Abstract:

The first part of talk deals with some models designed to cluster ordinal or mixed-type data (continuous and ordinal) through a finite mixture model that is only partially observed. Since the corresponding likelihoods involve multidimensional integrals, the full maximum likelihood estimation becomes computationally infeasible as the number of ordinal variables increases. To solve the problem, the model parameter estimation is based on a composite likelihood approach. In the second part, a multivariate hidden Markov model (HMM) for mixed-type variables is introduced. As some of the considered variables may not contribute to the clustering structure, a hidden Markov-based model is built such that discriminative and noise dimensions can be recognized. The variables are considered to be linear combinations of two independent sets of latent factors where one contains the information about the cluster structure, following an HMM, and the other one contains noise dimensions distributed as a multivariate normal (and it does not change over time). The resulting model is parsimonious, but its computational burden may be cumbersome. To overcome any computational issue, a composite likelihood approach is introduced to estimate model parameters. The model is applied to a real dataset derived from the first five waves of the Chinese Longitudinal Healthy Longevity Survey. The model is able to identify the discriminant variables and capture the cluster structure changing over time parsimoniously. In the last part of the talk, a broad overview on some other research themes is given: applications of the composite likelihood to spatial and demographical data and some other clustering models.