Available at www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 127 (2005) 309–321
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Optimal rank set sampling estimates for a population proportion

Jeff T. Terpstra*, Eric J. Nelson

Department of Statistics, North Dakota State University, P.O. Box 5575, Waldron 201 H, Fargo, ND 58105, USA

Received 4 February 2003; accepted 7 September 2003

Abstract

This paper examines two different classes of estimates for a population proportion based on an unbalanced rank set sample. Specifically, the two classes correspond to the maximum likelihood estimator (MLE) and a weighted average (WA) estimate. Both estimators are asymptotically normal, so standard inference procedures can still be implemented. Furthermore, these results can be used to develop optimal allocation schemes for both estimators. The performances of the optimal estimators are studied in terms of both finite sample and asymptotic relative efficiency. In general, the MLE is more efficient than the WA estimate. Lastly, the practicality of the optimal sampling plans is addressed and illustrated via an example.

Keywords: Bernoulli data; Efficiency; Mean; MLE; Optimal allocation; Proportion; Rank set sample

1. Introduction

Rank set sampling (RSS) is a sampling procedure that can be viewed as a generalization of the simple random sample (SRS). It is primarily used in studies where the quantification of sampling units is, or may be, complex. For instance, sampling units may be difficult to measure, time consuming, and or costly. On the other hand, ranking the sampling units (e.g. by visual inspection) may be relatively easy. Additionally, it is well-known that population parameters can be estimated more efficiently using a RSS as opposed to a SRS.

McIntyre (1952) has been credited with introducing the idea of a RSS. Since the beginning, much has been written about both parametric and nonparametric RSS

* Corresponding author. Tel.: +1-701-231-8188; fax: +1-701-231-8734.
E-mail address: jeff.terpstra@ndsu.nodak.edu (J.T. Terpstra).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.09.020

inference procedures. In fact, a complete literature review is impractical. Nevertheless, reviews pertaining to this literature can be found in Patil et al. (1994, 1999), Bohn (1996), Kaur et al. (1996), and Muttlak and Al-Saleh (2000).

One area that has received a considerable amount of attention is unbalanced and optimal RSS designs. For example, Barabesi (2001) and Kaur et al. (2000b) discuss the rank set sample sign test under unequal allocation. Kaur et al. (1997, 2000a) and Ozturk and Wolfe (2000) present optimal RSS plans which account for different shape characteristics of the underlying distribution. Some general parametric results pertaining to URSS are given in Stokes (1995), Chen and Bai (2000), and Bhoj (2001). Finally, Chen (2001) discusses optimal rank set sampling schemes for inferences on population quantiles.

All of the above-mentioned RSS papers assume that the variable of interest is continuous. However, there is no fundamental reason why the RSS protocol cannot be applied to discrete random variables. In fact, Barnett and Barreto (2001) have successfully used the RSS methodology to estimate a Poisson parameter. Furthermore, Terpstra (2002) has demonstrated that the RSS is equally effective in estimating population proportions. For instance, suppose a researcher is interested in estimating the prevalence of college athletes who use steroids; or perhaps the proportion of adults who are obese. In both instances an expert would (perhaps by visual inspection) be able to identify steroid users and obese adults. In the case of steroid use, this would be cheaper and less time consuming than conducting urine tests for each athlete. All of the results presented by Terpstra (2002) pertain to the balanced rank set sample (BRSS).

This paper extends Terpstra's (2002) results to the unbalanced rank set sample (URSS). As we shall see, these extensions can be used to develop optimal RSS designs. The paper is outlined as follows. Section 2 discusses the methodological background of URSS in the context of Bernoulli data. The results presented in this section are used to derive optimal URSS schemes in Section 3. An example which illustrates the practicality and effectiveness of the optimal URSS procedure is given in Section 4. Section 5 presents finite sample and asymptotic relative efficiency comparisons for various designs and estimators. Finally, a discussion and some concluding remarks are given in Section 6.

2. Methodological background

Let us begin by discussing the URSS plan as it pertains to Bernoulli data. The first step is to select a positive integer, say k, to denote what is typically called the set size. Next, k sample sizes, say $n_1, n_2, \ldots, n_k$, are determined where $N = \sum_{i=1}^{k} n_i$ represents the total number of quantified observations in the URSS. Now, for $i = 1, 2, \ldots, k$, the investigator randomly selects $n_i k$ subjects from the population and then randomly divides them into n_i subgroups of size k. For each subset of k individuals, an expert then determines (without actually quantifying) which subjects are trait free ($X = 0$) and which ones actually possess ($X = 1$), the characteristic of interest. We assume throughout that the expert does not make any judgment errors. The effects of judgment error in a BRSS have been studied by Terpstra (2002). We expect that