Mathematics, Statistics, and Computer Science   **@ UIC**

Special Colloquium

Statistical Inference in Large Discrete Graphical Models via Quadratic Programming

Zhao Ren (University of Pittsburgh)

Abstract: The high dimensional graphical model, a powerful tool for studying conditional dependency relationship of random variables, has attracted great attention especially in biological network analysis with different types of omics data. While significant progress has been achieved recently in computing confidence intervals and p-values of each edge for Gaussian graphical model (GGM), the usage of GGM on important discrete-type omics data can be statistically and biologically inappropriate. On the other hand, discrete-type graphical models were proposed to tailor the network analysis of count-valued data, but the statistical inference of these models is not well studied. In this talk, we investigate statistical inference of each edge for large Ising and modified Poisson-type graphical models. The key role in most existing inferential methods is played by a linear projection method to de-bias an initial regularized estimator. Major drawback of this approach in those discrete-type graphical models is that an extra sparsity assumption on the linear projection coefficient is required, which cannot be checked in practice. In addition, efficiency often is compromised by the usage of sample splitting in these methods. To solve these challenges, we first propose a novel estimator of each edge for Ising model via quadratic programming and show that our estimator is asymptotically normal without the above mentioned extra sparsity condition. Our proof applies a novel low dimensional maximum likelihood method for the de-bias procedure and a data swap technique to avoid loss of efficiency. In addition, we further show that whenever the extra sparsity condition is satisfied, our estimator is adaptively

Monday, January 27 at 3:00 PM in 636 SEO

efficient and achieves the Fisher information. Otherwise, we still provide a restricted Fisher information as a lower bound. We then extend our approach to modified Poisson-type graphical models for both edge-wise and global statistical inference. The practical merit of the proposed method is demonstrated by an application to a novel RNA-seq gene expression data set in childhood atopic asthma in Puerto Ricans. Compared to sole estimation and statistical inference of GGM, our method provides more biologically meaningful results.

Monday, January 27 at 3:00 PM in 636 SEO