THE GRADUATE COLLEGE OF THE
UNIVERSITY OF OKLAHOMA HEALTH SCIENCES CENTER

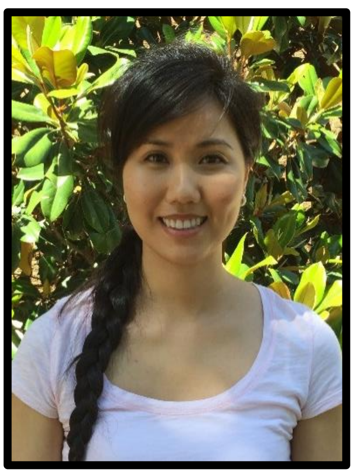
ANNOUNCES THE FINAL EXAMINATION OF

Ngoc Quyen Thi Duong

FOR THE DEFENSE OF THE DOCTOR OF PHILOSOPHY DEGREE
GRADUATE COLLEGE

Department of Biostatistics and Epidemiology

Thursday, August 15, 2019 | 9:00 am | Hudson College of Health Building, Room 204



Bayesian Posterior Kernel Density Estimation Using a Conjugate Gamma Prior

COMMITTEE IN CHARGE:
Michael P. Anderson, PhD; Sara K. Vesely, PhD
Ying Zhang, MD, PhD; Janis E. Campbell, PhD; Clara H. Song, MD

ABSTRACT: Due to advancements in technology, a larger number of predictors can be obtained for a subject in a short period of time. Such data are known as high dimensional data and they pose unique challenges for traditional statistical analysis. The application of traditional methods to these data are not optimal because they either fail to converge and do not produce estimates, or they require substantial dimension reduction techniques that result in discarding large portions of the observed data which in turn leads to a loss of information. The methodology proposed in this dissertation uses all of the available data to produce posterior probabilities that can be used to classify subjects to groups. We employ a Bayesian classification and prediction model to predict the risk of a given outcome using high dimensional longitudinal data. Kernel density estimates are computed and augmented with a Gamma prior distribution to account for the correlation among predictors. Classification performance was assessed by calculating sensitivity and specificity at various time points, and simulations were used to measure bias and mean squared error in the posterior probability calculations. Gamma posterior kernel density estimate (GPKDE) is compared with the normal posterior kernel density estimate (NPKDE). When samples are drawn from a normal distribution or a truncated normal distribution, some of the trends show that for both the GPKDE and NPKDE methods as mean differences increased, sensitivities and specificities increased. As correlation level increases, sensitivities and specificities also increased. As time point increased, over time it gathers more information so sensitivities and specificities also increased. Results for a hundred simulations is the same as a thousand simulations when data are high-dimensional. The proposed methodology performs well in terms of small bias and high sensitivity and specificity when data are bounded below by zero and right-skewed.