PhD Dissertation Presentation

Bayesian Analyses of the Unrelated Question Design for Multiple Sensitive Questions

We use hierarchical Bayesian models to implement multiple sensitive questions into the unrelated question design for small areas (or clusters). Interest is focused on inference about finite population proportions of individuals with various sensitive characteristics for small areas.

Given binary response data from two or more sensitive questions from many small areas, we use a hierarchical Dirichlet-multinomial model to estimate the sensitive proportions. The computation is difficult, so a blocked Gibbs sampler is used to sample the joint posterior density and the posterior distributions of finite population proportions can be obtained. We apply our method to college cheating data to estimate the finite population proportions of students with different sensitive features for each of several classes. We also use a simulation study to validate our method, and we investigate the effects on posterior inference of increasing the number of areas and the correlation between the sensitive items.

When there are a large number of areas, our procedure is computationally intensive. Another problem is that the Dirichlet distribution models negative correlated probabilities and this is inflexible. To solve these problems, we propose a normal model after appropriately transforming the parameters. Then based on the new parameter setting, we are able to either use a full Gibbs sampler or an integrated nested normal approximation to make posterior inference about the finite population proportions of students cheating in different courses. This model has much fewer parameters, and therefore, there are gains in precision when the finite population proportions are estimated. It can also include covariates in a straightforward manner, when available.

Finally, we propose that the randomized response procedure can be used to provide masked public-used data, which is an important activity for many government agencies.