Graphical Model Structure Inference Using Trees

Loïc Schwaller, Stéphane Robin, Michael Stumpf

INRA MIA/AgroParisTech
Imperial College London

April 11, 2014

Graphical Models

- $\mathbf{X} = (X_1, ..., X_p)$ random variables

Graphical Models

- $\mathbf{X} = (X_1, ..., X_p)$ random variables
- $G = (V, E)$ **undirected** graph with $V = \{1, ..., p\}$

Graphical Models

- $\mathbf{X} = (X_1, ..., X_p)$ random variables
- $G = (V, E)$ **undirected** graph with $V = \{1, ..., p\}$

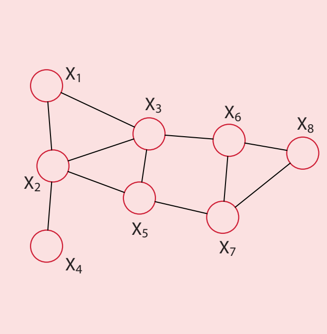
Definition

A graphical model following G is a probabilistic model for which the conditional dependence structure of $\mathbf{X}$ is given by G.

Inference Problem

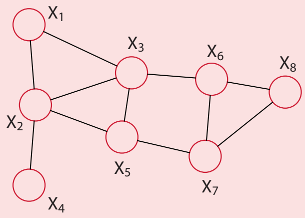
Inference Problem

Graph



Inference Problem

Graph

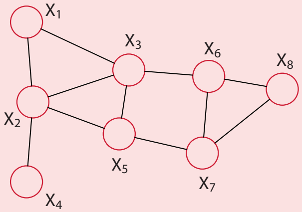


PDF

$p(X_1, \ldots, X_8)$

Inference Problem

Graph



PDF

$p(X_1, ..., X_8)$

Sample

$$\mathbf{x}^{(1)} = (x_1^{(1)}, ..., x_8^{(1)})$$
$$\mathbf{x}^{(2)} = (x_1^{(2)}, ..., x_8^{(2)})$$
$$...$$
$$\mathbf{x}^{(n)} = (x_1^{(n)}, ..., x_8^{(n)})$$

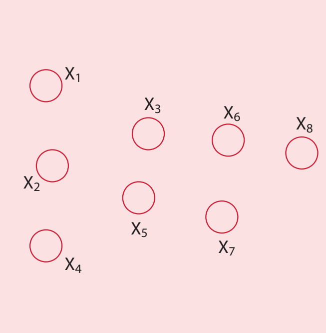
Inference Problem

Sample

$$\mathbf{x}^{(1)} = (x_1^{(1)}, ..., x_8^{(1)})$$
$$\mathbf{x}^{(2)} = (x_1^{(2)}, ..., x_8^{(2)})$$
$$...$$
$$\mathbf{x}^{(n)} = (x_1^{(n)}, ..., x_8^{(n)})$$

Inference Problem

Graph

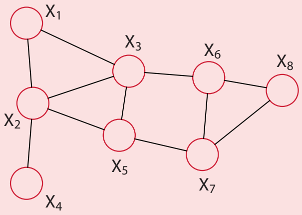


Sample

$$\mathbf{x}^{(1)} = (x_1^{(1)}, ..., x_8^{(1)})$$
$$\mathbf{x}^{(2)} = (x_1^{(2)}, ..., x_8^{(2)})$$
$$...$$
$$\mathbf{x}^{(n)} = (x_1^{(n)}, ..., x_8^{(n)})$$

Inference Problem

Graph



Inference

$\longleftarrow$

Sample

$$\mathbf{x}^{(1)} = (x_1^{(1)}, \ldots, x_8^{(1)})$$
$$\mathbf{x}^{(2)} = (x_1^{(2)}, \ldots, x_8^{(2)})$$
$$\ldots$$
$$\mathbf{x}^{(n)} = (x_1^{(n)}, \ldots, x_8^{(n)})$$

Rationale

$D = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(n)}\}$
G (unknow) underlying graph

Rationale

$D = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(n)}\}$
G (unknow) underlying graph

$\mathcal{G} = \{\text{graphs on } V\}$

Rationale

$D = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(n)}\}$
G (unknow) underlying graph

$\mathcal{G} = \{\text{graphs on } V\}$
For all $G' \in \mathcal{G}$,

$$\omega(G', D) = p(D|G')p(G')$$

Rationale

$D = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(n)}\}$
G (unknow) underlying graph

$\mathcal{G} = \{\text{graphs on } V\}$
For all $G' \in \mathcal{G}$,

$$\omega(G', D) = p(D|G')p(G')$$

For an edge $\{i, j\}$,
if $\mathbf{1}_{\{i,j\}}(G') = 1$ when $\{i, j\} \in E_{G'}$ and 0 otherwise

$$p(\{i, j\} \in E_G|D) = \frac{1}{Z(D)} \sum_{G' \in \mathcal{G}} \omega(G', D)\mathbf{1}_{\{i,j\}}(G')$$

Practical Issues

$\mathcal{G}$ is an enormous set.

p vertices $\longrightarrow \frac{p(p-1)}{2}$ possible edges $\longrightarrow |\mathcal{G}| = 2^{p(p-1)/2}$

Example p=10, $|\mathcal{G}| \approx 3.5 \cdot 10^{13}$

- Impossible to exhaustively explore $\mathcal{G}$ as soon as we consider more than a few vertices.

How we deal with it

- Consider a subset $\mathcal{T} \subset \mathcal{G}$.

How we deal with it

- Consider a subset $\mathcal{T} \subset \mathcal{G}$.

Spanning Tree

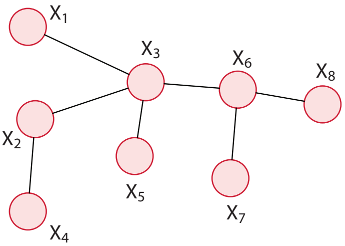
A **spanning tree** T on the set of vertices V is a connected graph with no cycles.

How we deal with it

- Consider a subset $\mathcal{T} \subset \mathcal{G}$.

Spanning Tree

A **spanning tree** T on the set of vertices V is a connected graph with no cycles.

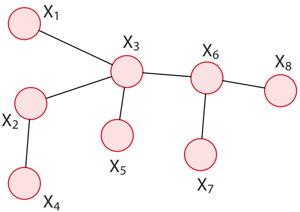


$\mathcal{T} = \{\text{spanning trees}\}$

$|\mathcal{T}| = p^{p-2}$

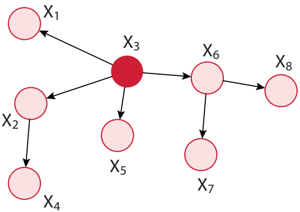
Example p=10, $|\mathcal{T}| = 1 \cdot 10^8$

Tree distribution



$$p(x_1, ..., x_p) = \prod_{i \in V} p_i(x_i) \prod_{\{i,j\} \in E_T} \frac{p_{ij}(x_i, x_j)}{p_i(x_i) p_j(x_j)}$$

Tree distribution



$$p(x_1, ..., x_p) = \prod_{i \in V} p_i(x_i) \prod_{\{i,j\} \in E_T} \frac{p_{ij}(x_i, x_j)}{p_i(x_i) p_j(x_j)}$$

$$p(x_1, ..., x_p) = \prod_{i \in V} p_{i|pa(i)}(x_i | x_{p(i)})$$

Suppose that

$$\forall T \in \mathcal{T},\ p(D|T) = \frac{1}{Z(D)} \prod_{\{i,j\} \in E_T} W_{ij}(D)$$

where $Z(D)$ is a normalizing constant

$$Z(D) = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} W_{ij}(D)$$

Suppose that

$$\forall T \in \mathcal{T},\ p(D|T) = \frac{1}{Z(D)} \prod_{\{i,j\} \in E_T} W_{ij}(D)$$

where $Z(D)$ is a normalizing constant

$$Z(D) = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} W_{ij}(D)$$

$Z(D)$ (and other integrated quantities) can be efficiently computed thanks to algebra.

Laplacian Matrix

$W = (W_{ij})$ symmetric weight matrix ($\forall i, W_{ii} = 0$)

Laplacian Matrix

The Laplacian matrix $Q = (Q_{ij})$ relatively to the weights W is given by

$$Q_{ij} = \begin{cases} -W_{ij} & \text{if } i \neq j \\ \sum_j W_{ij} & \text{if } i = j \end{cases}$$

Matrix-Tree Theorem

Theorem

Let Q be the Laplacian matrix associated to weights W. Let $\overline{Q}_{ij}$ denote the $(i,j)^{th}$ minor of Q.

- *All $\overline{Q}_{ij}$ are equal.*
- *The following identity holds*

$$\sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} W_{ij} = \overline{Q}_{ij}$$

Seth Chaiken. *A Combinatorial Proof of the All Minors Matrix Tree Theorem.* 1982

Matrix-Tree Theorem

Suppose that

$$\forall T \in \mathcal{T},\ p(D|T) = \frac{1}{Z(D)} \prod_{\{i,j\} \in E_T} W_{ij}$$

Matrix-Tree Theorem

Suppose that

$$\forall T \in \mathcal{T},\ p(D|T) = \frac{1}{Z(D)} \prod_{\{i,j\} \in E_T} W_{ij}$$

- $Z(D)$ is computed as a direct result of the M-T theorem.

Matrix-Tree Theorem

Suppose that

$$\forall T \in \mathcal{T},\ p(D|T) = \frac{1}{Z(D)} \prod_{\{i,j\} \in E_T} W_{ij}$$

- $Z(D)$ is computed as a direct result of the M-T theorem.
- If we set $W^* = W$ except for $W^*_{kl} = W^*_{lk} = 0$

Matrix-Tree Theorem

Suppose that

$$\forall T \in \mathcal{T},\ p(D|T) = \frac{1}{Z(D)} \prod_{\{i,j\} \in E_T} W_{ij}$$

- $Z(D)$ is computed as a direct result of the M-T theorem.
- If we set $W^* = W$ except for $W^*_{kl} = W^*_{lk} = 0$

$$Z^*(D) = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} W^*_{ij} = \sum_{\substack{T \in \mathcal{T} \\ \{k,l\} \notin E_T}} \prod_{\{i,j\} \in E_T} W_{ij}$$

$$p(\{k,l\} \in E_G|D) = 1 - \frac{Z^*(D)}{Z(D)}$$

Bayesian Framework

- Choose a family of emission distributions (multinomial, gaussian...)
- Consider for a given tree T every possible distribution in this family that can be factorised on T.

Bayesian Framework

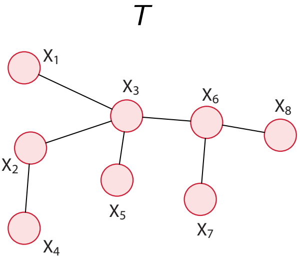
- Choose a family of emission distributions (multinomial, gaussian...)
- Consider for a given tree T every possible distribution in this family that can be factorised on T.

$T \in \mathcal{T}$
$\mathcal{F}_T$ a family of tree distributions parametrised by a set of parameters θ_T

$$p(D|T) = \int p(D|\theta_T, T)p(\theta_T|T)d\theta_T$$

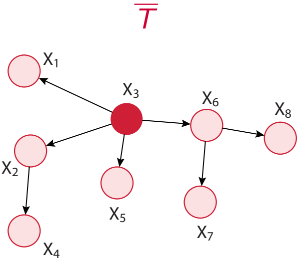
Bayesian Framework

$$p(x|T) = \int p(x|\theta_T, T)p(\theta_T|T)d\theta_T$$



Bayesian Framework

$$p(x|T) = \int p(x|\theta_{\overline{T}}, T)p(\theta_{\overline{T}}|T)d\theta_{\overline{T}}$$

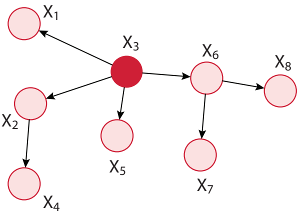


- Directed parametrisation

Bayesian Framework

$$p(x|T) = \int \left[\prod_{i \in V} p(x_i | x_{pa(i)}, \theta_{i|pa(i)})\right] p(\theta_{\overline{T}}|T) d\theta_{\overline{T}}$$

$\overline{T}$

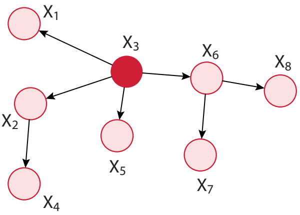


- Directed parametrisation
- Tree distribution

Bayesian Framework

$$p(x|T) = \int \left[\prod_{i \in V} p(x_i|x_{pa(i)}, \theta_{i|pa(i)})\right] \prod_{i \in V} p(\theta_{i|pa(i)}|T) d\theta_{\overline{T}}$$

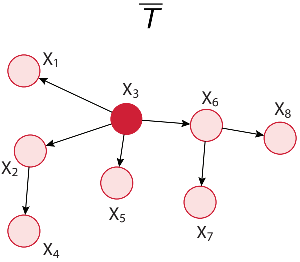
$\overline{T}$



- Directed parametrisation
- Tree distribution
- Global parameter independence

Bayesian Framework

$$p(x|T) = \prod_{i \in V} \int p(x_i | x_{pa(i)}, \theta_{i|pa(i)}) p(\theta_{i|pa(i)} | T) d\theta_{i|pa(i)}$$

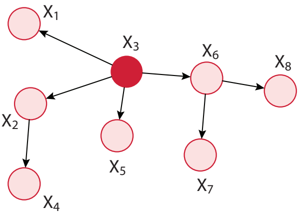


- Directed parametrisation
- Tree distribution
- Global parameter independence

Bayesian Framework

$$p(x|T) = \prod_{i \in V} \int p(x_i | x_{pa(i)}, \theta_{i|pa(i)}) p(\theta_{i|pa(i)}) d\theta_{i|pa(i)}$$

$\overline{T}$

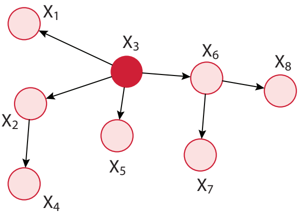


- Directed parametrisation
- Tree distribution
- Global parameter independence
- Prior modularity

Bayesian Framework

$$p(x|T) = \prod_{i \in V} p(x_i | x_{pa(i)})$$

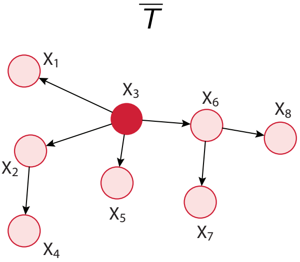
$\overline{T}$



- Directed parametrisation
- Tree distribution
- Global parameter independence
- Prior modularity

Bayesian Framework

$$p(x|T) = \prod_{i \in V} p(x_i) \prod_{\{i,j\} \in E_T} \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$$

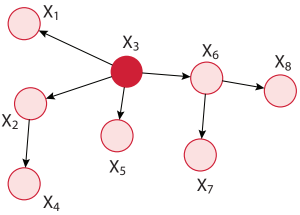


- Directed parametrisation
- Tree distribution
- Global parameter independence
- Prior modularity

Bayesian Framework

$$p(x|T) = \frac{1}{Z(x)} \prod_{\{i,j\} \in E_T} W_{ij}(x)$$

$\overline{T}$



- Directed parametrisation
- Tree distribution
- Global parameter independence
- Prior modularity

Assumptions

- Global parameter independence

$$p(\theta_{\overline{T}}|T) = \prod_{i \in V} p(\theta_{i|pa(i)}|T)$$

Dan Geiger and David Heckerman. *Parameter Priors for Directed Acyclic Graphical Models and the Characteriration of Several Probability Distributions.* 1999

Assumptions

- Global parameter independence

$$p(\theta_{\overline{T}}|T) = \prod_{i \in V} p(\theta_{i|pa(i)}|T)$$

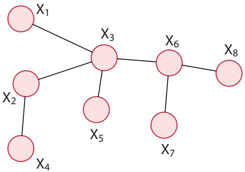
- Prior modularity

$$p(\theta_{i|pa(i)}|T) = p(\theta_{i|pa(i)})$$

Dan Geiger and David Heckerman. *Parameter Priors for Directed Acyclic Graphical Models and the Characteriration of Several Probability Distributions.* 1999

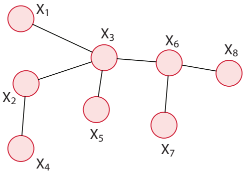
Multinomial case

$\mathbf{X} \in \{1, \ldots, r\}^p$
$T \in \mathcal{T}$



Multinomial case

$\mathbf{X} \in \{1, ..., r\}^p$
$T \in \mathcal{T}$



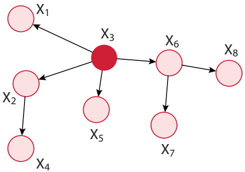
Undirected parametrisation

$$\theta_i(l) = P_i(l) \qquad \forall i \in V, l \in [r]$$
$$\theta_{ij}(l, k) = P_{ij}(l, k) \qquad \forall \{i,j\} \in E_T, (l,k) \in [r]^2$$

$$\theta_T = \{\theta_{ij}(k, l) | (i, j) \in E_T, (l, k) \in [r]^2\}$$

Multinomial case

$\mathbf{X} \in \{1, ..., r\}^p$
$T \in \mathcal{T}$



Directed parametrisation

$u \in V$ root $\longrightarrow$ directed tree $\overline{T}$

$$\theta_{i|j}(l|k) = P_{i|j}(l|k) \qquad \forall \{i,j\} \in E_{\overline{T}}, (l,k) \in [r]^2$$

$$\theta_{\overline{T}} = \{\theta_{i|j}(k|l) | (i,j) \in E_{\overline{T}}, (l,k) \in [r]^2\}$$

Multinomial case

Prior

- Dirichlet distribution

$$\theta_{i|j}|X_j = k \sim D(N'_{ij}(1,k), ..., N'_{ij}(r,k))$$

$\forall (i,j) \in V^2$

$$N'_{ij}(l,k) = N'_{ji}(k,l)$$
$$\sum_{k=1}^{r} N'_{ij}(l,k) = N'_i(l)$$
$$\sum_{l=1}^{r} N'_i(l) = N'$$

Multinomial case

Prior

- Dirichlet distribution

$$\theta_{i|j}|X_j = k \sim D(N'_{ij}(1,k), ..., N'_{ij}(r,k))$$

$\forall (i,j) \in V^2$

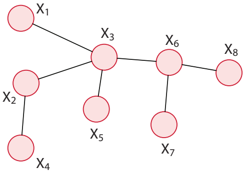
$$N'_{ij}(l,k) = N'_{ji}(k,l)$$

$$\sum_{k=1}^{r} N'_{ij}(l,k) = N'_i(l)$$

$$\sum_{l=1}^{r} N'_i(l) = N'$$

- Define prior distribution for the parameters of all tree distributions.

Gaussian Graphical Model



$\mathbf{X} \in \mathbb{R}^p$
$T \in \mathcal{T}$

$$\mathbf{X} \sim \mathcal{N}(\mu, \Lambda^{-1})$$

with normal-Wishart conjugate prior.

$$\Lambda = \begin{pmatrix} * & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & * & * & * & \cdot & \cdot & \cdot & \cdot \\ * & * & * & \cdot & * & * & \cdot & \cdot \\ \cdot & * & \cdot & * & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & * & \cdot & * & \cdot & \cdot & \cdot \\ \cdot & \cdot & * & \cdot & \cdot & * & * & * \\ \cdot & \cdot & \cdot & \cdot & \cdot & * & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & * & \cdot & * \end{pmatrix}$$

Simulations

RAF Network[3]

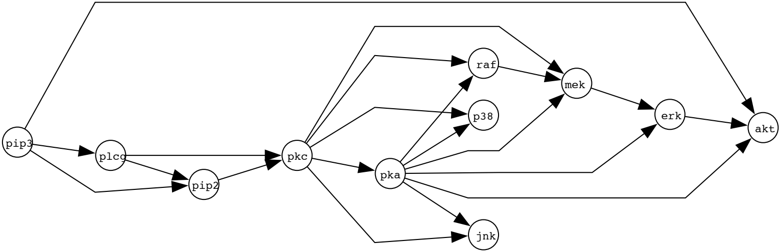


Figure : Cellular signalling network describing the interactions of 11 phosphorylated proteins and phospholipids in human immune cells.

[3] Adriano V. Werhli, Marco Grzegorczyk, and Dirk Husmeier. *Comparative Evaluation of Reverse Engineering Gene Regulatory Networks with Relevance Networks, Graphical Gaussian Models and Bayesian Networks*. 2006.

Simulation Scheme

$\mathbf{A} \longleftarrow$ undirected adjacency matrix of the RAF network
$\mathbf{\Lambda} \longleftarrow$ partial correlation matrix

- $\Lambda \longleftarrow 0.5\mathbf{A}$
- $\Lambda_{ii} \longleftarrow \sum_j \Lambda_{ij}$
- normalisation

$\mathbf{\Sigma} \longleftarrow \mathbf{\Lambda}^{-1}$ covariance matrix

$$x^{(i)} \sim \mathcal{N}(0, \Sigma) \quad \text{i.i.d.}$$

$i = 1, ..., n$
$n = 10, 20, ..., 200$
10 repetitions per sample size

Inference Methods

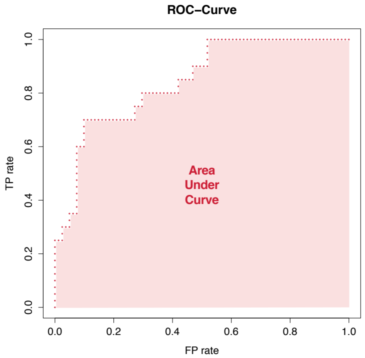
- Relevance Network (RN)
  - Correlation coefficient
- Gaussian Graphical Models (GGM)
  - Partial correlation coefficient
  - Regularized shrinkage estimate[4]

- Tree-based Method
  - Gaussian distribution, Normal-Wishart prior

[4] J. Schäfer and K. Strimmer. *An empirical Bayes approach to inferring large-scale gene association networks.* 2005.

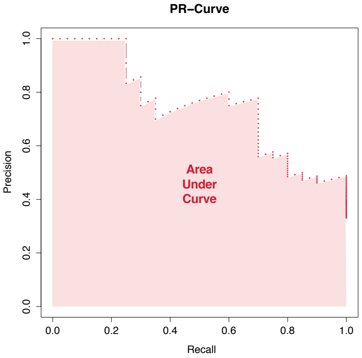
Assessment

ROC Curve



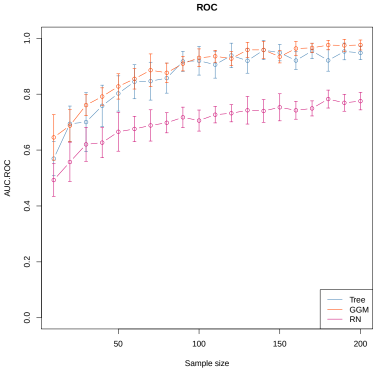
$$TPR = \frac{TP}{P} \qquad FPR = \frac{FP}{N}$$

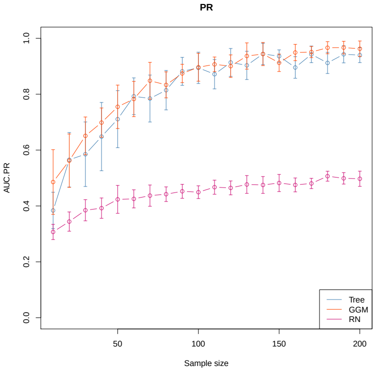
PR Curve



$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{P}$$

Results


ROC
AUC.ROC
0.0
0.2
0.4
0.6
0.8
1.0
50
100
150
200
Sample size
Tree
GGM
RN


PR
AUC.PR
0.0
0.2
0.4
0.6
0.8
1.0
50
100
150
200
Sample size
Tree
GGM
RN

Conclusion

- Exploration of a subset of all graphs.
  - Spanning trees.
- Algebraic theorem to integrate on this set.
  - Matrix-Tree theorem.

- Can be used in a Bayesian framework.

Perspectives

- But not only (mutual information, etc).
- General method
  - Network Inference
  - Network Comparison

D_1, D_2 two datasets
$D = D_1 \cup D_2$

$$\frac{Z(D)}{Z(D_1)Z(D_2)}$$

References

Seth Chaiken. *A Combinatorial Proof of the All Minors Matrix Tree Theorem*. 1982.

Dan Geiger and David Heckerman. *Parameter Priors for Directed Acyclic Graphical Models and the Characteriration of Several Probability Distributions*. 1999.

J. Schäfer and K. Strimmer. *An empirical Bayes approach to inferring large-scale gene association networks*. 2005.

Adriano V. Werhli, Marco Grzegorczyk, and Dirk Husmeier. *Comparative Evaluation of Reverse Engineering Gene Regulatory Networks with Relevance Networks, Graphical Gaussian Models and Bayesian Networks*. 2006.

Multinomial case

Prior

- Dirichlet distribution

$$\theta_{i|j}|X_j = k \sim D(N'_{ij}(1,k), ..., N'_{ij}(r,k))$$

$\forall (i,j) \in V^2$

$$N'_{ij}(l,k) = N'_{ji}(k,l)$$

$$\sum_{k=1}^{r} N'_{ij}(l,k) = N'_i(l)$$

$$\sum_{l=1}^{r} N'_i(l) = N'$$

Update

$$D = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)})$$

$$\alpha_{k,i,j,l,l'} = \mathbf{1}(\{x_i^{(k)} = l\} \cap \{x_j^{(k)} = l'\})$$

$$N_{i,j}(l,l') = \sum_{k=1}^{p} \alpha_{m,i,j,l,l'}$$

$$N_j(l') = \sum_{l=1}^{r} N_{i,j}(l,l')$$

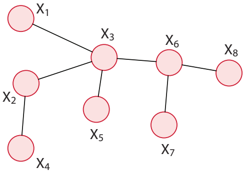
Marginal data distribution

$$p(D|T) = \frac{1}{Z} \prod_{\{i,j\} \in E_T} W_{ij}$$

$$W_{ij} = \prod_{l,l' \in \Omega} \frac{\Gamma(N'_i(l))}{\Gamma(N'_i(l) + N_i(l))} \frac{\Gamma(N'_j(l'))}{\Gamma(N'_j(l') + N_j(l'))} \frac{\Gamma(N'_{i,j}(l,l') + N_{i,j}(l,l'))}{\Gamma(N'_{i,j}(l,l'))}$$

Gaussian Graphical Model

$\mathbf{X} \in \mathbb{R}^p$
$T \in \mathcal{T}$



$$\mathbf{X} \sim \mathcal{N}(\mu, \Lambda^{-1})$$

$$\Lambda = \begin{pmatrix} * & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & * & * & * & \cdot & \cdot & \cdot & \cdot \\ * & * & * & \cdot & * & * & \cdot & \cdot \\ \cdot & * & \cdot & * & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & * & \cdot & * & \cdot & \cdot & \cdot \\ \cdot & \cdot & * & \cdot & \cdot & * & * & * \\ \cdot & \cdot & \cdot & \cdot & \cdot & * & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & * & \cdot & * \end{pmatrix}$$

Prior

- **Prior distribution**
  Normal-Wishart on (μ, Λ).

$$\Lambda \sim \mathcal{W}(\Phi, \alpha)$$
$$\mu|\Lambda \sim \mathcal{N}(\nu, (\alpha_\mu \Lambda)^{-1})$$

- **Hyperparameters**

ν	mean of μ
α_μ	relative precision of μ
α	degrees of freedom of the precision of Λ
Φ	positive-definite parametric matrix

Update

$$\nu' \longleftarrow \frac{\alpha_\mu \nu + n\bar{\mathbf{x}}}{\alpha_\mu + n}$$

$$\alpha'_\mu \longleftarrow \alpha_\mu + n$$

$$\Phi' \longleftarrow \Phi + (n-1)S_n^2 + \frac{\alpha_\mu n}{\alpha_\mu + n}(\nu - \bar{\mathbf{x}})(\nu - \bar{\mathbf{x}})^T$$

$$\alpha' \longleftarrow \alpha + n$$

Marginal data distribution

$$p(D|T) = \frac{1}{Z} \prod_{\{i,j\} \in E_T} W_{ij}$$

$$W_{ij} = \left(\frac{|(\Phi'^{-1})_{(i,j)}|^{1+(\alpha-p+n)/2}}{(|(\Phi'^{-1})_j||(\Phi'^{-1})_i|)^{(\alpha-p+n+1)/2}} \right) \left(\frac{(|(\Phi^{-1})_j||(\Phi^{-1})_i|)^{(\alpha-p+1)/2}}{|(\Phi^{-1})_{(i,j)}|^{1+(\alpha-p)/2}} \right)$$