Saw Swee Hock Public Lecture

Bayesian Pyramids: Identifying Interpretable Structure Underlying High-dimensional Data

by Professor David B. Dunson

Arts and Sciences Distinguished Professor of Statistical Science and Mathematics
Duke University

July 6, 2022 (Wednesday)
8:00am - 9:00am

By Zoom: **https://hku.zoom.us/j/95103015302**
Meeting ID: 951 0301 5302

About the Talk

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.
Joint work with Yuqi Gu.

About the Speaker

David Dunson is Arts and Sciences Distinguished Professor of Statistical Science and Mathematics at Duke University. His research focuses on Bayesian statistical theory and methods motivated by high-dimensional and complex applications. This includes dimensionality reduction, scalable inference algorithms, latent factor models, and nonparametric approaches, particularly for high-dimensional, dynamic and multimodal data, including images, functions, shapes and other complex objects. Primary areas of application include neurosciences and brain network modeling, environmental health, ecology, and human fertility. Professor Dunson is a fellow of the American Statistical Association, the Institute of Mathematical Statistics and the International Society of Bayesian Analysis. He has won numerous awards, including the Mortimer Spiegelman Award in 2007, and most notably the 2010 COPSS President's Award. He also received the 2010 Myrto Lefkopoulou Distinguished Lectureship from Harvard University. Professor Duunson served as joint Editor of the Journal of the Royal Statistical Society, Series B.

Department of Statistics & Actuarial Science
The University of Hong Kong 統計及精算學系

All interested are welcome

ENQUIRY: 3917 8311, saas@hku.hk