Network Modeling in Systems Biology with R
Tutorial 1

Markov chain Monte Carlo (MCMC) sampling of Gaussian Bayesian networks for static (steady-state) data

Statistical Computing 2010, Reisensburg
21-June-2010

Miriam Lohr and Marco Grzegorczyk
Department of Statistics
TU Dortmund University

Regulatory networks and protein signalling pathways

Flow cytometry technology

Protein activation cascade

From Sachs et al., Science, 2005
Protein activation cascade

phosphorylation

nucleus

cell membrane

-> cell response

Protein activation cascade

possibly completely unknown

True regulatory pathway/network

possibly completely unknown
Statistical Task

Extract a network from a data matrix

variables $X_1, ..., X_n$
genes/proteins

Either $m$ independent (steady-state) observations of the system $X_1, ..., X_n$
Or time series of the system of length $m$: $(X_1, ..., X_n)_{t=1,...,m}$
**Static Bayesian networks**

- Marriage between graph theory and probability theory.
- Directed acyclic graph (DAG) represents conditional independence relations.
- Markov assumption leads to a factorization of the joint probability distribution:

\[
P(A, B, C, D, E, F) = P(A) \cdot P(B \mid A) \cdot P(C \mid A) \cdot P(D \mid B, C) \cdot P(E \mid D) \cdot P(F \mid C, D)
\]

**Incidence Matrix of a DAG**

Number the variables/nodes 1= A, 2= B, 3= C, etc.

\[
I = \begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

\[I(i,j)=1 \text{ if there is an edge from node } i \text{ to node } j\]
\[I(i,j)=0 \text{ if there is no edge from node } i \text{ to node } j\]

**Ancestor Matrix of a DAG**

Number the variables/nodes 1= A, 2= B, 3= C, etc.

\[
A = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0
\end{pmatrix}
\]

\[A(j,i)=1 \text{ if there is a path from node } i \text{ to node } j\]
\[A(j,i)=0 \text{ if there is no path from node } i \text{ to node } j\]

**Bayesian networks versus causal networks**

Bayesian networks represent conditional (in)dependency relations - not necessarily causal interactions.
Equivalence classes of BNs

\[ P(A) \cdot P(B \mid C) \cdot P(C \mid A) = P(A) \cdot P(B, C) \cdot P(C)^{-1} \cdot P(C, A) \cdot P(A)^{-1} \]

\[ = P(C \mid B) \cdot P(B) \cdot P(C)^{-1} \cdot P(A \mid C) \cdot P(C) \]

\[ = P(A \mid C) \cdot P(B) \cdot P(C \mid B) \]

\[ = P(A \mid C) \cdot P(B, C) \]

\[ = P(A \mid C) \cdot P(B \mid C) \cdot P(C) \]

\[ P(A, B) \neq P(A) \cdot P(B) \]

\[ P(A, B) = P(A) \cdot P(B) \]

\[ P(A, B) \neq P(A) \cdot P(B) \]

\[ P(A, B) = P(A) \cdot P(B) \]

\[ P(A, B) 
\]

CPDAG representations

\[ \text{DAG} \]

\[ \text{CPDAG} \]

\[ \text{interpretation} \]

\[ \text{superposition} \]

\[ \text{completed partially directed graphs (CPDAGs)} \]

\[ \text{P(A,B)} \neq \text{P(A)} \cdot \text{P(B)} \]

\[ \text{P(A,B)} = \text{P(A)} \cdot \text{P(B)} \]

\[ \text{P(A,B)} \neq \text{P(A)} \cdot \text{P(B)} \]

\[ \text{Static Bayesian networks} \]

\[ P(\text{graph} \mid \text{data}) = \frac{P(\text{data} \mid \text{graph}) \cdot P(\text{graph})}{P(\text{data})} \propto P(\text{data} \mid \text{graph}) \cdot P(\text{graph}) \]

\[ = P(\text{graph}) \cdot \int P(\text{data}, \theta(\text{graph}) \mid \text{graph}) d\theta(\text{graph}) \]

\[ \text{Parameterisation: Gaussian BGe scoring metric:} \]

\[ \text{data} \sim N(\mu, \Sigma) \]

\[ \mu \sim N(\mu^*, (vW)^{-1}) \text{ and } W \sim \text{Wishart}(T_0) \]

\[ \text{Static Bayesian networks} \]

\[ P(\text{graph} \mid \text{data}) = \frac{P(\text{data} \mid \text{graph}) \cdot P(\text{graph})}{P(\text{data})} \propto P(\text{data} \mid \text{graph}) \cdot P(\text{graph}) \]

\[ = P(\text{graph}) \cdot \int P(\text{data}, \theta(\text{graph}) \mid \text{graph}) d\theta(\text{graph}) \]

18

19

20
Static Bayesian networks

\[ P(\text{graph} \mid \text{data}) = \frac{P(\text{data} \mid \text{graph}) \cdot P(\text{graph})}{P(\text{data})} \propto P(\text{data} \mid \text{graph}) \cdot P(\text{graph}) \]
\[ = P(\text{graph}) \cdot \int P(\text{data}, \theta(\text{graph}) \mid \text{graph}) \, d\theta(\text{graph}) \]

**BGe metric:** closed form solution

Learning the network/graph structure

graph $\rightarrow$ score\(_{\text{BGe}}\)(graph)

<table>
<thead>
<tr>
<th>n</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>#DAGs</td>
<td>543</td>
<td>$3.7 \cdot 10^6$</td>
<td>$7.8 \cdot 10^{11}$</td>
<td>$4.2 \cdot 10^{18}$</td>
</tr>
</tbody>
</table>

**Idea:** Heuristically searching for the graph \( M^* \) that is most supported by the data \( P(M^* \mid \text{data}) > P(\text{graph} \mid \text{data}) \), e.g.: greedy search algorithm

Data are sparse $\rightarrow$ Intrinsic uncertainty of inference

Large data set D: Best network structure \( M^* \) well defined

Small data set D: Intrinsic uncertainty about \( M^* \)
**MCMC sampling of Bayesian networks**

**Better idea**: Bayesian model averaging via Markov Chain Monte Carlo (MCMC) simulations

Construct and simulate a Markov Chain \( M_t \), in the space of DAGs whose distribution converges to the graph posterior distribution as stationary distribution, i.e.:

\[
P(M_t = \text{graph} | \text{data}) \rightarrow P(\text{graph} | \text{data})
\]

\( t \rightarrow \infty \)

to generate a DAG sample: \( G_1, G_2, G_3, \ldots G_T \)

---

**Metropolis Hastings sampler**

A Metropolis Hastings MCMC sampling scheme consists of two parts.

(i) Given a graph \( G_{\text{old}} \), a new graph is proposed with a proposal probability \( Q(G_{\text{new}} | G_{\text{old}}) \).

(ii) The new graph is accepted with an acceptance probability \( A(G_{\text{new}} | G) \), or rejected otherwise.

In the structure MCMC sampling scheme a neighbour graph, that is a graph \( G_{\text{new}} \) that can be reached from \( G_{\text{old}} \) by one single edge operation, is randomly drawn from a discrete uniform distribution in the proposal move (i).

And in step (ii) the new graph is accepted with probability:

\[
A(G_{\text{new}} | G_{\text{old}}) = \min \left \{ 1, \frac{P(D | G_{\text{new}}) P(G_{\text{new}}) Q(G_{\text{old}} | G_{\text{new}})}{P(D | G_{\text{old}}) P(G_{\text{old}}) Q(G_{\text{new}} | G_{\text{old}})} \right \}
\]

---

**Structure MCMC sampling scheme**

(based on single edge operations)

1. Remove edge
2. Reverse edge
3. Add edge

---

**Metropolis Hastings sampler**

A Metropolis Hastings MCMC sampling scheme consists of two parts.

(i) Given a graph \( G_{\text{old}} \), a new graph is proposed with a proposal probability \( Q(G_{\text{new}} | G_{\text{old}}) \).

(ii) The new graph is accepted with an acceptance probability \( A(G_{\text{new}} | G) \), or rejected otherwise.

In the structure MCMC sampling scheme a neighbour graph, that is a graph \( G_{\text{new}} \) that can be reached from \( G_{\text{old}} \) by one single edge operation, is randomly drawn from a discrete uniform distribution in the proposal move (i).

And in step (ii) the new graph is accepted with probability:

\[
A(G_{\text{new}} | G_{\text{old}}) = \min \left \{ 1, \frac{P(D | G_{\text{new}}) P(G_{\text{new}}) Q(G_{\text{old}} | G_{\text{new}})}{P(D | G_{\text{old}}) P(G_{\text{old}}) Q(G_{\text{new}} | G_{\text{old}})} \right \}
\]

<table>
<thead>
<tr>
<th>Likelihood ratio</th>
<th>Prior ratio</th>
<th>Hastings ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(D</td>
<td>G_{\text{new}}) P(G_{\text{new}}) Q(G_{\text{old}}</td>
<td>G_{\text{new}}) )</td>
</tr>
</tbody>
</table>
**Metropolis Hastings sampler**

A Metropolis Hastings MCMC sampling scheme consists of two parts.

(i) Given a graph $G_{\text{old}}$, a new graph is proposed with a proposal probability $Q(G_{\text{new}}|G_{\text{old}})$.

(ii) The new graph is accepted with an acceptance probability $A(G_{\text{new}}|G_{\text{old}})$, or rejected otherwise.

In the structure MCMC sampling scheme a neighbour graph, that is a graph $G_{\text{new}}$ that can be reached from $G_{\text{old}}$ by one single edge operation, is randomly drawn from a discrete uniform distribution in the proposal move (i).

And in step (ii) the new graph is accepted with probability:

$$A(G_{\text{new}}|G_{\text{old}}) = \min \left\{ 1, \frac{\text{Score}_{BBG}(G_{\text{new}})}{\text{Score}_{BBG}(G_{\text{old}})} \frac{Q(G_{\text{old}}|G_{\text{new}})}{Q(G_{\text{new}}|G_{\text{old}})} \right\}$$

**Hastings ratio**

$Q(G_i|G_{i-1}) \neq Q(G_{i-1}|G_i)$ is possible

**Metropolis Hastings algorithm**

structure MCMC for Bayesian networks

**Initialisation:** Start from an arbitrary initial graph $G$ (e.g. the empty-seeded graph) and set $G_0 = G$.

**Iteration:** For $i=1,\ldots,T$
- Obtain a new graph $G_i$ from the proposal distribution $Q(G_i|G_{i-1})$
- Accept the new graph with probability $A(G_i|G_{i-1})$ where $A(\cdot,\cdot)$ has to be specified as described above; otherwise reject $G_{i-1}$ leave the Markov chain state unchanged; symbolically: $G_i = G_{i-1}$.

**END**

Discard an initial ‘burn-in’ period to allow the Markov chain to reach stationarity, i.e. to converge. For example discard the first $I<T$ MCMC samples.

**Output:** An MCMC sample from the posterior distribution $P(G|D)$, symbolically: $\text{MCMC sample: } G_{1,\ldots,T}$.
Marginal edge posterior probabilities

Use the DAG (CPDAG) sample for estimating the marginal posterior probability of "directed edge relation features"

\[
P(A \rightarrow B) = \frac{1}{(T-I)} \sum_{i=1}^{T} I(G_i)
\]

where \(I(G_i)\) is 1 if the CPDAG of \(G_i\) contains the directed edge \(A \rightarrow B\), and 0 otherwise

Convergence of MCMC sampling

The DAG sample \(G_1,G_2,G_3,\ldots,G_T\) is generated via Markov Chain Monte Carlo (MCMC) simulations so that the Markov Chain \((M_t)_t\) converges to the graph posterior distribution:

\[
P(M_t=\text{graph}|\text{data}) \rightarrow P(\text{graph}|\text{data})
\]

\[t \rightarrow \infty\]

In practice: \(t\) is not infinite!!!
Network reconstruction accuracy

true regulatory network

Thresholding

marginal edge posterior probabilities

data

Concrete network predictions

Sensitivity: 1/2 = 0.5
Specificity: 4/4 = 1.0

Sensitivity: 2/2 = 1.0
Specificity: 3/4 = 0.75

Receiver Operator Characteristic (ROC) curve

ROC Curve

Ideal Test
Actual Test
No Predictive Value

Sensitivity (%)

Specificity (%)

AUC scores
Area under Receiver Operator Characteristic (ROC) curve

AUC = 0.5
AUC = 1
0.5 < AUC ≤ 1

Outlook to practical application
Gold-standard RAF pathway according to Sachs et al. (2004)

Outlook

Cellular signalling cascade which consists of 11 phosphorylated proteins and phospholipids in human immune systems cell (true network known from the literature)

Generate synthetic Gaussian network data

where $\epsilon_{pip3}$ is a Gaussian with expectation $\mu=0$ and variance $\sigma^2=1$

We generate $m$ independently and identically distributed (iid) realisations for $pip3$, and we standardise the $m$ observations:

$$pip3 \leftarrow \text{zscore}(pip3) := \frac{(pip3 - \text{mean}(pip3))}{\text{std}(pip3)}$$
**Generate synthetic Gaussian network data**

Having sampled \( m \) realisations for parent node \( \text{pip} 3 \), we sample the regression coefficient \( \beta_1 \) from a uniform distribution on \([0.5,2]\) with a randomly drawn sign +/-.

And we generate \( m \) iid realisations for \( \text{plcg} \) as follows:

\[
\text{plcg} = \beta_1 \cdot \text{pip} 3 + \varepsilon_{\text{plcg}}
\]

where the noise term \( \varepsilon_{\text{plcg}} \) is a Gaussian with expectation \( \mu = 0 \) and variance \( \sigma^2 \).

Standardise the \( m \) values for \( \text{plcg} \):

\[
\text{zscore}(\text{plcg}) = \frac{(\text{plcg} - \text{mean}(\text{plcg}))}{\text{std}(\text{plcg})}
\]

We go to the next node, etc.

Each node is described as a linear combination of its parent nodes. The regression coefficients are randomly sampled, and the noise terms are Gaussian distributed. We standardise the data to avoid that the signals become stronger and stronger. The parameter \( \sigma^2 \) can be used to vary the signal-to-noise (SNR) ratio:

**E.g.:**

\[
\text{pip} 2 = \beta_2 \cdot \text{pip} 3 + \beta_3 \cdot \text{plcg} + \varepsilon_{\text{pip} 2}
\]

\[
\text{std}(\text{signal}) = \text{std}(\beta_2 \cdot \text{pip} 3 + \beta_3 \cdot \text{plcg})
\]

\[
\text{N}(0,1) \quad \text{N}(0,1)
\]
Generate synthetic Gaussian network data

**E.g.:** \( \text{pip}^2 = \beta_2 \cdot \text{pip}^3 + \beta_3 \cdot \text{plcg} + \epsilon_{\text{pip}^2} \)

\[
\text{std} \left( \text{noise} \right) = \text{std} \left( \epsilon_{\text{pip}^3} \right) = \sigma
\]

E.g.: \( \text{pip}^2 = \beta_2 \cdot \text{pip}^3 + \beta_3 \cdot \text{plcg} + \epsilon_{\text{pip}^2} \)

\[
\text{SNR} = \frac{\text{std} \left( \text{signal} \right)}{\text{std} \left( \text{noise} \right)} = \frac{\text{std} \left( \beta_2 \cdot \text{pip}^3 + \beta_3 \cdot \text{plcg} \right)}{\sigma}
\]

**Task:** We will try to infer the Raf-pathway graph topology from a synthetically generated data set