle then repeats m times and yields a ranked set sample of size $N = mk$. For $k = 3$, the sampling procedure is illustrated in Figure 1.1.

The essence of RSS is conceptually similar to the classical stratified sampling. RSS can be considered as post-stratifying the sampling units according to their ranks in a sample. Although the mechanism is different from the stratified sampling, the effect is the same in that the population is divided into sub-populations such that the units within each sub-population are as homogeneous as possible. In fact, we can consider any mechanism, not necessarily ranking the units according to their X values, which can post-stratify the sampling units in such a way that it does not result in a random permutation of the units. The mechanism will then have similar effect to the ranking mechanism considered above. In the next chapter, we will discuss several such mechanisms. Therefore, we can extend the notion of RSS to a more general version.