Detecting change-points in the structure of a network: Exact Bayesian inference

S. Robin

Joint work with L. Schwaller

Emerging Topics in Biological Networks and Systems Biology, Oct. 2017, Uppsala
Example: Gene regulatory network along time

Data: [AFI+02]

\[ Y_{jt} = \text{expression of } j \text{ at time } t \]
Example: Gene regulatory network along time

**Data:** [AFI+02]

\[ Y_{jt} = \text{expression of } j \text{ at time } t \]

'**Model**':

\[ G_t = \text{gene regulatory network at time } t \]
\[ = \text{graphical model of } Y_t = (Y_{jt})_j \]
Example: Gene regulatory network along time

Data: \([\text{AFI}^+02]\)

\[ Y_{jt} = \text{expression of } j \text{ at time } t \]

'Model':

\[ G_t = \text{gene regulatory network at time } t \]
\[ = \text{graphical model of } Y_t = (Y_{jt})_j \]

Questions:

- Is \( G_t \) constant along time or is there some 'gene rewiring'?
- If not, when does it change?
- And what is the network within each period of time?
Example of output

Data: $n = 67$ time points, $p = 11$ genes, four expected regions
Example of output

Data:  $n = 67$ time points, $p = 11$ genes, four expected regions

Posterior probability of change-points:
Example of output

Data: \( n = 67 \) time points, \( p = 11 \) genes, four expected regions

Posterior probability of change-points:

Inferred networks:
Motivating example

Similar problems

Ecology:

\[ Y_{jt} = \text{abundance of species } j \text{ at time } t \text{ in a given medium} \]

\[ G_t = \text{interaction structure between species at time } t. \]

▶ Time-evolving species interaction network?
Similar problems

Ecology:

\[ Y_{jt} = \text{abundance of species } j \text{ at time } t \text{ in a given medium} \]

\[ G_t = \text{interaction structure between species at time } t. \]

▶ Time-evolving species interaction network?

Neuroscience:

\[ Y_{jt} = \text{activity of brain region } j \text{ at time } t \]

\[ G_t = \text{connectivity structure between regions at time } t. \]

▶ Time-evolving connectivity network?
Outline

Bayesian inference with discrete parameters

Change-point detection

Network inference

Detecting changes in a graphical model

Discussion
A reminder on Bayesian inference

Typical Bayesian model:

Parameter $\Theta \sim p(\Theta) = \text{prior distribution}$

Data $Y \sim p(Y|\Theta) = \text{likelihood}$

Aim: evaluate $p(\Theta|Y) = \text{posterior distribution}$

where

$$p(\Theta|Y) = \frac{p(\Theta)p(Y|\Theta)}{p(Y)}$$

Requires to integrate over the whole parameter space:

$$p(Y) = \int p(\Theta)p(Y|\Theta) \, d\Theta$$
Bayesian inference with mixed parameters

Mixed parameter: $\Theta = (\theta, T)$

- $\theta = \text{(means, variances, correlations)}$: continuous parms,
- $T = \text{(segmentation, graph)}$: discrete parms,
Bayesian inference with mixed parameters

Mixed parameter: $\Theta = (\theta, T)$

- $\theta = \text{(means, variances, correlations): continuous parms,}$
- $T = \text{(segmentation, graph): discrete parms,}$

Even if $\int d\theta$ raises no issue, we are left with

$$p(Y) = \sum_{T \in \mathcal{T}} \int p(Y, \theta, T) \, d\theta = \sum_{T \in \mathcal{T}} p(Y, T)$$

- Intractable when $\#\mathcal{T}$ grows (super-)exponentially with $n$ or $p$
Bayesian inference with discrete parameters

Examples.

- Change-point detection:
- Network inference = structure inference
- Combination of both
Bayesian inference with discrete parameters

Examples.
- Change-point detection:
- Network inference = structure inference
- Combination of both

Main approaches for Bayesian inference.
- Stochastic: Monte-Carlo sampling, MCMC, SMC, ...
- Approximation: variational Bayes, INLA, ...
- Hard headed: 'exact' computation
Outline

Bayesian inference with discrete parameters

Change-point detection

Network inference

Detecting changes in a graphical model

Discussion
A change-point detection model

Segmentation $T =$ set of adjacent segments. $\mathcal{T}^K = \mathcal{T}_{1:n}^K$ set of all possible segmentations.
A change-point detection model

Segmentation $T = \text{set of adjacent segments}$. $\mathcal{T}^K = \mathcal{T}_{1:n}^K$ set of all possible segmentations.
A change-point detection model

Segmentation $T = \text{set of adjacent segments}$. $\mathcal{T}^K = \mathcal{T}_{1:n}^K$ set of all possible segmentations.

\[
p(Y | T) = \prod_{r \in T} \int p(Y^r | \theta_r) p(\theta_r) \, d\theta_r = \prod_{r \in T} p(Y^r), \quad Y^r = (Y_t)_{t \in r}
\]
A change-point detection model

Segmentation $T = \text{set of adjacent segments}$. $\mathcal{T}^K = \mathcal{T}_{1:n}^K$ set of all possible segmentations.

\[
p(Y | T) = \prod_{r \in T} \int p(Y^r | \theta_r) p(\theta_r) \, d\theta_r = \prod_{r \in T} p(Y^r), \quad Y^r = (Y_t)_{t \in r}
\]

Segmentation space:

\[
\# \mathcal{T}_{1:n}^K = \binom{n-1}{K-1} \approx \left(\frac{n}{K}\right)^K
\]
Some quantities of interest

Under mild assumptions (incl. \( p(T) \propto \prod_{r \in T} a_r \)).

Marginal likelihood:

\[
p(Y \mid K) = \sum_{T \in \mathcal{T}^K} p(T \mid K)p(Y \mid T) \propto \sum_{T \in \mathcal{T}^K} \prod_{r \in T} a_r p(Y^r)
\]

with normalizing constant \( \sum_{T \in \mathcal{T}^K} \prod_{r \in T} a_r \).
Some quantities of interest

Under mild assumptions (incl. $p(T) \propto \prod_{r \in T} a_r$).

Marginal likelihood:

$$p(Y | K) = \sum_{T \in T^K} p(T | K) p(Y | T) \propto \sum_{T \in T^K} \prod_{r \in T} a_r p(Y^r)$$

with normalizing constant $\sum_{T \in T^K} \prod_{r \in T} a_r$.

Posterior distribution of a change-point.

$$\Pr\{\tau_k = t | Y, K\} \propto \left( \sum_{T \in T^K_{1:t}} \prod_{r \in T} a_r p(Y^r) \right) \left( \sum_{T \in T^K_{t+1:n}} \prod_{r \in T} a_r p(Y^r) \right)$$
Some simple algebra

Computing the sum of

\[ f_{1, \tau_1} \times f_{\tau_1+1, \tau_2} \times \cdots \times f_{\tau_{K-1}+1, n} \]

for all \( 1 \leq \tau_1 < \tau_2 < \cdots < \tau_{K-1} < n \) is the same as ...
Some simple algebra

Computing the sum of

$$f_{1, \tau_1} \times f_{\tau_1+1, \tau_2} \times \cdots \times f_{\tau_{K-1}+1, n}$$

for all $1 \leq \tau_1 < \tau_2 < \cdots < \tau_{K-1} < n$ is the same as ...

Computing the $K$-th power of the $(n + 1) \times (n + 1)$ upper-triangular matrix (with zero diagonal) $A$:

$$[A]_{i, j+1} = f_{i, j} \quad \Rightarrow \quad \sum_{T \in T_{1:n}^K} \prod_{r \in T} f_r = [A^K]_{1, n+1}$$
Some simple algebra

Computing the sum of

\[ f_{1,\tau_1} \times f_{\tau_1+1,\tau_2} \times \cdots \times f_{\tau_{K-1}+1,n} \]

for all \( 1 \leq \tau_1 < \tau_2 < \cdots < \tau_{K-1} < n \) is the same as ...

Computing the \( K \)-th power of the \((n+1) \times (n+1)\) upper-triangular matrix (with zero diagonal) \( A \):

\[
[A]_{i,j+1} = f_{i,j} \quad \Rightarrow \quad \sum_{T \in \mathcal{T}^K_{1:n}} \prod_{r \in T} f_r = [A^K]_{1,n+1}
\]

- All terms are computed in \( O(Kn^2) \).
- To compute \( p(Y \mid K) \), take \( f_r = a_r p(Y^r) \).
- Similar ideas in [Fea06].
- \( R \) package EBS (exact Bayesian segmentation) [CR14]
Gene regulatory network

Data: \( n = 67 \) time points, \( p = 11 \) genes, four expected regions [AFI'02]

Model:
\[
t \in r : \quad Y_t | \theta_r = (\mu_r, \Sigma_r) \sim \mathcal{N}(\mu_r, \Sigma_r)
\]
→ Saturated graphical model.
Gene regulatory network

Data: \( n = 67 \) time points, \( p = 11 \) genes, four expected regions [AFI⁺02]

Model:

\[ t \in r : \ Y_t | \theta_r = (\mu_r, \Sigma_r) \sim \mathcal{N}(\mu_r, \Sigma_r) \]

→ Saturated graphical model.

Posterior probability of change-points: dotted line
Outline

Bayesian inference with discrete parameters

Change-point detection

Network inference

Detecting changes in a graphical model

Discussion
A reminder on (undirected) graphical model

Means that

\[ p(Y_1, \ldots, Y_8) \propto \psi_1(Y_1, Y_2, Y_3) \]
\[ \psi_2(Y_1, Y_4) \psi_3(Y_1, Y_5) \]
\[ \psi_4(Y_2, Y_6) \psi_5(Y_3, Y_8) \]

which implies that

\[ Y_4 \perp Y_3 \mid Y_1 \]
\[ (Y_6, Y_7) \perp Y_3 \mid Y_2 \]
\[ \ldots \]
A reminder on (undirected) graphical model

Means that

\[ p(Y_1, \ldots, Y_8) \propto \psi_1(Y_1, Y_2, Y_3) \]
\[ \psi_2(Y_1, Y_4) \psi_3(Y_1, Y_5) \]
\[ \psi_4(Y_2, Y_6) \psi_5(Y_3, Y_8) \]

which implies that

\[ Y_4 \perp Y_3 \mid Y_1 \]
\[ (Y_6, Y_7) \perp Y_3 \mid Y_2 \]
\[ \ldots \]

- \( G \) reveals the structure of conditional independences between the variables \( Y_1, \ldots, Y_p \)
- 'Network inference' problem: Based on \( \{(Y_{i1}, \ldots, Y_{ip})\}_i \) iid \( \sim p \), infer \( G \).
Tree-structures network

Tree assumption: the graph $G$ is a spanning tree $T$. 

▶ Consistent with the usual assumption that the graph is sparse (although much stronger).

▶ Not true in general, but may be sufficient for the inference on local structures, such as the existence of a given edge.
Tree-structures network

Tree assumption: the graph $G$ is a spanning tree $T$. 

Consistent with the usual assumption that the graph is sparse (although much stronger).

Not true in general, but may be sufficient for the inference on local structures, such as the existence of a given edge.
Tree-structures network

Tree assumption: the graph $G$ is a spanning tree $T$. 

Consistent with the usual assumption that the graph is sparse (although much stronger).

Not true in general, but may be sufficient for the inference on local structures, such as the existence of a given edge.
Tree-structures network

Tree assumption: the graph $G$ is a spanning tree $T$.

- Consistent with the usual assumption that the graph is sparse (although much stronger).

- Not true in general, but may be sufficient for the inference on local structures, such as the existence of a given edge.
Bayesian inference for tree-structured network [SRS15]

\[ p(Y \mid T) \text{ Markov wrt } T \]

\[ p(Y \mid T) = \prod_j p(Y_j) \prod_{(j,k) \in T} \frac{p(Y_j, Y_k)}{p(Y_j)p(Y_k)} \propto \prod_{(j,k) \in T} \psi_{jk} \]

where \( d_j \) is the degree (number of neighbors in \( T \)) of node \( j \).
Bayesian inference for tree-structured network \cite{SRS15}

\[ p(Y \mid T) \text{ Markov wrt } T \]

\[
p(Y \mid T) = \prod_j p(Y_j) \prod_{(j,k) \in T} \frac{p(Y_j, Y_k)}{p(Y_j)p(Y_k)} \propto \prod_{(j,k) \in T} \psi_{jk}\]

where \( d_j \) is the degree (number of neighbors in \( T \)) of node \( j \).

Assumptions:

- The prior on \( T \) factorizes over the edges: \( p(T) \propto \prod_{(j,k) \in T} b_{jk} \)
- The prior on \( \theta \) is hyper-Markov (see \cite{DL93})
- Holds for multinomial-Dirichlet, normal-Wishart, Gaussian copulas.
Quantities of interest

Marginal distribution.

\[ p(Y) = \sum_{T \in \mathcal{T}} p(T)p(Y | T) \propto \sum_{T \in \mathcal{T}} \prod_{j,k} b_{jk} \psi_{jk} \]

where \( \mathcal{T} \) stands for the set of all spanning trees.
Quantities of interest

Marginal distribution.

\[
p(Y) = \sum_{T \in \mathcal{T}} p(T)p(Y \mid T) \propto \sum_{T \in \mathcal{T}} \prod_{j,k} b_{jk} \psi_{jk}
\]

where \( \mathcal{T} \) stands for the set of all spanning trees.

Posterior probability for an edge to be absent.

\[
\Pr\{(j, k) \notin T \mid Y\} \propto \sum_{T \in \mathcal{T} : (j, k) \notin T} \prod_{j,k} b_{jk} \psi_{jk}
\]
Quantities of interest

Marginal distribution.

\[ p(Y) = \sum_{T \in \mathcal{T}} p(T)p(Y \mid T) \propto \sum_{T \in \mathcal{T}} \prod_{j,k} b_{jk} \psi_{jk} \]

where \( \mathcal{T} \) stands for the set of all spanning trees.

Posterior probability for an edge to be absent.

\[ \Pr\{(j, k) \notin T \mid Y\} \propto \sum_{T \in \mathcal{T}: (j, k) \notin T} \prod_{j,k} b_{jk} \psi_{jk} \]

Typical form:

\[ \sum_{T \in \mathcal{T}} \prod_{(j, k) \in T} f_{jk}, \quad \text{with} \quad \#\mathcal{T} = p^{p-2}. \]
Summing over spanning trees

Matrix-tree theorem. [Cha82]

- $F = [f_{jk}]$: a symmetric matrix with $f(j, j) = 0, f_{jk} > 0$;
- $\Delta = [\Delta_{jk}]$ its Laplacian: $\Delta_{jj} = \sum_k f_{jk}, \Delta_{jk} = -f_{jk}$.
Summing over spanning trees

Matrix-tree theorem. [Cha82]

- $F = [f_{jk}]$: a symmetric matrix with $f(j,j) = 0$, $f_{jk} > 0$;
- $\Delta = [\Delta_{jk}]$ its Laplacian: $\Delta_{jj} = \sum_k f_{jk}$, $\Delta_{jk} = -f_{jk}$.

Then the minors $|\Delta^{uv}|$ of $\Delta$ are equal and

$$|\Delta^{uv}| = \sum_{T \in \mathcal{T}} \prod_{(j,k) \in T} f_{jk}.$$
Summing over spanning trees

Matrix-tree theorem. [Cha82]

- \( F = [f_{jk}] \): a symmetric matrix with \( f(j,j) = 0, f_{jk} > 0 \);
- \( \Delta = [\Delta_{jk}] \) its Laplacian: \( \Delta_{jj} = \sum_k f_{jk}, \Delta_{jk} = -f_{jk} \).

Then the minors \( |\Delta^{uv}| \) of \( \Delta \) are equal and

\[
|\Delta^{uv}| = \sum_{T \in T} \prod_{(j,k) \in T} f_{jk}.
\]

- Quantities of interest can be computed at the cost of computing a determinant, ie \( O(p^3) \).
- Already used in [MJ06,Kir07] for tree learning.
- For edge probability, set \( f_{jk} = 0 \) (see [Kir07])
- R package Saturnin (spanning trees used for network inference) [SRS15]
Tree averaging

\[ P\{ T = T_1 | Y \} \]
Tree averaging

\[
P\{T = T_1|Y\} \quad P\{T = T_2|Y\}
\]
Tree averaging

\[ P\{ T = T_1 | Y \} \quad P\{ T = T_2 | Y \} \quad P\{ T = T_3 | Y \} \]
Tree averaging

\[
P\{T = T_1 | Y\} \\
P\{T = T_2 | Y\} \\
P\{T = T_3 | Y\} \\
P\{T = T_4 | Y\}
\]
Tree averaging

Edge posterior probabilities:

\[ P\{(j, k) \in T | Y\} \]
Tree averaging

\[ P\{ T = T_1 | Y \} \]
\[ P\{ T = T_2 | Y \} \]
\[ P\{ T = T_3 | Y \} \]
\[ P\{ T = T_4 | Y \} \]

Edge posterior probabilities:

\[ P\{ (j, k) \in T | Y \} \]

Thresholding probabilities:

\[ P\{ (j, k) \in T | Y \} \]
Simulations: Comparison with sampling among DAGs

[NPK11]: MCMC sampling over the directed acyclic graphs (multinomial case)

Area under the curves: top=ROC, bottom=PR
light grey = multinomial trees (2.2”), dark grey: multinomial DAGs (1393”)

Tree

Erdös-Rényi $p_c = 2/p$

Erdös-Rényi $p_c = 4/p$

Erdös-Rényi $p_c = 8/p$
Illustration: Raf pathway

Flow cytometry data for $p = 11$ proteins from the Raf signaling pathway [SPP+05]

'ground truth'

posterior probabilities

most likely tree

second most likely tree
Outline

Bayesian inference with discrete parameters

Change-point detection

Network inference

Detecting changes in a graphical model

Discussion
Problem:

Consider \( p \) variables observed along time;
Consider the graph \( G_t \) supporting the graphical model at time \( t \);
Does the graph \( G_t \) remain the same along time?

Examples:

1. Gene regulatory network along the Drosophila life cycle?
2. Connections between brain regions along different tasks?
Change-point in a graphical model

Problem: [SR16]

- Consider $p$ variables observed along time;
- Consider the graph $G_t$ supporting the graphical model at time $t$;
- Does the graph $G_t$ remain the same along time?

Examples:
1. Gene regulatory network along the Drosophila life cycle?
2. Connections between brain regions along different tasks?
Change-point in a graphical model

Problem: [SR16]

- Consider $p$ variables observed along time;
- Consider the graph $G_t$ supporting the graphical model at time $t$;
- Does the graph $G_t$ remain the same along time?

Examples:

1. Gene regulatory network along the *Drosophila* life cycle?

2. Connections between brain regions along different tasks?
Model

\[ K = \text{number of segments} \]

\[ R = (r_k)_k = \text{segmentation} \]

\[ T = (T_r)_r = \text{set of trees} \]
\[ T_r = \text{graphical model in segment } r \]

\[ \theta = (\theta_r)_r = \text{parameter in each segment} \]
\[ \text{typically: } \text{support}(\Sigma_r^{-1}) = T_r \]

\[ Y_t = (Y_{jt}) = \text{data collected at time } t \]
Model

\[ K = \text{number of segments} \]

\[ R = (r_k)_k = \text{segmentation} \]

\[ T = (T_r)_r = \text{set of trees} \]

\[ T_r = \text{graphical model in segment } r \]

\[ \theta = (\theta_r)_r = \text{parameter in each segment} \]

\[ \text{typically: support}(\Sigma_r^{-1}) = T_r \]

\[ Y_t = (Y_{jt}) = \text{data collected at time } t \]

Typical quantity of interest: (under factorisable prior assumptions)

\[ p(Y|K) \propto \sum_R \sum_{T=(T_r)_{r \in R}} \prod_r a_r \prod_{jk \in T_r} b_{jk} p(Y_{jk}^r) \]
Handling two sums

<table>
<thead>
<tr>
<th></th>
<th>Space size</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation</td>
<td>(\approx (n/K)^K)</td>
<td>(O(Kn^2))</td>
</tr>
<tr>
<td>Spanning trees</td>
<td>(p^{(p-2)})</td>
<td>(O(p^3))</td>
</tr>
<tr>
<td>Both</td>
<td>(\approx (n/K)^K p^K(p-2))</td>
<td>(O(\max{K, p^3} n^2))</td>
</tr>
</tbody>
</table>

Quantities of interest can be computed in \(O(p^3 n^2)\):

- \(P(\text{change-point at time } t \mid K, Y)\)
- \(P(\text{edge } (j, k) \text{ present at time } t \mid K, Y)\)
- \(P(\text{edge } (j, k) \text{ present at all } t \mid Y)\)
- \(P(K \text{ segments } \mid Y)\)
- \(+ \text{ Network comparison } P( T_1 = T_2 \mid Y_1, Y_2)\)
Gene regulatory network
Gene regulatory network

Data: \( n = 67 \) time points, \( p = 11 \) genes, four expected regions
Gene regulatory network

Data: \( n = 67 \) time points, \( p = 11 \) genes, four expected regions

Posterior probability of change-points:
Gene regulatory network

Data: \( n = 67 \) time points, \( p = 11 \) genes, four expected regions

Posterior probability of change-points:

Inferred networks:
FMRI data [CHA⁺12]

FMRI data collected on 20 patients:
\( p = 5 \) brain regions, 
\( n = 215 \) time-points.

Task changes at 
\( t = 60 \) and 120.

Top: 5 patients analyzed separately.

Bottom: joint analysis of the same 5 patients.
Outline

Bayesian inference with discrete parameters

Change-point detection

Network inference

Detecting changes in a graphical model

Discussion
Summary

To summarize.

- Exact Bayesian inference can be achieved for some fairly complex models with discrete parameter.
- Do not have to care about sampling and convergence.
- No systematic way to check when similar algebraic shortcuts exist → ad-hoc developments.
Future works

- Dealing with dependency along time.

- Influence of the prior: $p(T)$ depends on $n$ and/or $p$.

- Solve numerical issues raised by the exact evaluation of all probabilities.
References


**Graphical model framework**

**Property [Hammersley-Clifford].** \( p(Y) = p(Y_1, \ldots, Y_p) \) is Markov wrt the (decomposable) graph \( G \) iff it factorizes wrt the maximal cliques of \( G \):

\[
p(Y) \propto \prod_{C \in \mathcal{C}(G)} \psi_c(Y^c), \quad Y^c = (Y_j)_{j \in C}.
\]

\[ \rightarrow \quad G \text{ reveals the structure of conditional independences between the variables } Y_1, \ldots, Y_p. \]
Bayesian inference

Factorability assumptions

- Independent parameters in each segment:
  \[ p(\theta | T) = \prod_{r \in T} p(\theta_r) \]

- Data are independent from one segment to another
  \[ p(Y | T, \theta) = \prod_{r \in T} p(Y_r | \theta_r) \]

- Prior distribution for the segmentation:
  \[ p(T | K) = \prod_{r \in T} a_r, \quad \text{e.g. } a_r = n_r^\alpha \]
Hyper-Markov prior

Graphical model:
\[ p(Y | \theta, T) \text{ factorizes wrt edges of } T \]
(Markov wrt \( T \))
Hyper-Markov prior

Graphical model:
\( p(Y | \theta, T) \) factorizes wrt edges of \( T \)
(Markov wrt \( T \))

Desirable prior:
\( p(\theta | T) \) factorizes wrt edges of \( T \) as well
(Hyper-Markov wrt \( T \))
Hyper-Markov prior

Graphical model:
\( p(Y \mid \theta, T) \) factorizes wrt edges of \( T \) 
(Markov wrt \( T \))

Desirable prior:
\( p(\theta \mid T) \) factorizes wrt edges of \( T \) as well 
(Hyper-Markov wrt \( T \))

Averaging over \( T \):
This should hold for any tree \( T \)

\( \theta_1 \)
\( \theta_2 \)
\( \theta_3 \)
\( \theta_{12} \)
\( \theta_{13} \)
\( Y_1 \)
\( Y_2 \)
\( Y_3 \)
\( Y_4 \)
\( Y_5 \)

(\( \theta_4, \theta_5, \theta_{14}, \theta_{15} \) not drawn.)
Hyper-Markov prior

Graphical model:
\[ p(Y | \theta, T) \] factorizes wrt edges of \( T \)
(Markov wrt \( T \))

Desirable prior:
\[ p(\theta | T) \] factorizes wrt edges of \( T \) as well
(Hyper-Markov wrt \( T \))

Averaging over \( T \):
This should hold for any tree \( T \)

Compatible family of strong Markov hyper-dist. [DL93]:
\[ \to \text{ multinomial-Dirichlet (conjugacy)}, \]
\[ \to \text{ normal-Wishart (conjugacy)}, \]
\[ \to \text{ Gaussian copulas (numerical integration), ...?} \]
Posterior probability of an edge

The existence of an edge between variables $Y_j$ and $Y_k$ can be assessed by

$$\Pr\{(j, k) \in T \mid Y\} \propto \sum_{T \ni (j, k)} p(T)p(Y \mid T)$$

which depends on the prior $p(T)$.

The prior probability $\Pr\{(j, k) \in T\}$ can be tuned

- with the prior coefficient $b_{jk}$
- or set to an arbitrary value using an edge-specific probability change.
Posterior probability of an edge

The existence of an edge between variables $Y_j$ and $Y_k$ can be assessed by

$$\Pr\{ (j, k) \in T \mid Y \} \propto \sum_{T \ni (j, k)} p(T)p(Y \mid T)$$

which depends on the prior $p(T)$.

The prior probability $\Pr\{ (j, k) \in T \}$ can be tuned

- with the prior coefficient $b_{jk}$
- or set to an arbitrary value using an edge-specific probability change.

All posterior probabilities can still be computed in $O(p^3)$ [Kir07].

$\rightarrow$ R package Saturnin (spanning trees used for network inference) [SRS15]
Simulations: ROC curves for edge detection
For various graph topologies ($p = 25, n = 25, 50, 200, B = 100$ simulations)
Some simulations

Tree

Erdös ($\pi = 2/p$)

Erdös ($\pi = 4/p$)

## Algebraic properties

<table>
<thead>
<tr>
<th>Bayesian inference</th>
<th>Maximum likelihood</th>
</tr>
</thead>
</table>

*Any other example?*

S. Robin (INRA / AgroParisTech)
## Algebraic properties

<table>
<thead>
<tr>
<th>Change-point detection</th>
<th>Bayesian inference</th>
<th>Maximum likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum_m \prod_{r \in m} p_r$</td>
<td>$\rightarrow$ Matrix power</td>
<td>$\rightarrow$ Dynamic programming</td>
</tr>
<tr>
<td>$\max_m \sum_{r \in m} \log p_r$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Algebraic properties

<table>
<thead>
<tr>
<th>Change-point detection</th>
<th>Bayesian inference: $\sum_m \prod_{r \in m} p_r$</th>
<th>Maximum likelihood: $\max_m \sum_{r \in m} \log p_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rightarrow$ Matrix power</td>
<td>$\rightarrow$ Dynamic programing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tree-structured network inference</th>
<th>Bayesian inference: $\sum_T \prod_{(jk) \in T} p_{jk}$</th>
<th>Maximum likelihood: $\max_T \sum_{(jk) \in T} \log p_{jk}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rightarrow$ Matrix-tree theorem</td>
<td>$\rightarrow$ Max. spanning tree</td>
</tr>
</tbody>
</table>
# Algebraic properties

<table>
<thead>
<tr>
<th>Change-point detection</th>
<th>Bayesian inference</th>
<th>Maximum likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum_m \prod_{r \in m} p_r$</td>
<td>$\rightarrow$ Matrix power</td>
<td>$\rightarrow$ Dynamic programming</td>
</tr>
<tr>
<td>Tree-structured network inference</td>
<td>$\sum_T \prod_{(jk) \in T} p_{jk}$</td>
<td>$\rightarrow$ Matrix-tree theorem</td>
</tr>
<tr>
<td>Algebra</td>
<td>sum-product</td>
<td>max-sum</td>
</tr>
</tbody>
</table>
## Algebraic properties

<table>
<thead>
<tr>
<th>Change-point detection</th>
<th>Bayesian inference</th>
<th>Maximum likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum_m \prod_{r \in m} p_r$</td>
<td>$\rightarrow$ Matrix power</td>
<td>$\rightarrow$ Dynamic programming</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tree-structured network inference</th>
<th>$\sum_T \prod_{(jk) \in T} p_{jk}$</th>
<th>$\rightarrow$ Matrix-tree theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rightarrow$ Max. spanning tree</td>
<td></td>
</tr>
</tbody>
</table>

### Algebra

- Sum-product
- Max-sum

**Any other example?**