Graphical Modeling for Discrete Random Variables with Application to Tissue Microarray (TMA) Experiments

Corinne Dahinden, dahinden@stat.math.ethz.ch

Seminar für Statistik, ETH Zürich and Competence Center for Systems Physiology and Metabolic Diseases, University/ETH Zürich

Keywords: Graphical models, Hierarchical models, Interactions, Lasso, Log-linear models, Variable selection

AMS: 68R01, 62P10

Abstract

Tissue microarrays (TMA) are composed of hundreds of tissue sections from different patients arrayed on a single glass slide. With the use of immunohistochemical staining, they provide a high-throughput method of analyzing potential biomarkers on large patient samples. The assessment of the expression level of a biomarker is usually performed by the pathologist on a categorical scale.

The analysis of the interaction of these biomarkers and in particular the estimation of the graph structure associated with the underlying discrete random variables, are of biological importance. Questions such as how the interaction pattern changes with progressing tumor grade or with survival time are of direct biological interest. However, the estimation of the interaction structure requires sophisticated techniques. Our approach is to fit an ℓ_1-regularized log-linear model assuming a multinomial sampling scheme in order to obtain the graphical model. The regularization becomes necessary as after cross-tabulation of the samples in contingency tables, many cell entries remain zero, leading to so-called sparse contingency tables, where standard procedures fail to work.

We compare our approach with other methods for graphical modeling. Moreover, biological validation of the estimated interaction structure is done by mapping to known biochemical pathways.