

# A Resampling Method for Variance Estimation with Unequal Probability Sample

Kyuseong Kim

*Department Computer Science and Statistics*

*The University of Seoul, Seoul, 130-743*

*Republic of Korea*

*kskim@uoscc.uos.ac.kr*

Jukyung Jung

*Samsung Fire and Marine Insurance*

*Seoul, 100-191*

*Republic of Korea*

*hippasos@samsung.co.kr*

## 1. Introduction

Variance estimation methods for non-linear statistics such as ratio, regression, and correlation coefficients have been widely studied under complex sampling design. Unlike iid sample complex sample may be correlated each other and may have different selection probabilities, conventional resampling methods developed from iid sample can not be directly adapted to complex sample.

Lots of resampling methods originated from iid bootstrap have been proposed for complex sample. Under stratified random sampling without replacement, rescaling method (Rao and Wu, 1988), mirror match method (Sitter, 1992) and extended without replacement bootstrap method (Sitter, 1992) are suggested. These methods are concerned with effects of without replacement sampling.

In the following, we examined resampling methods for unequal probability sample in single stage as well as multi-stage sampling.

## 2. Resampling methods for unequal probability sample

For the probability proportional to size(PPS) sample with replacement in single stage, we consider two resampling methods as follows. The first, we apply the same PPS procedure of the original sample from the population to subsampling from the original sample. Also the same type of PPS estimator is obtained. The second, we can regard PPS sample as iid after dividing the observed value  $y_i$  by its selection probability  $p_i$ , and then traditional bootstrap method can be applied to the scaled data. The following simulation results show that the second method

works well, which is as efficient as linearization method and jackknife method. But the the first method seriously underestimate the variance for the mean and slightly overestimate of the ratio. In the Table 1, RB stands for relative bias of the estimator, RV for relative bias of variance and RS for relative stability.

**Table 1. Comparison of variance estimation methods for PPS sample**

Methods	mean			ratio		
	RB	RV	RS	RB	RV	RS
linearization	0.052	0.813	41.512	0.028	2.392	49.100
jackknife	0.052	0.813	41.512	0.027	2.239	49.399
resampling 1	0.034	-37.272	31.357	0.129	6.744	57.194
resampling 2	-0.011	0.634	47.500	-0.012	3.822	54.196

For the general linear unbiased estimator under fixed size design, Rao and Wu (1988) proposed a resampling method for estimating the variance of the estimator, which is developed in the single-stage sampling. Extending the method to multi-stage sampling where the first sampling unit is selected with unequal probability without replacement, we construct a resampling method as follows. First of all, we consider a general linear estimator for the population total,

$$\hat{Y} = \sum_{i \in s} w_{si} \hat{Y}_i$$

where  $\hat{Y}_i$  is unbiased estimator of the  $i$ th cluster total  $Y_i$ . Then we derived a necessary form of nonnegative quadratic unbiased variance estimator of  $\hat{Y}$ . And we found a bootstrap replicate adding two components to  $\hat{Y}_i$  such that the variance of the replicate under the replication distribution become the derived unbiased variance estimator. After this, we followed the general bootstrap replication procedure.

## REFERENCES

Kovar, J.G., Rao, J.N.K. and Wu, C.F.J.(1988). Bootstrap and other methods to measure errors in survey estimates. *The Canadian J. of Stat.* , **16**, 25-45.

Rao, J.N.K. and Wu, C.F.J. (1988). Resampling inference with complex survey data. *J. the Amer. Stat. Asso.*, **83**, 231-241.

Vijayan, K., Mukhopadhyay, P. and Bhattacharyya, S. (1995). On non-negative unbiased estimation of quadratic forms in finite population sampling. *Austral. J. of Statist.* , **37**, 169-178.