

Bootstrap Inference for Network Construction

Shuang Li

Fred Hutchinson Cancer Research Center

Biostatistics & Biomathematics

Location: MH 107

Thursday, February 23, 4:00pm-5:00pm

Abstract:

Network structure learning based on Gaussian graphical modeling (GGM) is essentially a model selection problem. Regularized algorithms are used to accommodate high-dimension scenarios, whereas there is a difficulty in finding the right amount of regularization. Traditional techniques for regularization tuning, such as BIC and cross validation, are prediction-criterion-based and do not fit well for this unsupervised case. We propose a method (BINCO) to choose models by directly estimating the false discovery rates (FDRs). It uses model aggregation (i.e., repeatedly estimate the network based on perturbed data) and builds a mixture model for the selection frequencies of candidate edges in the network. With a reasonable fitting for the mixture model under certain amount of regularization, it gives good estimates of FDRs for models at each frequency cutoff, and then the best model is transparent as the largest set of edges across all frequency cutoffs under a range of regularization with its FDR not exceeding a controlled level. BINCO is very general as it only depends on the selection frequencies and hence has a wide range of applications. We show the performance of BINCO on both simulated and real data using the *space* algorithm.