School of Mathematical and Statistical Sciences
Arizona State University

M E M O R A N D U M

DATE: 03/19/2024

TO: Faculty and Students

FROM: Professor(s) Chair/Co-Chairs of Defense for the PhD Committee Members

P. Richard Hahn
Palak Jain

in Applied Mathematics

Jingyu He
Ming-Hung Kao
Shiwei Lan
Shuang Zhou

DEFENSE ANNOUNCEMENT

Candidate: Palak Jain

Defense Date: Wednesday. April 10, 2024

Defense Time: 4:00 PM

Virtual Meeting Link: https://asu.zoom.us/j/84731788008 In Person: WXLR 304 (Tempe)

Title: Algorithms for Non- and Semi-parametric Bayesian Conditional Density Estimation on Large Datasets

Please share this information with colleagues and other students, especially those studying in similar fields. Faculty and students are encouraged to attend. The defending candidate will give a 40-minute talk, after which the committee members will ask questions. There may be time for questions from those in attendance. However, guests are primarily invited to attend as observers and will be excused when the committee begins its deliberations or if the committee wishes to question the candidate privately.

ABSTRACT
-See next page-

PO Box 871804 Tempe, AZ 85287-1804
(480) 965-3951 Fax: (480) 965-8119
http://math.asu.edu

ALGORITHMS FOR NON- AND SEMI-PARAMETRIC BAYESIAN CONDITIONAL DENSITY ESTIMATION ON LARGE DATASETS

ABSTRACT

This dissertation considers algorithmic techniques for non- or semi-parametric Bayesian estimation of density functions or conditional density functions. Specifically, computational methods are developed for performing Markov chain Monte Carlo (MCMC) simulation from posterior distributions over density functions when the dataset being analyzed is quite large, say, millions of observations on dozens or hundreds of covariates. The motivating scientific problem is the relationship between low birth weight and various factors such as maternal attributes and prenatal circumstances.

Low birth weight is a critical public health concern, contributing to infant mortality and childhood disabilities. The dissertation utilizes birth records to investigate the impact of maternal attributes and prenatal circumstances on birth weight through a statistical method called density regression. However, the challenges arise from the estimation of the density function, the presence of outliers, irregular structures in the data, and the complexity of selecting an appropriate method.

To address these challenges, the study employs a Bayesian Gaussian mixture model inspired by kernel density methods. Additionally, it develops a fast MCMC algorithm tailored to handle the computational demands of large datasets. A targeted sample selection procedure is introduced to overcome difficulties in analyzing weakly informative data.

To further enhance the study's approach to addressing challenges, a sophisticated clustering methodology is incorporated. The study leverages the creation of clusters based on different sizes, emphasizing the scalability and complexity of cluster formation within a dichotomous state space. Valid clusters, representing unique combinations of data points from distinct feature states, offer a granular understanding of patterns in the dataset.