Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

ERIC JANOFSKY

Department of Statistics
The University of Chicago

Exponential Series Approaches for Nonparametric Graphical Models

TUESDAY, April 28, 2015, at 11:00 AM

Eckhart 117, 5734 S. University Avenue

ABSTRACT

Markov Random Fields (MRFs) or undirected graphical models are parsimonious representations of joint probability distributions. Variables correspond to nodes of a graph, with edges between nodes corresponding to conditional dependencies. For a pairwise MRF, the joint density factorizes as a product over edges of the graph. This thesis studies high-dimensional, continuous-valued pairwise MRFs. We are particularly interested in approximating pairwise densities whose logarithm belongs to a Sobolev space. For this problem we propose the method of exponential series [Crain, 1974; Barron and Sheu, 1991], which approximates the log density by a finite-dimensional exponential family with the number of sufficient statistics increasing with the sample size.

We consider two approaches to estimating these models. The first is regularized maximum likelihood. This involves optimizing the sum of the log-likelihood of the data and a sparsity-inducing regularizer. We provide consistency and edge selection guarantees for this method. We then propose a variational approximation to the likelihood based on tree- reweighted, nonparametric message passing. This approximation allows for upper bounds on risk estimates, leverages parallelization and is scalable to densities on hundreds of nodes. We show how the regularized variational MLE may be estimated using a proximal gradient algorithm. We demonstrate our method's efficacy in density estimation and model selection in comparison to other approaches in the literature using simulated data and MEG signal data.

We then consider estimation using regularized score matching. This approach uses an alternative scoring rule to the log-likelihood, which obviates the need to compute the normalizing constant of the distribution. For general continuous-valued exponential families, we provide parameter and edge consistency results. As a special case we detail a new approach to sparse precision matrix estimation which has statistical performance competitive with the graphical lasso [Yuan and Lin, 2007] and computational performance competitive with the state-of-the-art glasso algorithm [Friedman et al., 2008]. We then describe results for model selection in the nonparametric pairwise model using exponential series. The regularized score matching problem is shown to be a convex program; we provide scalable algorithms based on consensus Alternating Direction Method of Multipliers (ADMM, [Boyd et al., 2011]) and Coordinate-wise Descent. We compare our method to others in the literature as well as the aforementioned TRW estimator using simulated data.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: https://lists.uchicago.edu/web/arc/statseminars.