Estimation: Approximated Maximum Likelihood

- Maximum likelihood would be a typical approach but it entails intractable \( \mathcal{L}(\Theta) \).
- Instead, replace \( \mathcal{L}(\Theta) \) with a tractable upperbound \( \tilde{\mathcal{L}}(\Theta) \):

\[
\tilde{\mathcal{L}}(\Theta) = \log \mathbb{E}_{\mathcal{X}} \left[ \prod_{x \in \mathcal{X}} B_x(x) \right] = \sum_{x \in \mathcal{X}} \log \mathbb{E}_{\mathcal{X}} \left[ B_x(x) \right] = \sum_{x \in \mathcal{X}} \log \frac{1}{\theta_x} = \sum_{x \in \mathcal{X}} \log \theta_x.
\]

- Here, \( \mathcal{X} \) is the averaged sufficient statistic \( B_x(x) / \mathcal{Z} \) over the \( n \) samples. And the group lasso \( \tilde{\mathcal{L}}(\Theta) = \sum_{x \in \mathcal{X}} | \theta_x |^{1/2} \) encourages the \( x \)-th block, for every \( x \neq y \), to be a zero matrix.
- Derive a convex upperbound \( \mathcal{L}_r(\Theta) \) (Sketch: Express \( \mathcal{L}(\Theta) \) as an entropy \( H(\mathcal{X}) \), simplify it into \( H(\mathcal{X}) \), and use the fact that Gaussian distribution gives maximum entropy among any pdf under the same covariance matrix.)

\[
\text{Theorem. For a PE-MRF, the approximated negative maximum log-likelihood problem becomes}
\]

\[
\min_{\theta \in \mathbb{R}^k} \mathbb{E}_{\mathcal{X}} \left[ \left( \theta \right)^T \mathcal{X} \right] - \log \text{det} \mathbb{E}_{\mathcal{X}} \left[ \mathcal{X} \right] + \lambda \sum_{k=1}^p \sum_{j=1}^{\max_{i \in I} |I_j|} \theta_{jk}.
\]

- \( \mathcal{X} \) is the natural parameter with node-parameter \( \eta \in \mathbb{R}^n \) and edge-parameter \( \tilde{\eta}_{ij} \in \mathbb{R}^n \).
- The log-partition function \( \mathcal{L}(\Theta) \) should be found.

- Property.
- PE-MRF is exponential family where sufficient statistics and natural parameter interact linearly and nodes have a pairwise relationship.
- Include well-known distributions such as GMRF, Ising model, discrete model.
- The \( \{\tilde{\eta}_{ij}^{\theta_{jk}}\} \) explicitly reveal underlying Markov structure.

Experiments

- **Synthetic Data.**
  - Experiment on (synthetic) heterogeneous network with 32 nodes: 8 Bernoulli, 8 gamma, 8 Gaussian, and 8 three-dimensional Dirichlet.
  - Plot the ROC curves for edge recovery percentage, compared with VS-MRF.

- **For inferring the 100 node Markov network, our PE-MRF solver can takes under 30 seconds (whereas VS-MRF takes over 31 hours).**

- **Heterogeneous Genomic Networks.**
  - Level III public data from the Cancer Genome Atlas for 290 breast cancer patients.
  - Contains expression profiles for 500 genes (Gaussian) and 314 miRNAs (Poisson).
  - Consider a 5-core of the gene-gene subnetwork to demonstrate its utility.
  - For validation, observe how many gold standard edges exist between 0/1 core genes from external data.

Conclusions

- Improve a PE-MRF model, a subclass of the exponential family that is well-suited for heterogeneous multivariate distributions.
- Formulate an approximated maximum likelihood problem by deriving upper bound on the log partition function.
- Develop an ADMM algorithm with closed-form updates.
- The estimator guarantees to recover underlying Markov structure consistently.
- Our results, as well as the widespread applications with heterogeneous data sources, lead to many extensions of this work. For example, instead of inferring a single PE-MRF network, we could use the time-stamped observations to estimate a time-varying network because it is possible that Markov structure changes over time.


References.


