Gene Network inference for high-dimensional problems

A. Mohammadi and E. Wit

March 28, 2013
**Motivation**

Flow cytometry data with 11 proteins from Sachs et al. (2005)
RESULT FOR CELL SIGNALING DATA
Problem in Bayesian graph estimation

\[ p(G|data) = \frac{p(G)p(data|G)}{\sum_{G \in G} p(G)p(data|G)} \]

Trans-dimensional MCMC in general

- Reversible-jump MCMC
- Birth-death MCMC

Our solution

- We proposed birth-death MCMC method for undirected graph estimation
- Implement to R: BDgraph package
Gaussian graphical model

Respect to graph $G = (V, E)$ as

$$\mathcal{M}_G = \left\{ \mathcal{N}_p(0, \Sigma) \mid K = \Sigma^{-1} \text{ is positive definite based on } G \right\}$$

Pairwise Markov property

$$X_i \perp X_j \mid X_{V \backslash \{i,j\}} \iff k_{ij} = 0,$$

$$K = \begin{bmatrix} k_{11} & k_{12} & k_{13} & 0 \\ k_{22} & 0 & 0 & 0 \\ k_{33} & k_{34} & 0 & 0 \\ k_{44} & 0 & 0 & 0 \end{bmatrix}$$
BIRTH-DEATH PROCESS


- Birth-death MCMC: Stephens (2000) in mixture models

Birth-death process in GGM

- Adding new edge in birth and deleting edge in death time
SIMPLE CASE

Graph specific element of BDMCMC method

Examples
Simple case
SIMPLE CASE

\[ K = \begin{bmatrix} k_{11} & k_{12} & k_{13} & 0 \\ k_{22} & 0 & 0 \\ k_{33} & k_{34} \\ k_{44} \end{bmatrix} \]

\[ K_{12}^{-} = \begin{bmatrix} k_{11}^* & k_{12}^* & k_{13}^* & 0 \\ k_{22}^* & 0 & k_{24}^* \\ k_{33}^* & k_{34}^* \\ k_{44}^* \end{bmatrix} \]
Canvergency

Preston (1976): Backward Kolmogorov
Under Balance condition, process converges to unique stationary distribution.

Mohammadi and Wit (2013): BDMCMC in GGM
Stationary distribution = Posterior distribution of (G,K)
So, relative sojourn time in graph $G = p(G|data)$
**Proposed BDMCMC Algorithm**

**Step 1:** (a). Calculate birth and death rates

\[ \beta_{\xi}(K) = \lambda_b, \quad \text{new link } \xi = (i, j) \]

\[ \delta_{\xi}(K) = \frac{b_{\xi}(k_{\xi}) p(G^-_{\xi}, K^-_{\xi} | x)}{p(G, K | x)} \lambda_b, \quad \text{existing link } \xi = (i, j) \]

(b). Calculate waiting time,

(c). Simulate type of jump, birth or death

**Step 2:** Sampling new precision matrix: \( K^+_\xi \) or \( K^-_\xi \)
PROPOSED PRIOR DISTRIBUTIONS

Prior for graph

- Discrete Uniform
- Truncated Poisson according to number of links

Prior for precision matrix

- G-Wishart: \( W_G(b, D) \)
  \[
p(K|G) \propto |K|^{(b-2)/2} \exp \left\{ -\frac{1}{2} tr(DK) \right\}
  \]
  \[
  I_G(b, D) = \int_{\mathcal{P}_G} |K|^{(b-2)/2} \exp \left\{ -\frac{1}{2} tr(DK) \right\} dK
  \]
**Computing Death Rates**

\[
\delta_{\xi}(K) = \frac{I_G(b, D)}{I_{G_{\xi}}(b, D)} \left( \frac{|K - \xi|}{|K|} \right)^{(b^* - 2)/2} \exp \left\{ -\text{tr}(D^*(K-\xi - K))/2 \right\} \gamma_b
\]

\[
\frac{I_G(b,D)}{I_{G-\xi}(b,D)} = 2\sqrt{\pi} t_{ii} t_{jj} \frac{\Gamma((b+\nu_i)/2)}{\Gamma((b+\nu_i - 1)/2)} \frac{E_G[f_T(\psi^\nu)]}{E_{G-\xi}[f_T(\psi^\nu)]}
\]
BDgraph package

- Graph estimation for high-dimensional cases
- Graph estimation for low-dimensional cases
**Simulation: 8 Nodes**

\[ M_G = \{ N_8(0, \Sigma) | K = \Sigma^{-1} \in \mathbb{P}_G \} \]

\[
K = \begin{bmatrix}
1 & .5 & 0 & 0 & 0 & 0 & 0 & .4 \\
1 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]
## SOME RESULT

### Effect of Sample size

<table>
<thead>
<tr>
<th>Number of data</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(\text{true graph} \mid \text{data})$</td>
<td>0.018</td>
<td>0.067</td>
<td>0.121</td>
<td>0.2</td>
<td>0.22</td>
<td>0.35</td>
</tr>
<tr>
<td>false positive</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>false negative</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

![Diagrams](image_url)
**Simulation: Circle graph with 120 nodes**

$$
\mathcal{M}_G = \left\{ \mathcal{N}_{120}(0, \Sigma) | K = \Sigma^{-1} \in \mathbb{P}_G \right\},
$$

- $n = 2000 \ll 7260$
- Priors: $K \sim W_G(3, I_{120})$ and $G \sim TU$ (all possible graphs)
- 10000 iterations and 5000 iterations as burn-in

**Result**

- Time: 4 hours
- $p(\text{true graph} \mid \text{data}) = 0.09$ which is most probable graph
**SUMMARY**