Mixed Graphical Models

Landon Buechner

STAT 616 - Dr. Bhattacharya

Texas A&M University

2020

Outline

- Markov Random Fields
- The Exponential Family
- Mixed Graphical Models
- Learning the Structure of Mixed Graphs
- Simulation Study
- Real World Applications

Historical Background

Define $G = (V, E)$ as the undirected graph over $X = (X_1, \ldots, X_p)$ with node set $V = \{1, \ldots, p\}$ and edge set $E = V \times V$. Note that X is a p-dimensional random vector with each variable taking values in a set $\mathcal{X}_r$.

Besag (1974) showed that a joint distribution can be fully specified in terms of node-neighborhood conditional distributions.

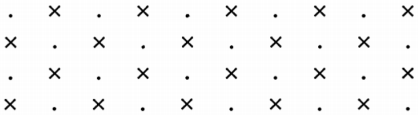


FIG. 1. Coding pattern for a first-order scheme.

$$\prod p_{i,j}(x_{i,j};\, x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}),$$

Exponential MRF's

Theorem (Besag 1974 + Hammersley-Clifford + Yang 2014)

If X has marginal distributions arising from the exponential family and satisfies the Markov independence property then the joint distribution can factorized over the set maximal cliques $\mathcal{C}$ of the graph G.

Let $\phi_c(X_c)$ be a sufficient statistic dependent on the X_c that correspond to the clique $c \in \mathcal{C}$, then the Markov Random Field (MRF) family of distributions has the form:

$$P(X) \propto \exp\left\{ \sum_{c \in \mathcal{C}} \phi_c(X_c) \right\}$$

The Exponential Family

Denote a random variable Z with distribution that is a member of the exponential family as $Z \in \mathcal{F}$. These distributions have the form

$$P(Z) = \exp\left\{\eta^T B(Z) + C(Z) - D(\eta)\right\}$$

with canonical parameter η, sufficient statistics $B(Z)$, base measure $C(Z)$, and log-partition function $D(\eta)$.

- What is the natural parameter space $\mathcal{N} = \{\eta : D(\eta) < \infty\}$?
- Note that in general given a set Θ, we can write $\eta = \eta(\theta)$ where $\eta : \Theta \to \mathcal{N}$ is a bijection.

Mixed Exponential MRF's

Graphical Models via Univariate Exponential Family Distributions (Yang et al. 2012)

For node $r \in V$, consider the distribution of X_r conditioned on the remaining $p-1$ random variables $X_{V\setminus r}$. Note that $\eta_r \equiv \eta_r(X_{V\setminus r}, \theta)$.

$$P(X_r | X_{V\setminus r}) = \exp\left\{\eta_r B_r(X_r) + C_r(X_r) - D_r(\eta_r)\right\}$$

Mixed Graphical Models via Exponential Families (Yang et al. 2014)

- What if each conditional distribution came from a unique distribution in $\mathcal{F}$?
- What restrictions would this impose on $\mathcal{N}$?

Theorem (Mixed Exponential MRF - Yang et al. 2014)

The collection of $P(X_r|X_{V\setminus r}) \in \mathcal{F}$ form a pairwise MRF over the random vector X that is Markov with respect to the graph $G = (V, E)$ if and only if the functions $\{\eta_r(\cdot)\}_{r\in V}$ that specify the conditional distributions have the form:

$$\theta_r + \sum_{(r,t)\in E} \theta_{rt} B_t(X_t)$$

With corresponding joint distribution

$$P(X) = \exp\left\{ \sum_{r\in V} \theta B_r(X_r) + \sum_{(r,t)\in E} \theta_{rt} B_r(X_r) B_t(X_t) + \sum_{r\in V} C_r(X_r) - A(\theta) \right\}$$

with log-partition function

$$A(\theta) = \log \int_X \exp\left\{ \sum_{r\in V} \theta_r B_r(X_r) + \sum_{(r,t)\in E} \theta_{rt} B_r(X_r) B_t(X_t) + \sum_{r\in V} C_r(X_r) \right\} dX$$

Mixed Exponential MRF - Consider the possibilities!

Here are some examples, assuming the data is properly transformed

1. Gaussian-Multinomial:
    - Flight delays and weather condition (categorical)
2. Poisson-Bernoulli:
    - Gene expression and mutation state (binary)
3. Exponential-Bernoulli
    - Application use duration and membership status (binary)

Gaussian-Potts Model

Let $\{ Y_r\}$ be a set of univariate Gaussian random variables with domain $\mathcal{Y} = \mathbb{R}$ and known variance σ_r^2. Additionally, let $\{Z_r\}$ be a set of multinomial random variables taking values in the finite set $\mathcal{Z}_r = \{0, 1, 2, \ldots\}$.

$\mathcal{F}$	$B(\cdot)$	$C(\cdot)$
Gaussian	$\frac{Y_r}{\sigma_r}$	$-\frac{Y_r^2}{2\sigma_r^2}$
Multinomial	$\mathbb{I}[\, Z_r = j\,]$	0

Let (V_Y, E_Y), (V_Z, E_Z), and (V_{YZ}, E_{YZ}) be the two homogeneous sub-graphs and the heterogeneous sub-graph respectively.

Marginally we have a Gaussian graphical model and a Potts model!

Gaussian-Potts Model

Prepare your eyes!

Gaussian-Potts Model - A "*Manichean*" MRF

Parameters:

- $\theta \leftrightarrow$ Multinomial node potential
- $\alpha \leftrightarrow$ Gaussian node potential
- $\lambda \leftrightarrow$ Cross type potential

Identity Functions:

- $\mathcal{I}^{j}_{r'} = \mathbb{I}[\, Z_{r'} = j\,]$
- $\mathcal{I}^{jk}_{(r',t')} = \mathbb{I}[\, Z_{r'} = j\,,\; Z_{t'} = k\,]$

Joint density:

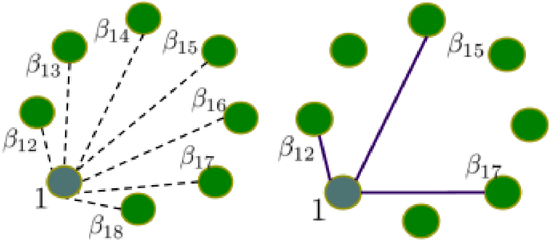
$$P(Y, Z) \propto \exp\left\{ \sum_{\substack{r' \in V_Z \\ j \in \mathcal{Z}}} \theta^{j}_{r'}\, \mathcal{I}^{j}_{r'} + \sum_{\substack{(r',t') \in E_Z \\ j,k \in \mathcal{Z}}} \theta^{jk}_{r't'}\, \mathcal{I}^{jk}_{(r,r')} + \sum_{\substack{(r,r') \in E_{YZ} \\ j \in \mathcal{Z}}} \lambda_{rr'}\, Y_r\, \mathcal{I}^{j}_{r'} \right.$$
$$\left. + \sum_{r \in V_Y} \alpha_r Y_r + \sum_{(r,t) \in E_Y} \frac{\alpha_{rt}}{\sigma_r \sigma_t} Y_r Y_t - \sum_{r \in V_r} \frac{Y_r^2}{2\sigma_r^2} \right\}$$

Parameter Estimation For Homogeneous MRF's

1. Gaussian Graphical Models
    - Node-neighborhood (Meinshausen and Bühlmann 2006)
    - Graphical Lasso (Hastie and Tibshirani 2008), (Friedman et al. 2008)
2. Ising/Potts Models
    - Node-neighborhood (Ravikumar et al. 2010)
    - Group Penalized Lasso (Jalali et al. 2011)

Parameter Estimation: Intuition

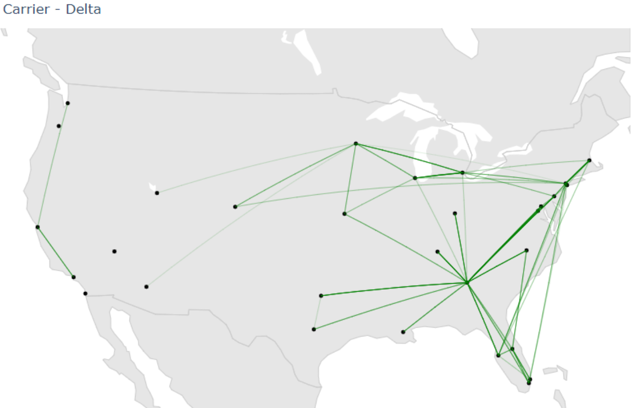
$$\log\det\Theta - \operatorname{tr}(S\Theta) - \rho||\Theta||_1,$$

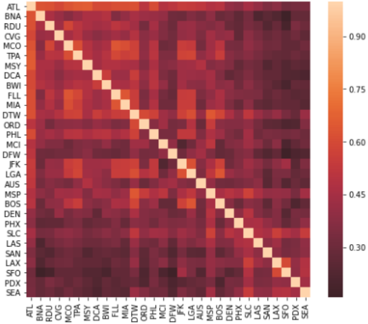


$$\beta^*_{ij} = \max\{\beta_{ij}, \beta_{ji}\}$$

Credit: Hastie and Tibshirani (2008) and CMU Lecture notes: 10-708 PGMs

Parameter Estimation: Intuition





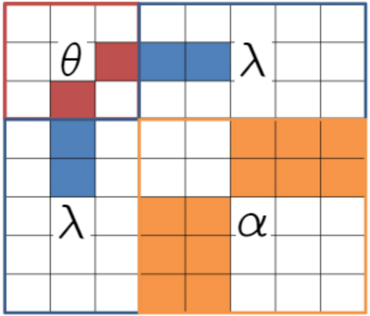
Learning the Structure of Mixed Graphical Models (Lee and Hastie 2014)

- Object of interest $\boldsymbol{\Theta} = \{\theta, \alpha, \lambda\}$
- Maximum likelihood estimation is hard - $D(\eta_r)$

Pseudolikelihood (Besag 1975):

$$\tilde{\ell}(\boldsymbol{\Theta} \mid y, z) = -\sum_{r}^{p} \log p(y_r \mid y_{\setminus r},\ z\,; \boldsymbol{\Theta}) - \sum_{r'}^{q} \log p(z_{r'} \mid z_{\setminus r'},\ y\,; \boldsymbol{\Theta})$$

Estimation: Gaussian-Potts



Symmetric matrix represents the parameters Θ *of the model.*

Credit: Lee and Hastie 2014

Parameter Estimation: Gaussian-Potts

Partition $X = (Y, Z)$ where there are p Gaussian and q Multinomial nodes

1. $\theta_{rt} = 0 \iff X_r \perp X_t \,|\, (X_{V\setminus\{r,t\}},\ Z)$
2. $\lambda_{rr'} = 0$ for all $\mathcal{Z}_{r'} \iff X_r \perp Z_{r'} \,|\, (X_{V\setminus r},\ Z_{V\setminus r'})$
3. $\alpha_{r't'} = 0$ for all values in $\mathcal{Z}_{r'}$ and $\mathcal{Z}_{t'} \iff X_r \perp Z_{r'} \,|\, (Z_{V\setminus\{r',t\}},\ Y)$

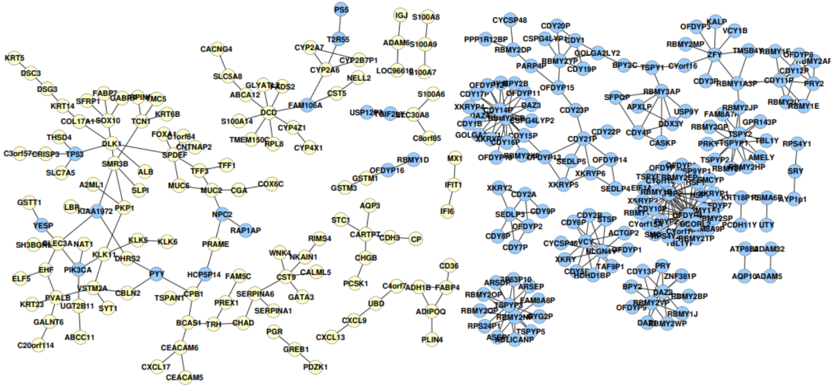
Group-Lasso (Uncalibrated):

$$\min_{\Theta} \ell_\omega(\Theta) = \ell(\Theta) + \omega\left(\sum_{r=1}^{p} \sum_{t=1}^{r-1} |\theta_{rt}| + \sum_{r=1}^{p} \sum_{r'=1}^{q} \|\lambda_{rr'}\|_2 + \sum_{r'=1}^{q} \sum_{t'=1}^{r'-1} \|\alpha_{r't'}\|_F \right)$$

Remaining Work

- Simulation studying comparing the node-neighborhood selection estimation procedure with the group-lasso proposed by Lee and Hastie (2014).
- Test model on real world data
- Spend too much time on network visualization

Questions?



Breast cancer genetic network (Poisson-Ising) Credit: Yang et al. 2014

References

Yang et al. 2014, Graphical Models via Univariate Exponential Family Distributions

Yang et al. 2012, Mixed Graphical Models via Exponential Families

Besag 1974, Spatial Interaction and the Statistical Analysis of Lattice Systems

Besag 1975, Statistical Analysis of Non-Lattice Data

Jalali et al. 2011, On Learning Discrete Graphical Models using Group-Sparse Regularization

Lee et al. 2014, Learning the Structure of Mixed Graphical Models

Wainwright and Jordan 2008, Graphical Models, Exponential Families, and Variational Inference

Ravikumar et al. 2010, High-Dimensional Ising Model Selection Using ℓ_1 Regularized Logistic Regression

Friedman et al. 2007, Sparse Inverse Covariance Estimation with the Graphical Lasso