Inferring cellular regulatory networks with Bayesian Model Averaging for Linear Regression (BMALR)

Xun Huang$^a$ and Zhike Zi$^{a,b,*}$

$^a$BIOS Centre for Biological Signalling Studies, University of Freiburg, 79104, Freiburg, Germany.
$^b$Present address: Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
$^*Corresponding author, E-mail: zhike.zi@molgen.mpg.de

This Supplementary Information contains the supplementary text for the derivation of Equation 6 in the main text, other technical details for calculating probability density of the data, as well as Supplementary Figures.

1. The derivation of probability density in Gaussian Bayesian network

Given a Gaussian Bayesian network, we can uniquely generate a multivariate (nonsingular) normal distribution and vice versa$^{1,2}$. Suppose we have an $n$-dimensional random vector $(X)$, which follows a multivariate normal distribution:

$$\rho(X \mid u, r) = N(X; u, r^{-1})$$

(S1)

where the parameter $u$ is the mean vector of this distribution, and the parameter $r$ is the inverse covariance matrix of this distribution. To make the model decomposable, the global parameter independence and parameter modularity should be satisfied$^3$. Therefore, the parameter $u$ is assumed to follow a multivariate normal distribution, and the parameter $r$ is assumed to follow a Wishart distribution$^1$. Then, the prior distributions of $u$ and $r$ are given below:

$$\rho(u \mid r) = N(u; \mu, (v \cdot r)^{-1})$$

(S2)

$$\rho(r) = \text{Wishart}(r; \alpha, \beta)$$

(S3)

where the hyperparameter $\mu$ is the mean vector of the distribution of $u$. The hyperparameter $v$ is a positive number. The hyperparameter $\alpha$ is the degree of freedom of the Wishart distribution, and $\alpha$ should be greater than $(n-1)$. The hyperparameter $\beta$ is an $n$-by-$n$ scale matrix, and should be positive-definite.

If a dataset $D$ is generated by the random vector $X$, which satisfies the distribution mentioned above, there is a close-formed expression of $\rho(D_W)$ that is the probability density of the data restricted to the subset $W$ of the variables in $X$. In addition, we use the parameter $M$ to stand for the sample size of $D$, $l_W$ for the number of elements in $W$. Then, the probability density of $D_W$ will satisfy the following equation$^{1,2}$:

$$\rho(D_W) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{v}{v+M}\right)^{\frac{\alpha}{2}} \left(c(l_W, \alpha)\right)^{\frac{\nu}{2}} \left(\beta_{\nu, \alpha+M}\right)^{\frac{\nu^2}{2 \alpha}} \left|_{\alpha+M}\right|^\frac{\nu}{2}$$

(S4)

where

$$c(\nu, \alpha) = 2^{\nu/2} \pi^{-(n+1)/4} \prod_{i=1}^{n} \Gamma(\frac{\alpha+i-1}{2})^{-1}$$

(S5)

$$\beta_{\nu, \alpha+M} = \left((\beta^{-1})^{(\nu)}\right)^{-1}$$

(S6)
\[ \beta'_w = \beta_w + s_w + \left( \frac{nM}{n + M} \right) (\bar{X}_w - \mu^{(w)}) (\bar{X}_w - \mu^{(w)})^T \]  
(S7)

\[ s_w = \sum_{i=1}^{n} (X_{w,i} - \bar{X}_w) (X_{w,i} - \bar{X}_w)^T \]  
(S8)

\[ \alpha_w = \alpha - n + l_w \]  
(S9)

2. Default parameter settings for probability density in Gaussian Bayesian network

To calculate the probability density of data \( \rho(D_w) \) with the Equation S5 listed above, we need to specify four parameters \( \mu, \nu, \alpha, \) and \( \beta \), which are the hyperparameters shown in Equation S3 and Equation S4. The parameter \( \mu \) is the mean vector of the parameter \( u \) in Equation S2. In BMALR, we use the mean vector of the real sample \( (X) \) as the default value for the parameter \( \mu \). The parameter \( \nu \) is the size of the hypothetical sample upon which we base our prior belief concerning the value of \( u^{1-2} \). Therefore, we use the sample size \( M \) as the default value for the parameter \( \alpha \). The parameter \( \beta \) can be regarded as covariance matrix in the hypothetical sample upon which we based on our prior belief of \( u \). It is an \( n \)-by-\( n \) positive definite matrix. We assume the prior probability of all the edges is uniformly distributed for simplification. With such a flat prior, \( \beta \) will be a diagonal matrix. Therefore, we set the default value of \( \beta \) as \( H \), where \( H \) is the diagonal matrix of the covariance matrix of the sample data.

In summary, the parameter settings for calculating probability density of the data, \( \rho(D_w) \), is listed as below:

\[ \mu = \bar{X} \]  
(S10)

\[ \nu = M \]  
(S11)

\[ \alpha = n \]  
(S12)

\[ \beta = H \]  
(S13)

where \( H \) is a \( n \)-by-\( n \) diagonal matrix, and \( H(i,j) = \text{variance}(X_i) \), for \( i = 1, 2, \ldots, n \).

With the default parameter settings in Equation S10-S13, Equation S4 can reformulate as below:

\[ \rho(D_w) = \left( \frac{1}{2\pi} \right)^{\frac{M}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \]  
(S14)

where \( E^{(i)} \) is a \( l_w \)-by-\( l_w \) identity matrix.

The likelihood that the local structure \( G_{Pa} \) generates the data \( D_{Pa,X} \), \( P(D_{Pa,X}|G_{Pa}) \) can be rewritten as:

\[ P(D_{Pa,X}|G_{Pa}) = \frac{\rho(D_{Pa,X})}{\rho(D_{Pa})} \]

\[ = \left( \frac{1}{2\pi} \right)^{\frac{M}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \]  
(S15)

To further increase the efficiency, we make an approximation: \( \frac{c(l_{Pa,X},l_{Pa,X})c(l_{Pa,X},l_{Pa,X})}{c(l_{Pa,X},l_{Pa,X})c(l_{Pa,X},l_{Pa,X})} = 1 \), therefore,

\[ P(D_{Pa,X}|G_{Pa}) = \left( \frac{1}{2\pi} \right)^{\frac{M}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2} \right)^{\frac{1}{2}} \]  
(S16)

The posterior probability of an edge feature \( f \) is computed with
\[ P(f | D) = \sum_{p \in S \cap x} f(G_{p, x}) P(G_{p, x} | D_{p, x}) \]
\[ = \sum_{p \in S \cap x} f(G_{p, x}) P(D_{p, x} | G_{p, x}) P(G_{p, x}) \]
\[ = \sum_{p \in S \cap x} P(D_{p, x} | G_{p, x}) P(G_{p, x}) \]  
\[ (S17) \]

It is worth noting that replacing \( P(D_{p, x} | G_{p, x}) \) in equation S17 with equation S16 leads to the cancellation of the item \( \left( \frac{1}{2\pi} \right)^M \left( \frac{1}{2} \right)^M \left( H^{(X)} \right)^{\frac{M}{2}} \). Therefore, the item \( \left( \frac{1}{2\pi} \right)^M \left( \frac{1}{2} \right)^M \left( H^{(X)} \right)^{\frac{M}{2}} \) can be ignored in the implementation of the algorithm. Therefore, we made the following simplification for calculating \( \rho(D_W) \):

\[ \rho(D_w) \propto \left( \frac{1}{2\pi} \right)^{\frac{M}{2}} \left( \frac{1}{2} \right)^{\frac{M}{2}} \left( \frac{H^{(W)}}{2} \right)^{\frac{M}{2}} \left( E^{(W)} + M \cdot R^{(W)} \right)^{\frac{1}{2}} \]  
\[ \propto E^{(W)} + M \cdot R^{(W)} \]  
\[ (S18) \]

We noticed that when the parameter \( M \) is too large (e.g. > 100), loss of precision often appears in the calculation of equation S18. In practice, one can set \( M \) with a value between 20 and 100 to avoid this problem and get an approximate value for equation S18 in case that the sample size is too large. In BMALR, we set \( M = 50 \).

3. References

4. Supplementary Figures

**Supplementary Figure 1**  PR curves for DREAM4 *in silico* Size100 Multifactorial Challenge. BMALR* denotes the results of BMALR with the log transformation of the datasets. (A) For gold standard network 1 (NET1). (B) For gold standard network 2 (NET2). (C) For gold standard network 3 (NET3). (D) For gold standard network 4 (NET4). (E) For gold standard network 5 (NET5).
**Supplementary Figure 2** ROC curves for DREAM4 in silico Size100 Multifactorial Challenge. BMALR\(^*\) denotes the results of BMALR with the log transformation of the datasets. (A) For gold standard network 1 (NET1). (B) For gold standard network 2 (NET2). (C) For gold standard network 3 (NET3). (D) For gold standard network 4 (NET4). (E) For gold standard network 5 (NET5).
Supplementary Figure 3 Similarity of the inference methods clustered by their predicted networks on the DREAM5 network inference sub-challenge.
**Supplementary Figure 4** The performance (area under precision-recall curve, AUPR) of individual methods and community methods with combinations of every two individual methods on DREAM4 *in silico* size 100 multifactorial sub-challenge. (A) For gold standard network 2 (NET2). (B) For gold standard network 3 (NET3). (C) For gold standard network 4 (NET4). (D) For gold standard network 5 (NET5). The dashed horizontal line denotes the highest performance level in all the methods.
Supplementary Tables

**Supplementary Table 1** The influence of \textit{maxFanIn} on the performance of BMALR in DREAM4 \textit{in silico} size 100 multifactorial sub-challenge.

<table>
<thead>
<tr>
<th>MaxFanIn</th>
<th>NET1 AUPR</th>
<th>NET1 AUROC</th>
<th>NET2 AUPR</th>
<th>NET2 AUROC</th>
<th>NET3 AUPR</th>
<th>NET3 AUROC</th>
<th>NET4 AUPR</th>
<th>NET4 AUROC</th>
<th>NET5 AUPR</th>
<th>NET5 AUROC</th>
<th>Score</th>
<th>Time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.155</td>
<td>0.745</td>
<td>0.166</td>
<td>0.737</td>
<td>0.231</td>
<td>0.792</td>
<td>0.234</td>
<td>0.808</td>
<td>0.214</td>
<td>0.778</td>
<td>39.4</td>
<td>1.8E+00</td>
</tr>
<tr>
<td>3</td>
<td>0.151</td>
<td>0.742</td>
<td>0.163</td>
<td>0.735</td>
<td>0.232</td>
<td>0.798</td>
<td>0.231</td>
<td>0.806</td>
<td>0.214</td>
<td>0.790</td>
<td>39.4</td>
<td>5.7E+01</td>
</tr>
<tr>
<td>4</td>
<td>0.148</td>
<td>0.743</td>
<td>0.159</td>
<td>0.732</td>
<td>0.230</td>
<td>0.791</td>
<td>0.227</td>
<td>0.796</td>
<td>0.212</td>
<td>0.793</td>
<td>38.6</td>
<td>1.8E+03</td>
</tr>
</tbody>
</table>

**Supplementary Table 2** The influence of \textit{maxFanIn} on the performance of BMALR in DREAM5 network inference sub-challenge with \textit{in silico} datasets.

<table>
<thead>
<tr>
<th>maxFanIn</th>
<th>AUPR</th>
<th>AUROC</th>
<th>Optimal F-score</th>
<th>Time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.320</td>
<td>0.809</td>
<td>0.372</td>
<td>1.3E+02</td>
</tr>
<tr>
<td>3</td>
<td>0.321</td>
<td>0.809</td>
<td>0.374</td>
<td>8.3E+03</td>
</tr>
</tbody>
</table>

**Supplementary Table 3** The influence of \textit{maxFanIn} on the performance of BMALR in the benchmark of the T-cell signaling network.

<table>
<thead>
<tr>
<th>Method</th>
<th>AUPR</th>
<th>AUROC</th>
<th>Optimal F-score</th>
<th>p-value of Fisher’s exact test</th>
<th>Time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.343</td>
<td>0.720</td>
<td>0.500</td>
<td>2.70E-04</td>
<td>1.3E-01</td>
</tr>
<tr>
<td>3</td>
<td>0.345</td>
<td>0.726</td>
<td>0.500</td>
<td>2.70E-04</td>
<td>1.4E-01</td>
</tr>
<tr>
<td>4</td>
<td>0.337</td>
<td>0.709</td>
<td>0.465</td>
<td>9.44E-04</td>
<td>1.6E-01</td>
</tr>
<tr>
<td>5</td>
<td>0.338</td>
<td>0.709</td>
<td>0.462</td>
<td>9.44E-04</td>
<td>2.0E-01</td>
</tr>
<tr>
<td>6</td>
<td>0.338</td>
<td>0.709</td>
<td>0.462</td>
<td>9.44E-04</td>
<td>2.2E-01</td>
</tr>
<tr>
<td>7</td>
<td>0.338</td>
<td>0.709</td>
<td>0.462</td>
<td>9.44E-04</td>
<td>2.4E-01</td>
</tr>
</tbody>
</table>