



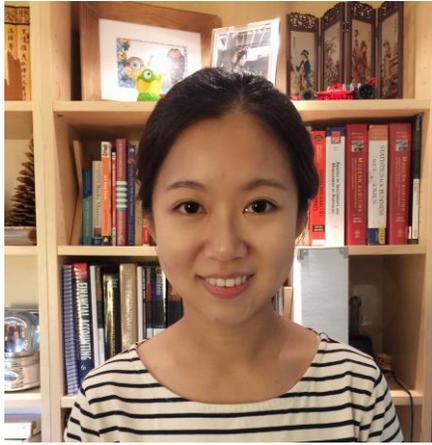
WPI

MATHEMATICAL SCIENCES

Lu Chen

PhD Candidate

Mathematical Sciences



Tuesday, November 20, 2018

Fuller Labs, Room 320

1:30 pm

Dissertation Committee

Professor Balgobin Nandram, Advisor,
WPI
Professor Jian Zou, WPI
Professor Andrew C. Trapp, WPI
Professor Joseph Sedransk,
Joint Program in Survey Methodology,
University of Maryland
Dr. Jai Won Choi, Statistical Consultant,
Meho Inc., Maryland

PhD Dissertation Defense

Hierarchical Bayesian Models for Polychotomous Data from Sub-Areas

Many population-based surveys have polychotomous categorical responses from a number of individuals in each household within small areas. An example is the second Nepal Living Standards Survey (NLSS II), in which health categorical data for each individual from sampled households (sub-areas) are available in sampled wards (small areas). To make inference about finite population proportions of individuals with different health statuses within small areas, we develop several hierarchical Bayesian models. There are three major aspects of this dissertation.

First, when the responses are binary, we consider a two-fold Beta-Binomial model without covariates and a sub-area logistic regression model with reliable auxiliary information. In this respect, there are two contributions. First, we extend an area model to a sub-area model. Second, because there are numerous sub-areas, it is time consuming to use standard Markov chain Monte Carlo (MCMC) methods to fit the joint posterior density. Therefore, we provide approximation methods that permit relatively much faster computations.

Second, when the survey responses are ordinal, we consider both area and sub-area hierarchical Bayesian probit models. A standard assumption is that the ordered categorical response is determined by an unobservable continuous variable. We discuss how to fit the model to avoid poor mixing problems in MCMC algorithm when simulating samples from the joint posterior distribution. This is a very common difficulty encountered in this type of computational problems.

Third, we discuss how to incorporate survey weights into all of our models. This is necessary because survey weights can increase variability when a correction is made for selection bias (e.g., nonresponse in the survey). Normalized composite likelihoods with survey weights are constructed for these models and a surrogate sampling approach is used in the models to predict the finite population proportions of small areas.

We use data in NLSS II to compare the sub-area and the area models for the three aspects of this dissertation. We show that the sub-area models are preferred over the area models that ignore the sub-areas within areas. Our sub-area models can borrow strength from both areas and sub-areas to obtain more efficient and precise estimates. The hierarchical structure of our models capture the variation in the NLSS II data reasonably well. A comparison of the weighted and unweighted models shows that covariates and survey weights might provide similar information. Our theoretical and methodological work can help provide small area official statistics for numerous surveys worldwide.