Estimation of Networks and Graphical Structures

Mikko J. Sillanpää

Research Unit of Mathematical Sciences
University of Oulu
mikko.sillanpaa@oulu.fi

In Gaussian graphical models, assuming a multivariate normal distribution with an unknown covariance matrix is a starting point for network estimation. An undirected network is typically constructed from the estimated precision matrix (inverse of the covariance matrix) whose elements exhibit conditional independences and dependences between variables. This is, non-diagonal elements of precision matrix are zero (and non-zero) at the same time as corresponding partial correlation coefficient is zero (and non-zero), respectively. This means in practice, if the network is estimated based on the covariance matrix rather than the precision matrix, one may obtain a larger number of connections that result from indirect associations than direct ones. However, this is not to say that only causative associations are included in the network constructed based on the precision matrix.

In biological applications, the number of variables (nodes) is usually much higher than the number of samples, which necessitates the use of different regularization methods, which may, e.g., assume sparsity - that many of the precision matrix elements will be zero. A small tutorial on the topic is provided, and some of the Bayesian and frequentist methods are shortly mentioned and discussed.