Generalized Linear Models for the Covariance Matrix of Longitudinal Data

Mohsen Pourahmadi

Department of Statistics

Texas A&M University

Abstract: Finding an *unconstrained* and *statistically interpretable* reparameterization of a general covariance matrix is still an open problem in statistics. Its solution is crucial for parsimonious and sparse modeling, and guaranteeing the positive-definiteness of an estimated covariance matrix in all areas of statistics dealing with correlated data, including the longitudinal (panel, functional, spectroscopic, repeated measure, ...) data. It is known that some estimated covariance matrices are not necessarily positive-definite, and recently due to popularity of generalized estimating equations (GEE) and SAS PROC Mixed, there has been a growing tendency to pick a covariance matrix for a data set from a long and expanding menu of covariance matrices, a task which is difficult even for the experts. In this presentation, pooling together ideas from regression and time series analysis, I will discuss a data-based, general-purpose method extending the framework of generalized linear models (GLMs) to covariance matrices, where a link function is introduced through the Cholesky decomposition. It reduces the difficult and unintuitive task of modeling a covariance matrix to that of modeling a sequence of (auto) regressions. Therefore, all existing regression machineries and approaches such as parametric, semiparametric, nonparametric, Bayesian, shrinkage (Ridge, Lasso, ...), etc. can be brought to the service of modeling covariances.