psd

DISSERTATION PRESENTATION AND DEFENSE

Variance Modeling in Several Settings

WHEN
April 27, 2022
10:00AM

WHERE
Jones Laboratory, Room 304

Nathan Gill, PhD candidate

As a result of advances in data collection technology and study design, modern longitudinal datasets can be much larger than they historically have been. So called "intensive" longitudinal datasets are rich enough to allow for detailed modeling of the variance of a response as well as the mean. The mixed-effects location-scale (MELS) regression model is now a common technique for analyzing intensive longitudinal data. The model allows fixed effects and subject-level random effects to influence the mean of a response as well as the variation around that mean. A limitation of the MELS model is that it is difficult to fit computationally. No closed form expression exists for the likelihood, and Gaussian quadrature is the current technique used to fit the model. This procedure can be time intensive, and in addition to being inconvenient for routine data analysis and model building, also makes resampling methods like the bootstrap infeasible.

To address these challenges, we introduce a new fitting technique for the MELS model, called FastRegLS, that is a great deal faster and more numerically stable than quadrature-based maximum likelihood, while still providing consistent parameter estimates. Because it is so much faster, FastRegLS has the added benefit of enabling bootstrap inference for the MELS model parameters.

Next, we extend the MELS model to allow for within-subject correlation structure. We focus on the AR(1) structure, but we take a two-stage estimation approach that allows other correlation structures to easily be substituted. Like FastRegLS, our estimation procedure produces statistically valid estimators without incurring unreasonable computational costs. A key aim of the approach is to estimate subject-specific autocorrelation parameters (treated as random effects) that can then be used, along with the best linear unbiased predictors from the MELS model, to group study subjects by common characteristics of their response profiles.

In the final section, we switch our focus from longitudinal data and MELS models to construction of Gaussian graphical models. We propose a scoring method for interpreting GGMs at the level of individual network paths based on the relative importance of such paths in determining the Pearson correlation between their terminal nodes. The method is validated using human metabolomics data, with

stat.uchicago.edu

THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS