

A Framework for Detecting Structural Heterogeneity in Latent Variable Models

Gabriel Wallin

School of Mathematical Sciences

Lancaster University

g.wallin@lancaster.ac.uk

Latent variable models are widely used in the social, behavioural, and health sciences to learn the latent structure underlying multivariate data. These models typically assume that the relationship between the set of latent variables and observed variables is identical for all measurement units. In practice, subpopulations may exist where the conditional distribution of a subset of observed variables given the latent variables differs systematically. Detecting such heterogeneity is challenging when both the subpopulations and the affected variables are unknown a priori. To address this problem, this talk presents a hybrid model that probabilistically assigns observations to discrete latent classes, where within each class, a continuous latent variable governs the observed variables. For each class, we estimate class-specific intercept and slope parameters that may deviate from a common baseline. We propose a regularised marginal likelihood estimator that enforces sparsity of these deviations, enabling simultaneous identification of latent classes and selection of heterogeneous variables via a proximal-gradient-based EM algorithm. The approach is illustrated using data from both a personality assessment and a large-scale educational test, where we identify groups that differ on specific variables beyond what is explained by the latent variable. Such patterns have important implications for the validity of these instruments. Connections to recent work on change-point analysis for latent variable models highlight a broader framework for detecting structural breaks in latent processes. This is joint work with Qi Huang (Purdue University).