Introduction

The Bayesian network [2, 3, 4] is a very powerful tool for estimating the gene network from microarray expression profiles. The estimated network is often susceptible to statistical sampling error, and thus Imoto et al. [3, 4] evaluated the reliability of estimation by calculating the bootstrap probabilities for the edges connecting genes. The bootstrap method, however, underestimates the probability values, and it sometimes leads to false "discovery". For improving the accuracy of the bootstrap probability, we propose the application of the newly developed multiscale bootstrap [5, 6] to the gene network estimation.

Method

1. Nonlinear Bayesian Network Model

Nonparametric Heteroscedastic Regression

We consider the additive regression model:

\[ m(x; \beta) = \sum_{j=1}^{p} \beta_j x_j + \epsilon \]

where \( \epsilon \) is a smooth function from \( \mathbb{R}^p \) to \( \mathbb{R} \) and \( \beta \) is a regression parameter.

The Bootstrap Algorithm with Multiscale Bootstrap Analysis of Gene Networks Based on Multiscale Bootstrap Method

1. What is Bayesian network?
   - DAG encoding the Markov assumption.
   - The joint density can be computed by the product of the conditional densities.
   - How can we capture the nonlinear relationships between genes?
   - How can we choose the optimal graph?

2. Bootstrap and Multiscale Bootstrap Edge Intensity

We measure the intensity of the edge by the bootstrap and multiscale bootstrap method. In the multiscale bootstrap method, we generate replicates \( X_1', \ldots, X_n' \) for several \( n \) values from the original gene expression data \( X_1, \ldots, X_n \). In other words, we alter the number of arrays from \( n \) to \( n' \) in the bootstrap replication. We will take \( n' = 10000 \).

The Bootstrap method

- We take \( n' = n \).
- The bootstrap edge intensity can be written as \( B_{n'}(\tau) \).

The multiscale bootstrap method

- We calculate \( B_{n'}(\tau) \) with several \( \tau \) values by altering \( n' \).

- The multiscale bootstrap intensity is expressed as \( B_{n'}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(d_i - c_i) \)

- We estimate \( d_i \) and \( c_i \) by fitting the theoretical curve \( B_{n'}(\tau) \) to the observed \( B_{n'}(\tau) \) values calculated by the multiscale bootstrap method.

Result

We applied the proposed method to S. cerevisiae gene expression data.

- We focus on 9 genes, which are involved or putatively involved in the heat shock response.
- We took \( n = 10000 \).

Table 1: Gene pairs with high multiscale bootstrap intensities

<table>
<thead>
<tr>
<th>gene1</th>
<th>gene2</th>
<th>( B_{n'}(\tau) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSP104</td>
<td>YER057C</td>
<td>0.676</td>
</tr>
<tr>
<td>YER057C</td>
<td>YGR042W</td>
<td>0.647</td>
</tr>
<tr>
<td>YGR042W</td>
<td>YDR037W</td>
<td>0.636</td>
</tr>
<tr>
<td>YDR037W</td>
<td>YGL073W</td>
<td>0.625</td>
</tr>
</tbody>
</table>

References