
Inférence de réseaux parcimonieux en présence de variables inobservées

Geneviève Robin^{*1}, Christophe Ambroise², and Stéphane Robin¹

¹Mathématiques et Informatique Appliquées (MIA) – Institut national de la recherche agronomique (INRA) : UMR518 – France

²Laboratoire de Mathématiques et Modélisation d'Evry (LAMME) – CNRS : UMR8071 – France

Résumé

In many biological and ecological case studies, the empirical covariance matrix of the variables of interest displays large blocks of uniform correlation. This suggests the existence of one or several unobserved (missing) variables having a simultaneous influence on a series of observed ones, and that we observe a sample drawn from a distribution where the unobserved variables have been marginalized out. The inference of underlying networks is compromised in this context because marginalizing variables yields locally dense structures that challenge the generally accepted assumption that biological networks are sparse. We present a procedure for inferring Gaussian graphical models from an independent sample in the presence of unobserved variables. Our model is based on spanning trees and the EM algorithm and accounts both for the influence of unobserved variables and the low density of the network. We treat the graph structure and the unobserved nodes as latent variables and compute posterior probabilities of edge appearance. We also compare our method to existing graph inference techniques on synthetic and flow cytometry data.

^{*}Intervenant