The Department of Educational Psychology's Research Methods, Measurement, & Evaluation (RMME) program and the Department of Statistics at the University of Connecticut present:

Bayesian Pyramids: Identifying Interpretable Deep Structure Underlying High-dimensional Data

Dr. David Dunson, Duke University

Friday, 3/26/2021, 12pm

https://uconn-cmr.webex.com/uconn-cmr/j.php?MTID=m09a58d2d0b8f3973e89583e46454fbfa

High-dimensional categorical data are routinely collected in biomedical and social sciences. It is of great importance to build interpretable models that perform dimension reduction and uncover meaningful latent structures from such discrete data. Identifiability is a fundamental requirement for valid modeling and inference in such scenarios yet is challenging to address when there are complex latent structures. We propose a class of interpretable discrete latent structure models for discrete data and develop a general identifiability theory. Our theory is applicable to various types of latent structures, ranging from a single latent variable to deep layers of latent variables organized in a sparse graph (termed a Bayesian pyramid). The proposed identifiability conditions can ensure Bayesian posterior consistency under suitable priors. As an illustration, we consider the two-latent-layer model and propose a Bayesian shrinkage estimation approach. Simulation results for this model corroborate identifiability and estimability of the model parameters. Applications of the methodology to DNA nucleotide sequence data uncover discrete latent features that are both interpretable and highly predictive of sequence types. The proposed framework provides a recipe for interpretable unsupervised learning of discrete data and can be a useful alternative to popular machine learning methods.

Dr. David Dunson is Arts & Sciences Distinguished Professor of Statistical Science and Mathematics at Duke University. His research focuses on developing methodology for analysis and interpretation of complex and high-dimensional data, with a particular emphasis on biomedical applications, Bayesian statistics, and probability modeling approaches. Methods development and theory is directly motivated by challenging applications in neuroscience, genomics, environmental health, and ecology, among others. Dr. Dunson received his BS in Mathematics from the Pennsylvania State University in 1994, and his PhD in Biostatistics from Emory University in 1997. He then spent a decade at the National Institute of Environmental Health Sciences before moving to Duke. His work has had a substantial impact, with ~55,000 citations on Google Scholar and an H-index of 80.

ONLINE INTERDISCIPLINARY SEMINARS ON STATISTICAL METHODOLOGY FOR SOCIAL AND BEHAVIORAL RESEARCH: Support for this seminar comes from Department of Educational Psychology's Research Methods, Measurement, & Evaluation (RMME) program and the Department of Statistics at the University of Connecticut (UCONN), the Statistical and Applied Mathematical Sciences Institute (SAMSI), and the New England Statistical Society (NESS). This seminar aims to promote connection between the statistics and social/behavioral science communities and encourage interdisciplinary research across faculty and students.

For announcements and WebEx live streaming links, please contact Tracy Burke (tracy.burke@uconn.edu). For questions related to the seminars, please feel free to contact the session organizers, Prof. Xiaojing Wang (xiaojing.wang@uconn.edu) and/or Prof. Betsy McCoach (betsy.mccoach@uconn.edu).

Additional Session Meeting Information:

Join by video system: Dial 1208212088@uconn-cmr.webex.com
You can also dial 173.243.2.68 and enter your meeting number.

Meeting # 120 821 2088
Password: RMMESTAT

Join by phone: +1-415-655-0002 US Toll
Access code: 120 821 2088