



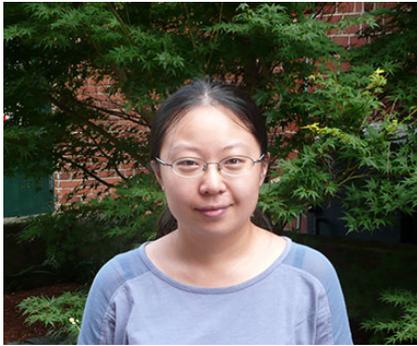
WPI

MATHEMATICAL SCIENCES

Yuan Yu

PhD Candidate

Mathematical Sciences



November 20, 2018

Fuller Labs, Room 311

9:30 – 11:30am

Dissertation Committee:

Professor Balgobin Nandram,
Advisor, WPI

Professor Jian Zou, WPI

Professor Joseph Sedransk, Joint
Program in Survey Methodology,
University of Maryland

Dr. Jai Won Choi, Statistical
Consultant, Meho Inc., Maryland

Professor Huong Higgins, WPI

PhD Dissertation Proposal Presentation

Title: Bayesian Analysis of Unrelated Question Design for Multiple Sensitive Questions from Small Areas

Abstract: Elicitation of answers from sensitive questions is a delicate issue, and even questions about basic demographics (e.g., age, race, sex) can be offensive or stigmatizing to some people. In sample surveys with sensitive questions, randomized response techniques have a huge advantage in estimating population quantities (e.g., proportion of people cheating on their tax returns) because they can reduce the bias caused by non-response or untruthful response, a measurement error.

The unrelated question design for estimating a single sensitive proportion is well studied and it is more efficient than Warner's original mirrored question design with the same sample size. Researchers have developed other techniques that are more efficient than the unrelated question design. However, these designs rely on large sample sizes to get admissible estimates and there are limited discussions about applications on data from small areas or clusters. Bayesian methods work well because they allow pooling of data from desperate areas and, moreover, they can utilize important prior information. On the other hand, few discussions have been made exploring the benefits of a combined design involving multiple items (e.g., two sensitive questions) under the Bayesian scheme. Therefore, in our study, given binary response data from two or more sensitive questions from many small areas, we use a hierarchical Bayesian Dirichlet-multinomial model with latent variables to estimate the sensitive proportions.

There are difficulties in running Markov chain Monte Carlo methods for this type of models in which latent variables must be introduced. However, a very clever blocked Gibbs sampler is used to get samples from the joint posterior density and the posterior distributions of finite population proportions can be obtained. We validate this procedure using surrogate data from the third National Health and Nutrition Examination Survey. We also provide a simulation study to investigate the effect of increasing the number of areas and the effect of increasing the correlation between the sensitive items. When there are a large number of areas, our procedure is computationally intensive.

Therefore, to make our procedure more useful, we propose using an integrated nested normal approximation to do the computation. It is expected that this new procedure will be much faster than the exact method and, moreover, it is as accurate as the exact method. We will explore these two scenarios.