Outline

1. Introduction
2. Static Bayesian networks
3. Dynamic Bayesian networks
4. Non-stationary gene regulatory processes
   4.1 BGM
   4.2 BGM_D
   4.3 BGM_D,2
Outline

1. Introduction
2. Static Bayesian networks
3. Dynamic Bayesian networks
4. Non-stationary gene regulatory processes
   4.1 BGM
   4.2 BGM
   4.3 BGM

Cell Biology

Microarray Chips

Expressions (activities) of thousands of genes in an experimental cell can be measured with Microarray Chips.
(Gen-)Regulatory Network

- Metabolite A
- Metabolite B
- Possibly completely unknown
Statistical Task

Extract a network from a data matrix

\[
\begin{array}{c}
\text{variables} \\
X_1, \ldots, X_n \\
\text{(genes)}
\end{array}
\begin{array}{c}
\left[\begin{array}{ccc}
X_{11} & X_{12} & \cdots & X_{1m} \\
X_{21} & X_{22} & \cdots & X_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
X_{n1} & X_{n2} & \cdots & X_{nm}
\end{array}\right]
\end{array}
\rightarrow
\begin{array}{c}
\text{cells} \\
\rightarrow
\end{array}
\begin{array}{c}
\text{variables X}_1, \ldots, X_n
\end{array}
\]

Outline

1. Introduction
2. Static Bayesian networks
3. Dynamic Bayesian networks
4. Non-stationary gene regulatory processes
   4.1 BGM
   4.2 BGM_D
   4.3 BGM_{D,2}
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:

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

A is a parent node of B
B is a child node of A
The parent node set of D is the set \{B, C\}
D is a common child node of B and C.
A has no parents. That is the parent set of A is the empty set.

A has no ancestor nodes (ancestors)
B has one ancestor node: A
C has one ancestor node: A
D has three ancestor nodes: A, B, and C
E and F have four ancestor nodes each: A, B, C, and D

Equivalence classes of BNs

- \( P(A, B) = P(A) \cdot P(B) \)
- \( P(A, B \mid C) = P(A \mid C) \cdot P(B \mid C) \)

Completed partially directed graphs (CPDAGs)
**Sytematic Overview – Equivalence classes**

3 node networks

---

**DAG and its CPDAG**

Directed Acyclic Graph (DAG)

CPDAG of the DAG on the left

CPDAGs (completed partially directed acyclic graphs) possess both direct and undirected edges.

---

**Bayesian networks versus causal networks**

Causal Network  ➔  Bayesian Network

**Bayesian networks with BGe**

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

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

Bayesian networks represent conditional (in)dependency relations - **not** necessarily causal interactions.
Bayesian networks with BGe

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

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

**Parameterisation:** Gaussian BGe scoring metric:
\[ \text{data} \sim \mathcal{N}(\mu, \Sigma) \]
with normal-Wishart distribution of the (unknown) parameters, i.e.:
\[ \mu \sim \mathcal{N}(\mu^*, (vW)^{-1}) \text{ and } W = \Sigma^{-1} \text{ with } W \sim \text{Wishart}(T_0) \]

\[ X_i \sim \mathcal{N}(\mu_i + \sum_{j=1}^{n} b_{ij} (x_j - \mu_j), \sigma_i^2) \quad b_{ij} \neq 0 \iff X_j \rightarrow X_i \]

---

**Learning the network structure**
\[ \text{graph} \rightarrow \text{score(\text{graph})} \]

<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

---

Bayesian networks with BGe

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

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

E.g. uniform distribution **BGe:** closed form solution

\[ = \text{score}_{\text{BGe}}(\text{graph} \mid \text{data}) \]

---

**Learning the network structure**

**Distribution of \( P(\text{graph} \mid \text{data}) \)**

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 \((M_t)_t\) 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: \(G_1, G_2, G_3, \ldots G_T\)

---

**Structure MCMC sampling scheme**

(based on single edge operations)

---

**CPDAG representations**

Utilise the DAG (CPDAG) sample for estimating the posterior probability of edge relation features:

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

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

1. Introduction
2. Static Bayesian networks
3. Dynamic Bayesian networks
4. Non-stationary gene regulatory processes
   4.1 BGM
   4.2 $BGM_b$
   4.3 $BGM_{b,2}$

Dynamic Bayesian networks

(a) recurrent network
(b) unfolded dynamic network

Dynamic Bayesian networks for time series data

Interpretation:

No need for the acyclicity constraint!!!
**Static/dynamic Bayesian networks**

**Static Bayesian networks**
- Important feature: Network has to be acyclic
- Implied factorisation: $P(A, B) = P(B|A)P(A|B)$

**Dynamic Bayesian networks**
- Network does not have to be acyclic
- Implied factorisation: $P(A(t), B(t)|A(t-1), B(t-1)) = P(B(t)|B(t-1))P(A(t)|A(t-1), B(t-1))$ for $t=2, \ldots, m$

**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: $G_1, G_2, G_3, \ldots, G_T$

**Example:** 2 genes $\rightarrow$ 16 different (dynamic) network structures

Identify the best network structure
- **Ideal scenario:** Large data sets, low noise

Best network: maximum score $P(D|M)$
Uncertainty about the best network structure

Limited number of experimental replications, high noise

Sample of high-scoring networks

Feature extraction, e.g. marginal posterior probabilities of the edges

High-confident edge

High-confident non-edge

Uncertainty about edges
Direct feedback loops (self-loops)

An edge pointing from a node to the same node, here "B→B"
B is its own parent node
A self-loop describes "autocorrelation"

Example: 2 genes → 16 different (dynamic) network structures

Graphs with self-loops may be considered as invalid

Outline

1. Introduction
2. Static Bayesian networks
3. Dynamic Bayesian networks
4. Non-stationary gene regulatory processes
   4.1 BGM
   4.2 BGM_D
   4.3 BGM_D,2
Reconstruction Accuracy

Is the extracted network a good prediction of the real relationships?

Reconstruction Accuracy

Evaluation of learning performance

(biological knowledge)

(gold standard network)

Probabilistic inference

true regulatory network

data

edge posterior probabilities

Thresholding

concrete network predictions

TP:1/2 FP:0/4

TP:2/2 FP:1/4

ROC Curve

From Perry Sprawls
Evaluation of Performance

**AUC scores**

Area under Receiver Operator Characteristic (ROC) curve

- **Random predictor**: AUC = 0.5
- **Perfect predictor**: AUC = 1
- **Realistic predictor**: $0.5 < \text{AUC} < 1$

Outline

1. Introduction
2. Static Bayesian networks
3. Dynamic Bayesian networks
4. Non-stationary gene regulatory processes
   - 4.1 BGM
   - 4.2 $\text{BGM}_D$
   - 4.3 $\text{BGM}_{D,2}$