Statistical Science
2014, Vol. 29, No. 1, 95–96
DOI: 10.1214/13-STS448
Main article DOI: 10.1214/12-STS413

Discussion of "Estimating the Distribution of Dietary Consumption Patterns"

Stephen E. Fienberg and Rebecca C. Steorts

Carroll describes an innovative model developed in Zhang et al. (2011) for estimating dietary consumption patterns in children, and a successful Bayesian solution for inferring the features of the model. The original authors went to great lengths to achieve valid frequentist inference via a Bayesian analysis that simplified the computational complexities encountered in standard frequentist approaches. Pragmatically, this led to a reasonable set of estimates, but their combination of Bayesian and frequentist tools and ideas stopped short of what we consider a full and proper Bayesian analysis. We ask two fundamental questions: How do we know that the model and estimation are valid? What role should the survey weights have played?

1. MODEL VALIDITY

The model of Zhang et al. (2011) is highly complex—how, without something like sensitivity analysis, are we to know that it is valid? As for inference, the original authors rely on the well-known (Bernstein–von Mises) asymptotic convergence of Bayesian posterior means and maximum likelihood estimates to develop standard errors using balanced repeated replication (BRR). We agree that their sample size is large for many purposes, however, when the inverse Fisher information is large, convergence can be slow. Moreover, this standard convergence result is known to slow down as the number of parameters grows, failing completely for nonparametric models. Can we rely on Bernstein–von Mises, at these sample sizes, for this very complex (and only semi-parametric) model? This is not clear to us.

Stephen E. Fienberg is Maurice Falk University Professor of Statistics and Social Science, Department of Statistics, Machine Learning Department, and Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: fienberg@stat.cmu.edu). Rebecca C. Steorts is Visiting Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: beka@cmu.edu).

2. SURVEY WEIGHTS

In Section 3.3, Carroll (2014) notes that the use of survey weights in Bayesian analyses is controversial, and then he proceeds to use them as reported in by the National Center of Health Statistics (NCHS) nonetheless to do a weighted analysis. Fienberg (2009) reminds us that in the NCHS survey context, weights are not just used to adjust for unequal selection probabilities, but are the product of at least three factors:

$$w_k = \frac{1}{\pi_k} \times (\text{nonresponse adjustment}) \\ \times (\text{post-stratification adjustment}).$$

The first factor is the inverse of the probability of selection, for example, taking into account stratification and clustering. The second factor inflates the sample results to adjust for nonresponse, typically by invoking the assumption that the missing data are missing at random, at least within chosen strata or post-strata. The third factor re-weights the population totals to add up to control totals coming from another source such as a census.

Gelman (2007) rightly states: "Survey weighting is a mess," and this is especially so from a Bayesian perspective. What weights if any should be used in a Bayesian analysis? In a simple stratification setting, and where we are estimating a mean or a total, weighting using $1/\pi_k$ has a Bayesian justification. For more complex situations, such as the one Carroll describes, the role of the survey weights is unclear. Bayesian benchmarking is a way to deal with the third component in the weight formula above, but Ghosh and Steorts (2013) point out the tricky nature of the choice of both loss function and benchmarking weights for small area estimation of complex surveys. In essence, Carroll and his collaborators appear to be creating a pseudo-likelihood that adjusts individual contributions by the weights and then they use a survey-weighted MCMC calculation with uncertainty estimation coming from balanced repeated replication. This seems unusually strange to us, and decidedly non-Bayesian in character.

Even if this pseudo-likelihood structure is correct, to be *fully Bayesian*, the weight w_k associated with the kth child *should be* a random variable. The weights should then have a prior distribution and a likelihood, and be estimated together with the other unknown parameters. At the very least, ignoring the variability in the weights will cause the estimate of population distributions to seem unduly precise. It may be that a proper Bayesian weighting justification of Carroll (2014) exists, but simply hoping that the frequentist approach to survey weighting carries over to the Bayesian setting without change seems problematic.

ACKNOWLEDGMENT

Supported in part by NSF Grant SES-1130706 to Carnegie Mellon University.

REFERENCES

CARROLL, R. (2014). Estimating the distribution of dietary consumption patterns. *Statist. Sci.* **29** 2–8.

FIENBERG, S. (2009). The relevance or irrelevance of weights for confidentiality and statistical analyses. *Journal of Privacy and Confidentiality* **1** 183–195.

GELMAN, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci.* **22** 153–164. MR2408951

GHOSH, M. and STEORTS, R. C. (2013). Two-stage Bayesian benchmarking as applied to small area estimation. *TEST* **22** 670–687. MR3122328

ZHANG, S., MIDTHUNE, D., GUENTHER, P. M., KREBS-SMITH, S. M., KIPNIS, V., DODD, K. W., BUCKMAN, D. W., TOOZE, J. A., FREEDMAN, L. and CARROLL, R. J. (2011). A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *Ann. Appl. Stat.* **5** 1456–1487. MR2849782