A shrinkage entropy approach for inferring high-dimensional association networks

Korbinian Strimmer[1]

[1] Institute of Medical Statistics, Statistics and Epidemiology, University of Leipzig, Germany

Association networks are a popular tool in genomics to uncover putative relationships among genes. In order to allow analysis of high-dimensional data sets diverse methodology for regularized inference of corresponding large-scale graphical models has been developed, see for instance [1, 2, 3, 4]. A drawback of these approaches is they all rely on partial correlation as measure of association, which in turn implies that only linear relationships among the considered variables are considered.

In my talk I discuss a generalization of the network inference procedure of [4]. Specifically, I present a shrinkage estimator for entropy and related quantities, such as mutual information. This estimator can be applied in the "small n, large p" domain, and allows to assess the conditional independence even for nonlinear relationships. Thus, the shrinkage entropy estimator offers the basis for the statistical learning of proper large-scale association networks. In my talk I discuss properties of this approach and illustrate entropy network learning by analyzing high-dimensional genomic data.

References

[1] Drton M and Perlman MD (2007) Multiple testing and error control in Gaussian graphical model selection. Stat. Sci. 22:430-449

[2] Jones B, Dobra A, Carvalho C, Hans C, Carter C and West M (2005) Experiments in stochastic computation for high-dimensional graphical models. Stat. Sci. 20: 388-400.

[3] Opgen-Rhein R and Strimmer K (2007) From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Syst. Biol. 1:37.

[4] Schäfer J and Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statist. Appl. Genet. Mol. Biol. 4:32.