RANKED-SET SAMPLE INFERENCE UNDER A SYMMETRY RESTRICTION

Omer Ozturk
The Ohio state University, Department of Statistics
1958 Neil Avenue
Columbus, OH USA
omer@stat.ohio-state.edu

1. INTRODUCTION

The ranked set sample is one of the cost efficient procedures for data collection. Sampling procedure involves selecting m^2 units from an infinite population F and dividing them into m sets each having m units. In each set, units are then ranked without actual quantification. The i-th ranked unit, $X_{[i]1}$, is measured in the i-th set, $i = 1, \cdots, m$. The process is repeated for n cycles to have nm quantified observations, $X_{[i]j}$, $i = 1, \cdots, m$, $j = 1, \cdots, n$. Stokes and Sager (1988) proposed an estimator for F based on ranked set sample, namely $F_n(x) = \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n}\delta(X_{[i]j} - x)$, where $\delta(a) = 1, 0$ as $a \le, > 0$. This paper introduces another estimator for the estimation of F under the symmetry assumption.

Let $\Omega(\theta)$ be the space of all continuous distributions that are symmetric about θ; i.e., $\Omega(\theta) = \{F(x) : F(x) = 1 - F(2\theta - x), \text{ for every } x\}$ and let $X_{[i]j}$, $i = 1, \cdots, m$; $j = 1, \cdots, n$ be a ranked set sample from a distribution $F \in \Omega(\theta)$. For known θ we propose to estimate $F \in \Omega(\theta)$ with $G_n(x;\theta) = \{F_n(x) + 1 - F_n(2\theta - x)\}/2$.

Theorem 1 *Assume* $F \in \Omega(\theta)$ *and* $F_{[i]}(x) = 1 - F_{[m+1-i]}(2\theta - x)$. *Then*

$$var\{F_n(x)\} - var\{G_n(x;\theta)\} = \begin{cases} \frac{1}{2nm^2}\sum_{i=1}^{m} F_{[i]}(x)\Delta_{[i]}(x;\theta) & x \le \theta \\ \frac{1}{2nm^2}\sum_{i=1}^{m}\{1 - F_{[i]}(x)\}\{2 - \Delta_{[i]}(x;\theta)\} & x > \theta, \end{cases}$$

where $\Delta_{[i]}(x;\theta) = 2 - F_{[i]}(2\theta - x) - F_{[i]}(x)$.

We note that the quantity $G_n(x;\theta)$ as it is defined is not truly an estimator of F unless θ is known.

Theorem 2 *Assume that* $\hat{\theta}$ *is an unbiased and scale invariant estimator with variance* τ^2. *Then*

$$var\{\sqrt{n}F_n(x)\} - var\{\sqrt{n}G_n(x;\hat{\theta})\} = \begin{cases} \frac{1}{2m^2}\sum_{i=1}^{m} F_{[i]}(x)\Delta_{[i]}(x;\theta) - \tau^2 f^2(x) & x \le \theta, \\ \frac{1}{2m^2}\sum_{i=1}^{m} \bar{F}_{[i]}(x)\{2 - \Delta_{[i]}(x;\theta)\} - \tau^2 f^2(x) & x > \theta, \end{cases}$$

where $\bar{F}_{[i]}(x) = 1 - F_{[i]}(x)$.

It is clear from the above theorem that the improvement for $G_n(x;\hat{\theta})$ depends on the estimator $\hat{\theta}$ through its asymptotic variance.

2. ESTIMATION OF THE POINT OF SYMMETRY

Let $W_n(\theta) = \int_{-\infty}^{\infty} \{F_n(y) - G_n(y;\theta)\}^2 dy$ be the Cramér-von Mises distance function, which measures the distance between $F_n(y)$ and $G_n(y;\theta)$. We estimate the the point of symmetry by minimizing $W_n(\theta)$ with respect to θ. Let $\hat{\theta}_W = \min_{\theta \in R} W_n(\theta)$. Then $\hat{\theta}_W$ is the median of the all Walsh averages and its limiting distribution is normal with mean θ and variance $\tau^2 = \frac{1}{6m(m+1)\{\int f^2(y)dy\}^2}$.

3. RESULTS

The proposed estimator, under the symmetry assumption, outperforms the Stokes and Sager(1988) estimator for all practical symmetric distribution. For a known center of symmetry a Kolmogorov-Smirnov type test is developed. This test is four times more efficient than its competitor based on Stokes and Sager estimator.

REFERENCES

Stokes, S. L., Sager, T. W. (1988). Characterizations of a ranked-set sample with application to estimating distribution functions. *J. Amer. Statist. Assoc.*, **83**, 374-381.

RESUME

This paper considers statistical inference for a ranked set sample under a symmetry restriction on the underlying distribution. We present new estimators for the distribution function and the center of symmetry. It is shown that these estimators outperform their competitors in the literature. Based on the proposed distribution function estimator, a Kolmogorov-Smirnov type test is developed and the construction of a confidence interval is discussed.