Statistical Science
2011, Vol. 26, No. 2, 257–259
DOI: 10.1214/11-STS346A
Main article DOI: 10.1214/10-STS346
© Institute of Mathematical Statistics, 2011

Discussion of "Impact of Frequentist and Bayesian Methods on Survey Sampling Practice: A Selective Appraisal" by J. N. K. Rao

Glen Meeden

It has been pointed out that when apologists for competing systems like capitalism and socialism or the frequentist and Bayesian approaches to survey sampling argue about the relative merits of their systems, they often compare their ideal to the other's reality. Since the ideal is always quite different than reality it is easy for each of them to score points. I wish to thank Rao for avoiding this trap and giving a fair reading to both sides in his survey. Beyond that, I particularly liked the sections on the early development of frequentist methods.

How should prior information about the population be used in survey sampling? It can inform how the sample is selected and is used when making inferences after the data have been collected. Formally, at each of the two stages, the frequentist and Bayesian approaches are quite different but practically, I believe, they are often more alike than is commonly supposed.

THE FREQUENTIST APPROACH

In theory, for non-model-based frequentists, the sampling design is the most important place to use prior information. In Section 2 Rao described some of the early fundamental advances in survey design based on the frequentist approach. He explained why in stratified sampling and in stratified two-stage cluster sampling, where one cluster within each stratum is drawn, self-weighting of the units is a very desirable property. Such examples led to the notion of assigning a weight, which is the reciprocal of its inclusion probability, to each unit in the sample. A unit's weight is the number of units in the population that it represents. A theoretical justification for this notion that is often given is that under the sampling design the resulting estimator is unbiased. Rao argues, however, that large sample consistency of an estimator is a more important property than unbiasedness. Although it is hard to find sensible estimators which are badly biased, I agree with him that unbiasedness in and of itself is not an important property. Whatever justification there is for the notion of a weight, it should not be based on unbiasedness.

What I have sometimes found puzzling about weights is that after the sample has been selected they are often adjusted. Information that may not have been used at the design stage is used to make the sampled units and their weights more accurately reflect what is known about the population of interest. Calibration and the model-assisted approach are two common methods for achieving this end. An estimator based on the adjusted weights will no longer be design-unbiased, but there is theory to show that it can be design-consistent. Practice, however, can be more complicated especially when there are missing observations. But more importantly, the whole reweighting technology seems to me to mix up an unconditional argument (selecting the sampling design) and a conditional argument (using population information to get a good estimate after the sample has been observed). I am not suggesting that such adjustments should not be done, only that there can be more art than science in finding a good set of weights.

I believe that frequentists would be better served in their analysis if they more explicitly recognized these two different stages in the inferential process. In the first stage, one uses the information that is relevant to select the sampling design. In the second stage, after the sample has been selected, one should ignore the design but use *all* the information when constructing an estimator. In effect that is what one does when the sampling weights are adjusted. However, in the second stage, the design weights need not be used explictly as long as all the information is being taken into account.

Glen Meeden is Professor, School of Statistics, University of Minnesota, 224 Church St. SE, Minneapolis, Minnesota 55455-0493, USA (e-mail: glen@stat.umn.edu).