Penalized maximum likelihood estimation of a sparse DAG

Sara van de Geer
Seminar for Statistics, ETH Zürich

Joint work with Peter Bühlmann

Let X be an $n \times p$ matrix of observations. A directed acyclic graph (DAG) models the observations as

$$X = XB_0 + E,$$

where each row of E is $\mathcal{N}(0, \Omega_0)$-distributed, with Ω_0 a diagonal matrix. Moreover, writing the columns of X as X_k, $k = 1, \ldots, p$, and those of E as ϵ_j, $j = 1, \ldots, p$, it is assumed that ϵ_j and X_k are independent whenever X_k is a parent of X_j. There are several ways to represent the DAG (B_0, Ω_0). We consider one with the minimal number of edges. We estimate the DAG using maximum likelihood with a penalty proportional to the number of edges. We assume that any representation of the DAG has at least a given proportion of its non-zero coefficients above the noise level, and that the number of edges per node is sufficiently smaller than $n/\log p$. We prove convergence in Frobenius norm of the penalized maximum likelihood estimator, and show that it has about the same number of edges as the true DAG.