Bayesian Learning of Genetic Network Structure in the Space of Differential Equation Models

Daniel Peavoy & Ata Kaban
School of Computer Science
The University of Birmingham

Intro

• Reverse-engineering gene regulatory networks from expressions of thousand genes over time
  – Numerous approaches (biology, statistics, machine learning)
  – Difficulty lies with large number of variable & unknowns
  – Simplification of the true complexity is inevitable
  – Gain understanding of the mechanisms of the system by a more detailed description of dynamics [a feasibility study]
Graphical models (Bayes nets)

**Nodes**: random variables (genes or proteins)

**Edges**: conditional probabilities

**Overall model**: Joint density

\[
p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2|x_1, x_3)p(x_3|x_2, x_4)p(x_4|x_2, x_5)p(x_5|x_3)
\]

**Drawbacks**:
- need to specify the form of all these conditional distributions
  Gaussian? – restricted to linear relations
  Multinomial? – data discretisation problematic
- structure learning is hard
Graphical representation for our modelling

Layer 1: translation
- observed mRNA
- hidden 'proteins'

Layer 2: TF → mRNA
- observed mRNA

Nodes:
- genes (capital letters)
- proteins (lower case)

Edges: reactions, modelled as ordinary differential equations

Overall model: coupled ODEs
- unknown structure
- unknown parameters of constituent ODEs

Task: infer structure (& parameters) from data

Synthetic data = simulated from such a model, with superimposed additive noise.
Nonlinear ODE building blocks

Basic building block: Michaelis-Menten eq, from the biological literature

\[
\frac{d[A]}{dt} = V_{Aa} \frac{[a]}{K_{Aa} + [a]}
\]

\[
\frac{d[A]}{dt} = V_{Aa} \left(1 - \frac{[a]}{K_{Aa} + [a]}\right)
\]

where \([.]\) denotes ‘concentration of’

V. and K. are parameters (governing boundedness)
• By combining M-M rate equations, one can build more complex dynamics. E.g.

\[
\frac{d[A]}{dt} = \frac{V_{Aa}[a]}{K_{Aa} + [a]} + V_{Ab} \left( 1 - \frac{[b]}{K_{Ab} + [b]} \right) - h_A[A]
\]

Promoter a and inhibitor b affecting mRNA A.
Dimer formation between proteins before acting as transcription factors for the next stage of gene expression

– Promotory dimer:

\[
[ab] = \frac{V_{ab}[a][b]}{K_{ab} + [a] + [b]} ; \quad \frac{d[A]}{dt} = \frac{V_{Aab}V_{ab} \frac{[a][b]}{K_{ab} + [a] + [b]}}{K_{Aab} + \frac{V_{ab}[a][b]}{K_{ab} + [a] + [b]}} = \frac{V_{Aab}[a][b]}{K_{Aab}K_{ab} + K_{Aab}[a] + K_{Aab}[b] + [a][b]}
\]

\(V_{ab}\) gets absorbed in the other parameters
– Inhibitory dimer:

\[
\frac{d[A]}{dt} = V_{Aab} \left( 1 - \frac{V_{ab} [a][b]}{K_{aabb} + [a][b]} \right)
\]

\[
= V_{Aab} \left( \frac{K_{Aab}K_{ab} + K_{Aab}[a] + K_{Aab}[b]}{K_{Aab}K_{ab} + K_{Aab}[a] + K_{Aab}[b] + [a][b]} \right)
\]

\(V_{ab}\) gets absorbed in the other parameters.

Response curve from inhibitory dimer \(ab\).
– Promotor inhibited by dimerisation

\[
\frac{d[A]}{dt} = \frac{V_{Aa}[a] \left( 1 - \frac{[b]}{K_{ab}+[a]+[b]} \right)}{K_{Aa} + [a] \left( 1 - \frac{[b]}{K_{ab}+[a]+[b]} \right)}
\]

\[
= \frac{V_{Aab}[a](K_{ab}+[a])}{K_{Aab}K_{ab} + K_{Aab}[a] + K_{Aab}[b] + K_{ab}[a] + [a]^2}
\]

Response curve from formation of dimer \( ab \) inhibiting protein \( a \).
Summing up the ‘affector’ types considered for our inference of model-combination

<table>
<thead>
<tr>
<th>Num</th>
<th>Affecto r type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No influence</td>
</tr>
<tr>
<td>1</td>
<td>Promotor</td>
</tr>
<tr>
<td>2</td>
<td>Inhibitor</td>
</tr>
<tr>
<td>3</td>
<td>Promotory dimer</td>
</tr>
<tr>
<td>4</td>
<td>Inhibitory dimer</td>
</tr>
<tr>
<td>5</td>
<td>Promotor inhibited by dimerisation</td>
</tr>
<tr>
<td>6</td>
<td>The reverse of 5</td>
</tr>
</tbody>
</table>

- In layer 1 (mRNA $\rightarrow$ protein): 1.
- In layer 2 (protein $\rightarrow$ mRNA): 0-6.
Bayesian framework for model inference

- Conditional data likelihood (given a model and parameters):
  \[ p(D | M, \theta) = \prod_{i=1}^{\text{#measurements}} \prod_{j=1}^{\text{#genes}} N(D_{ij} | \text{simulation}(M, \theta, \text{time}_i), \sigma^2) \]

- Parameters \( \theta = \{\text{init conditions, params of the ODEs}\} \)
- Generated noisy data from a model with 9 genes & 11 proteins, for validating the proposed inference procedure
- Defined a model space for search/inference, with 9 genes & 15 proteins, and pre-defined that at most 4 proteins are allowed to react with a gene.
- Inference of the model (structure)
  - asserted ‘complexity prior’ on models
  - Metropolis-Hastings sampling to generate new candidate models
- Parameter inference needed to evaluate candidate models acceptance probability
  - Used Gamma(1,3) prior on all parameters
  - Metropolis sampling to obtain parameter posteriors
  - Posterior Harmonic Mean Estimator to approximate the marginal likelihood
Model sampling

- Proposal distribution $p(M_2|M_1)$ for MH sampling of models = Uniform over models situated in the ‘neighbourhood’ of $M_1$

  \[ \alpha(M_2, M_1) = \frac{p(M_2|D) p(M_1|M_2)}{p(M_1|D) p(M_2|M_1)} \]

- $M_2$ accepted with probability:

  \[
  \begin{align*}
  0 - 1 \\
  0 - 2 \\
  \vdots \\
  2 - 0 \\
  2 - 1 \\
  0 - 3 \\
  \vdots \\
  0 - 6 \\
  \vdots \\
  2 - 6 \\
  3 - 4 \\
  \vdots \\
  5 - 6 \\
  3 - 3 \\
  5 - 5 \\
  3 - 0 \\
  4 - 1 \\
  \end{align*}
  \]

  - Changes to single edges
  - Single edges changing to dimer forming pair with a non-influencing node ($\text{Num} = 0$)
  - Dimer forming pair changing it’s action (e.g going from a promotor to inhibitor)
  - Dimer node selecting new partner node.
  - Pairs splitting to become independent

[Table of possible changes defining the neighbourhood of a model]
(a) Example parent model for a network with nine genes. Ten hidden variables participate although eleven are shown.

(b) Neighbouring network generated randomly showing the addition of an extra ligand and dimer.
Evaluating a model’s acceptance probability by parameter inference

$$\alpha(M_2, M_1) = \frac{p(M_2 | D) p(M_1 | M_2)}{p(M_1 | D) p(M_2 | M_1)}$$

ratio of model posteriors

- **Bayes factor (=ratio of marginal likelihoods)**

  $$\frac{p(M_2 | D)}{p(M_1 | D)} = \frac{p(M_2)}{p(M_1)} \frac{p(D | M_2)}{p(D | M_1)}$$

- **Parameter priors:** $P(M_i) \propto \frac{1}{\#\text{edges}(M_i)}$

- **Conditional data likelihood:**
  $$p(D | M) = \prod_{i=1}^{\#\text{measurements}} \prod_{j=1}^{\#\text{species}} N(D_{ij} | \text{simulation}(M_i, \theta_i, \text{time}_i), \sigma^2)$$
  $$\sigma^2 = 0.1$$

- **Marginal likelihoods (analytically intractable integrals) estimated using Posterior Harmonic Mean Estimator (Newton & Raftery).**

- **Parameter priors:** Gamma(1,3)

- **Model priors:**
  $$p(M_2 | D) = \frac{1}{\#\text{edges}(M_2)}$$

- **Parameter posteriors:**
  $$\theta^{(i)} \sim p(\theta | D, M)$$

- **Parameter posteriors:** estimates obtained by Metropolis sampling [10 separate Markov Chains, of 40,000 samples each]

Convergence diagnostics

Potential Scale Reduction Factor (PSRF) = \( \sqrt{\frac{\text{post_var_{upperbound}}}{\text{post_var_{lowerbound}}}} \)

Simulated dynamics from a high scoring model, with its best parameter set found (continuous lines) vs. the noisy data (marker symbols).
Assessment of the posterior distribution of model samples

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>96</td>
<td>86</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>18</td>
<td>18</td>
<td>100</td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
<td>96</td>
<td>100</td>
<td>21</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>21</td>
<td>21</td>
<td>100</td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>86</td>
<td>100</td>
<td>8</td>
<td>21</td>
<td>100</td>
<td>21</td>
<td>86</td>
<td>100</td>
<td>100</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>86</td>
</tr>
<tr>
<td>5</td>
<td>86</td>
<td>96</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>86</td>
<td>97</td>
<td>100</td>
<td>97</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>96</td>
<td>98</td>
<td>86</td>
<td>21</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>21</td>
<td>100</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>100</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>86</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>83</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>100</td>
<td>4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>83</td>
<td>21</td>
<td>100</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>9</td>
<td>21</td>
<td>21</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>31</td>
<td>97</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3: Mean “prediction” of edges expressed as a percentage. The rows are the genes and columns the proteins.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4: Indication of where edges from the best model agree with the original model. The rows are the genes and columns the proteins.
Structure recovery performance (measured by ROC) vs. increasing posterior probabilities. Higher posterior models correspond to better structures found.

\[ Sens = \frac{TP}{TP + FN}; \quad Spec = \frac{TN}{TN + FP}; \quad ROC = \frac{Sens}{1 - Spec} \]
Conclusions & further work

• The presented approach performs well at learning the network structure from synthetic data
• Model posteriors are in agreement with closeness to the true model structure
• This modelling was targeted towards protein-gene interactions within a cell
• It would be of interest to apply it in other contexts too
• We used some biological prior knowledge in the form of Michaelis-Menten eqs, and the prior on models favouring fewer edges
• The model space is still huge, and inserting more biological knowledge could further refine this and make the approach computationally less demanding
• At present the simulation (50 MH chains making 500 model samples each) took 5 days on a shared cluster
Related references


