Journal of the Korean Statistical Society ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Parametric inference based on judgment post stratified samples

Omer Ozturk [a], K.S. Sultan [b,*], M.E. Moshref [c]

[a] The Ohio State University, Department of Statistics, 1958 Neil Avenue, Columbus, OH 43210, United States
[b] King Saud University, Department of Statistics and OR, Riyadh 11451, Saudi Arabia
[c] Department of Mathematics, Faculty of Science, Al-Azhar University, Nasr City, Cairo 11884, Egypt

ARTICLE INFO

Article history:
Received 9 June 2016
Accepted 24 August 2017
Available online xxxx

AMS 2000 subject classifications:
62D05
94A20
62G05

ABSTRACT

In this paper, we consider a judgment post stratified (JPS) sample of set size H from a location and scale family of distributions. In a JPS sample, ranks of measured units are random variables. By conditioning on these ranks, we derive the maximum likelihood (MLEs) and best linear unbiased estimators (BLUEs) of the location and scale parameters. Since ranks are random variables, by considering the conditional distributions of ranks given the measured observations we construct Rao-Blackwellized version of MLEs and BLUEs. We show that Rao-Blackwellized estimators always have smaller mean squared errors than MLEs and BLUEs in a JPS sample. In addition, the paper provides empirical evidence for the efficiency of the proposed estimators through a series of Monte Carlo simulations.

1. Introduction

In settings where abundance of auxiliary information is available, judgment post-stratified (JPS) sample provides improved efficiency over its competitor simple random sampling designs. A JPS sample starts with a simple random sample (SRS) and uses available auxiliary information from additional sample units through a ranking process to determine the relative positions of the measured units in a set of size H. To construct a JPS sample from a population F, the experimenter first selects a simple random sample of size n and measures all of them. For each measured unit X_i; $i = 1, \ldots, n$, additional $H - 1$ units are selected to form a set of size H. Units in this set are ranked from smallest to largest without measurement and the rank (R_i) of the measured unit X_i is recorded. We assume that the rank of X_i is determined without an error in its set. The JPS sample then contains n fully measured units $\boldsymbol{X} = (X_1, \ldots, X_n)$ and n ranks $\boldsymbol{R} = (R_1, \ldots, R_n)$ associated with these fully measured units as additional information. This ranking information provides a mechanism to create a post stratified sample by putting homogeneous observations together in the same strata. A JPS sample in this setting can be considered as a post stratified sample since stratification is performed after a simple random sample has been collected. Increased efficiency of a JPS sample is then anticipated from the general theory of stratified sample in survey sampling designs.

Main difference between the JPS and the post-stratified sample is that post stratified sample requires a well-defined stratification variable. The measured units are then post stratified based on this stratification variable. The JPS sample, on the other hand, does not require a standard stratification variable. Post stratification can be done using visual inspection of units, judgment ranking, etc. It requires only monotonic relationship between the ranking information and variable of interest.

* Corresponding author.
E-mail address: ksultan@ksu.edu.sa (K.S. Sultan).

http://dx.doi.org/10.1016/j.jkss.2017.08.002
1226-3192/

Please cite this article in press as: Ozturk, O., et al., Parametric inference based on judgment post stratified samples. Journal of the Korean Statistical Society (2017), http://dx.doi.org/10.1016/j.jkss.2017.08.002.

There is a close connection between a JPS sample and a ranked set sample (RSS). A JPS sample conditionally on rank vector $\boldsymbol{R}$ becomes an unbalanced RSS sample. To construct an unbalanced RSS, one specifies a set size H and a vector of judgment class sample sizes, $\boldsymbol{n} = (n_1, \ldots, n_H)$, where n_h, $h = 1, 2, \ldots, H$, indicates the number of units with rank h to be selected for measurement. One then draws N independent simple random samples (sets) of size H and ranks the units within each set from smallest to largest using some method other than actual measurement. The hth judgment ranked unit is measured in n_h sets so that $\sum_{h=1}^{H} n_h = N$. The total sample then consists of $N = \sum_{h=1}^{H} n_h$ independent order statistics or judgment order statistics selected from H different judgment classes.

Descriptions of RSS and JPS designs reveal that there are two major differences between two designs. The first difference is in the nature of the ranking process. In an RSS sample, the ranking information is used prior to measurement to determine which ranked unit should be selected for full measurement in each set. The judgment rank and measured observation in a set are strongly attached to each other, and this strong attachment cannot be broken. Hence, RSS must be analyzed under this strong data structure. The procedures developed for simple random samples cannot be used. This could be a problem for settings where data is collected for multi purposes and inferential procedure for RSS has not been developed yet. To address this deficiency, MacEachern, Stasny, and Wolfe (2004) have proposed the JPS sample that uses the same ranking information that is used in RSS, but it is used post-experimentally after a simple random sample is collected. As a result, researchers who use JPS retain the option of using SRS-based inferential procedures.

The second difference is due to the distributional properties of the sample size vectors in JPS and RSS sampling designs. The sample size vector of judgment classes in an RSS sample, $\boldsymbol{n} = (n_1, \ldots, n_H)$, is a deterministic vector and must be determined prior to the construction of the sample while it is a random vector in a JPS sample. A JPS sample may be thought of as a randomized version of unbalanced RSS. It is equivalent to drawing N independent simple random samples of size H, ranking the units in each sample from smallest to largest, and then deciding at random which one unit in each sample to measure. Since the sample size vector $\boldsymbol{N}^\top = (N_1, \ldots, N_H)$ is a random vector, it may have severe unbalance among its entries, including zeros, where N_h is the number of measured units having rank h in the sample. Zero entries in $\boldsymbol{N}$ do not only reduce the efficiency, it may also lead to invalid inference, including biased estimators, inflated Type I error rates in tests and deflated coverage probabilities in confidence intervals.

Many authors have investigated the ranked set and JPS samples to draw statistical inference. Du and MacEachern (2008) have provided several estimators for the contrast parameters in a design of experiment in a JPS setting. Balakrishnan and Li (2008), Barnett and Moore (1997), Ozturk (2011) and Stokes (1995) have developed parametric inference for the parameters of location-scale family of distributions based on an RSS sample. Frey and Ozturk (2011) have developed a constrained estimation by using judgment post-stratification. Frey and Feeman (2012) have suggested an improved mean estimator for judgment post-stratification, Frey and Feeman (2013) have considered the problem of estimating the variance of a population using judgment post-stratification. Ozturk (2012) has combined the ranking information in judgment post stratified and ranked set sampling designs. Ozturk (2014) has investigated the statistical inference for population quantiles and variance in judgment post-stratified samples. Wang, Stokes, Lim, and Chen (2006) have developed a class of estimators for population mean based on concomitant of multivariate order statistics. Wang, Lim, and Stokes (2008) have used stochastic ordering constraint to construct estimators for the population mean. Chen, Ahn, Wang, and Lim (2014), Ozturk (2013) and Stokes, Wang, and Chen (2007) combined ranking information from different sources to develop statistical inference for population characteristics under a JPS sample. Up to date references for RSS and JPS sample can be found in Hollander, Wolfe, and Chicken (2014, Chapter 15) and Wolfe (2012).

In this paper, we propose parametric inference for a location and scale family of distributions F, $F(x, \boldsymbol{\theta}) = F(\frac{x-\mu}{\sigma})$, where $\boldsymbol{\theta}^\top = (\mu, \theta)$. Section 2 develops maximum likelihood estimators (MLE) for location (μ) and scale (σ) parameters of distribution F. Section 3 constructs best linear unbiased estimator (BLUE) for $\boldsymbol{\theta}$ by conditioning on the ranks of the measured observations. It is shown that the conditional variance of the BLUE estimators depend on the sample size vector $\boldsymbol{N}$. We construct unbiased estimators for the unconditional variance of BLUE estimators. Section 4 constructs Rao-Blackwellized estimators to improve the MLE and BLUE estimator. It is shown that Rao-Blackwellized estimators always perform better than regular MLE and BLUE estimators. Section 5 provides empirical evidence for the proposed estimators for small sample sizes. Finally, Section 6 provides some concluding remarks.

2. Maximum likelihood estimator

Let (X_j, R_j), $j = 1, \ldots, n$ be a JPS sample from a distribution F in a location and scale family, with location parameter μ and scale parameter σ. We construct maximum likelihood estimator of $\boldsymbol{\theta} = (\mu, \sigma)$ based on the JPS sample. The conditional distribution of X_i given $R_i = h$ is the same as the hth order statistics in a set of size H [see Arnold, Balakrishnan, and Nagaraja (1992)]

$$f_{X_i|R_i}(x; \boldsymbol{\theta}|h) = f_{(h)}(x; \boldsymbol{\theta}) = H \binom{H-1}{h-1} F^{h-1}(x; \boldsymbol{\theta})\{1 - F(x; \boldsymbol{\theta})\}^{H-h} f(x, \boldsymbol{\theta}).$$

Since R_i has discrete uniform distribution on integers $(1, \ldots, H)$, the joint distribution of (X_i, R_i) is given by

$$f_{X_i,R_i}(x, h; \boldsymbol{\theta}) = \frac{1}{H} f_{X_i|R_i}(x; \boldsymbol{\theta}|h).$$