Web-based Supplementary Materials for "Bayesian Methods for Predicting Interacting Protein Pairs Using Domain Information"

Inyoung Kim[1], Yin Liu[2], and Hongyu Zhao[1,3,*]

1 Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520, USA.

2 Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA.

3 Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA.

*To whom correspondence should be addressed:

Hongyu Zhao, Ph.D

Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, New Haven, CT 06520-8034.

Tel: (203) 785-6271

Fax: (203) 785-6912

E-Mail hongyu.zhao@yale.edu

APPENDIX

A The EM Algorithm

The loglikelihood of the complete data $Y_c = (Y_{obs}, Y_{miss}) = (O, D)$ is

$$\begin{aligned} LogL_c &= \sum_{ij} D_{mn}^{(ij)}[\log(\lambda_{mn}) + O_{ij}log\{h_{ij}(\Lambda)(1-f_n) + (1-h_{ij}(\Lambda))f_p\} \\ &+ (1-O_{ij})\log\{1-(h_{ij}(\Lambda)(1-f_n) + (1-h_{ij}(\Lambda))f_p)\}] \\ &+ (1-D_{mn}^{(ij)})[\log(1-\lambda_{mn}^{(ij)}) + O_{ij}log\{h_{ij}(\Lambda)(1-f_n) + (1-h_{ij}(\Lambda))f_p\} \\ &+ (1-O_{ij})\log\{1-(h_{ij}(\Lambda)(1-f_n) + (1-h_{ij}(\Lambda))f_p)\}]. \end{aligned}$$

For given $\theta^{(t-1)}$ obtained from the $(t-1)$th EM iteration, the E-step requires computations of the following:

$$E(D_{mn}^{(ij)}|O_{l_1l_2k} = o_{ijk}, \forall l_1, l_2, \theta^{(t-1)}) = \frac{\lambda_{mn}^{(t-1)}(1-f_n)^{O_{ijk}} f_n^{1-O_{ijk}}}{Pr(O_{ijk} = o_{ijk}|\theta^{(t-1)})} = \tau_{mn}^{(ij)}(\theta^{(t-1)})$$

and

$$\begin{aligned} Q(\theta^{(t)}) &= E_{\theta^{(t)}}\{logL_c|Y_{obs}\} \\ &= \sum_{ij} \tau_{mn}^{(ij)}(\theta^{(t)})[\log(\lambda_{mn}^{(t-1)}) + O_{ij}log\{h_{ij}(\Lambda^{(t)})(1-f_n) + (1-h_{ij}(\Lambda^{(t)}))f_p\} \\ &+ (1-O_{ij})\log\{1-\Big(h_{ij}(\Lambda^{(t)})(1-f_n) + (1-h_{ij}(\Lambda^{(t)}))f_p\Big)\}] \\ &+ (1-\tau_{mn}^{(ij)}(\theta^{(t-1)}))[\log(1-\lambda_{mn}^{(t)}) + O_{ij}log\{h_{ij}(\Lambda^{(t-1)})(1-f_n) + (1-h_{ij}(\Lambda^{(t)}))f_p\} \\ &+ (1-O_{ij})\log\{1-\Big(h_{ij}(\Lambda^{(t)})(1-f_n) + (1-h_{ij}(\Lambda^{(t)}))f_p\Big)\}]. \end{aligned}$$

The M-step consists of estimating λ_{mn} using equation (3) and estimating f_n and f_p by maximizing $Q(\theta^{(t)})$. For estimating the MLEs of f_n and f_p, we calculate the second derivatives of $Q(\theta^{(t)})$ with respect to f_n and f_p as the following:

$$\begin{aligned} \frac{\partial^2 Q(\theta^{(t)})}{\partial f_n^2} &= \sum[\frac{-O_{ij}h_{ij}(\Lambda^{(t)})^2}{(h(\Lambda^{(t)})(1-f_n) + (1-h_{ij}(\Lambda^{(t)}))f_p)^2} \\ &+ \frac{-(1-O_{ij})h_{ij}(\Lambda^{(t)})^2}{(1-h_{ij}(\Lambda^{(t)})(1-f_n) - (1-h_{ij}(\Lambda^{(t)}))f_p)^2}]; \end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 Q(\theta^{(t)})}{\partial f_p^2} &= \sum[\frac{-O_{ij}(1-h_{ij}(\Lambda^{(t)}))^2}{(h_{ij}(\Lambda^{(t)})(1-f_n)+(1-h_{ij}(\Lambda^{(t)}))f_p)^2} \\
&+ \frac{(1-O_{ij})(1-h_{ij}(\Lambda^{(t)}))^2}{(1-h_{ij}(\Lambda^{(t)})(1-f_n)-(1-h_{ij}(\Lambda^{(t)}))f_p)^2}]; \\
\frac{\partial^2 Q(\theta^{(t)})}{\partial f_n \partial f_p} &= \sum[\frac{-O_{ij}h_{ij}(\Lambda^{(t)})(1-h_{ij}(\Lambda^{(t)}))}{(h(\Lambda^{(t)})(1-f_n)+(1-h_{ij}(\Lambda^{(t)}))f_p)^2} \\
&+ \frac{(1-O_{ij})h_{ij}(\Lambda^{(t)})(1-h_{ij}(\Lambda^{(t)}))}{(1-h_{ij}(\Lambda^{(t)})(1-f_n)-(1-h_{ij}(\Lambda^{(t)}))f_p)^2}].
\end{aligned}
$$

However, if $\lambda_{mn} = 0$ for all m, n, the second derivatives are zeros. Therefore, the inverses of the Hessian matrices for f_p and f_p do not exist.

B The Proofs of Properties of λ_{mn}, f_{n_k}, and f_{p_k}

B.1 (*P.1*): The property of the posterior distribution of λ_{mn}

The posterior distribution of λ_{mn} is proportional to

$$
\begin{aligned}
[\Lambda|rest] &\propto L(O|f_{n_k}, f_{p_k}, \Lambda)L(\Lambda|f_{n_k}, f_{p_k}) \\
&\propto \prod_{ijk}[h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}]^{o_{ijk}} \\
&\quad [1-\{h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}\}]^{1-o_{ijk}} \\
&\quad f(\Lambda|f_{n_k}, f_{p_k}).
\end{aligned}
$$

Except for a constant, the loglikelihood is

$$
\begin{aligned}
\log L &= \sum_{ijk} O_{ijk}\log[h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}] \\
&\quad +(1-O_{ijk})\log[1-\{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}] + \log f(\Lambda|f_{n_k}, f_{p_k}).
\end{aligned}
$$

By chain rule, we obtain the following:

$$
\begin{aligned}
\frac{\partial \log L}{\partial \lambda_{mn}} &= \frac{\partial \log L}{\partial h_{ij}(\Lambda)}\frac{\partial h_{ij}(\Lambda)}{\partial \lambda_{mn}} \\
\frac{\partial^2 \log L}{\partial \lambda_{mn}^2} &= \frac{\partial^2 \log L}{\partial h_{ij}(\Lambda)^2}(\frac{\partial h_{ij}(\Lambda)}{\partial \lambda_{mn}})^2 + \frac{\partial \log L}{\partial h_{ij}(\Lambda)}\frac{\partial^2 h_{ij}(\Lambda)}{\partial \lambda_{mn}^2}.
\end{aligned}
$$

Since

$$\begin{aligned}
\frac{\partial \log L}{\partial h_{ij}(\Lambda)} &= \sum_{ijk} \frac{O_{ijk}(1-f_{n_k}-f_{p_k})}{h_{ij}(\Lambda)(1-f_{n_k})+\{1-h_{ij}(\Lambda)\}f_{p_k}} + \frac{-(1-O_{ijk})(1-f_{n_k}-f_{p_k})}{1-\{h_{ij}(\Lambda)(1-f_{n_k})+\{1-h_{ij}(\Lambda)\}f_{p_k}\}}, \\
\frac{\partial^2 \log L}{\partial h_{ij}^2(\Lambda)} &= \sum_{ijk} \frac{-O_{ijk}(1-f_{n_k}-f_{p_k})^2}{[h_{ij}(\Lambda)(1-f_{n_k})+\{1-h_{ij}(\Lambda)\}f_{p_k}]^2} + \frac{-(1-O_{ijk})(1-f_{n_k}-f_{p_k})^2}{[1-\{h_{ij}(\Lambda)(1-f_{n_k})+\{1-h_{ij}(\Lambda)\}f_{p_k}\}]^2} < 0,
\end{aligned}$$

we obtain

$$\frac{\partial^2 \log L}{\partial h_{ij}(\Lambda)^2}(\frac{\partial h_{ij}(\Lambda)}{\partial \lambda_{mn}})^2 < 0.$$

Therefore the logconcavity depends on the following formula

$$\frac{\partial \log L}{\partial h_{ij}(\Lambda)} \frac{\partial^2 h_{ij}(\Lambda)}{\partial \lambda_{mn}^2}. \tag{B.1}$$

B.1.1 Case 1: $a = 1$

Since $h_{ij}(\Lambda) = 1 - \prod_{(D_{mn}^{(ij)} \in P_{ijk})}(1-\lambda_{mn})$, we have

$$\begin{aligned}
h_{ij}^{'}(\Lambda) &= \frac{\partial h_{ij}(\Lambda)}{\partial \lambda_{mn}} = - \prod_{(D_{m'n'}^{(ij)} \in P_{ijk}),(m'n') \neq (mn)} (1-\lambda_{m'n'}); \\
h_{ij}^{''}(\Lambda) &= \frac{\partial^2 h_{ij}(\Lambda)}{\partial \lambda_{mn}^2} = 0.
\end{aligned}$$

Hence, equation (B.1) is zero. Therefore the posterior distribution of λ_{mn} is a logconcave function.

B.1.2 Case 2: $a \neq 1$

Since $h_{ij}(\Lambda) = 1 - \prod_{(D_{mn}^{(ij)} \in P_{ijk})}(1-\lambda_{mn}^a)$, we have

$$\begin{aligned}
h_{ij}^{'}(\Lambda) &= \frac{\partial h_{ij}(\Lambda)}{\partial \lambda_{mn}} = \prod_{(D_{m'n'} \in P_{ijk}),(m'n') \neq (mn)} (1-\lambda_{m'n'}^a) a \lambda_{mn}^{a-1}; \\
h_{ij}^{''}(\Lambda) &= \frac{\partial^2 h_{ij}(\Lambda)}{\partial \lambda_{mn}^2} = \prod_{(D_{m'n'}^{(ij)} \in P_{ijk}),(m'n') \neq (mn)} (1-\lambda_{m'n'}^a) a(a-1) \lambda_{mn}^{a-2} > 0.
\end{aligned}$$

Since we have

$$\frac{\partial \log L}{\partial h_{ij}} = \sum \frac{[O_{ijk} - \{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}][1-f_{n_k}-f_{p_k}]}{[h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}][1-\{h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}\}]},$$

and $E(O_{ijk}) = h_{ij}(\Lambda)(1-f_{n_k})+\{1-h_{ij}(\Lambda)\}f_{p_k}$, $E(\frac{\partial \log L}{\partial h_{ij}(\Lambda)}) = 0$. In addition, $\lim_{\lambda_{mn}\to 0} h^a_{ij}(\Lambda) = h^1_{ij}(\Lambda)$ and $\lim_{\lambda_{mn}\to 1} h^a_{ij}(\Lambda) = h^1_{ij}(\Lambda)$. Therefore, the posterior distribution of λ_{mn} is not a logconcave function but the behavior of mean and tails is the same as that of $h^1_{ij}(\Lambda)$.

B.2 (*P.2*): The property of the posterior distribution of f_{n_k}

The posterior distribution of f_{n_k} is proportional to

$$\begin{aligned}
[f_{n_k}|rest] &\propto L(O|f_{n_k}, f_{p_k}, \Lambda)L(f_{n_k}|\Lambda, f_{p_k})L(f_{p_k}|\Lambda) \\
&\propto \prod_{(P^k_{ij})} [h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}]^{o_{ijk}} \\
&\quad [1-\{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}]^{1-o_{ijk}} \\
&\quad f(f_{n_k}|\Lambda, f_{p_k}).
\end{aligned}$$

Except for a constant, the loglikelihood is

$$\begin{aligned}
\log L &= \sum_{ijk} O_{ijk} \log[h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}] \\
&\quad +(1-O_{ijk}) \log[1-\{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}] + \log f(f_{n_k}|\Lambda, f_{p_k}).
\end{aligned}$$

The first and second derivatives of the posterior distribution with respect to f_{n_k} are

$$\begin{aligned}
\frac{\partial \log L}{\partial f_{n_k}} &= \sum_{ijk} \frac{-O_{ijk}h_{ij}(\Lambda)}{h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}} + \frac{(1-O_{ijk})h_{ij}(\Lambda)}{1-\{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}}, \\
\frac{\partial^2 \log L}{\partial f^2_{n_k}} &= \sum_{ijk} \frac{-O_{ijk}h_{ij}(\Lambda)^2}{[h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}]^2} + \frac{-(1-O_{ijk})h_{ij}(\Lambda)^2}{[1-\{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}]^2} \\
&< 0.
\end{aligned}$$

Therefore, the posterior distribution of f_{n_k} is always a logconcave function regardless of $h_{ij}(\Lambda)$.

B.3 (*P.3*): The property of the posterior distribution of f_{p_k}

The posterior distribution of f_{p_k} is proportional to

$$[f_{p_k}|rest] \propto L(O|f_{n_k}, f_{p_k}, \Lambda)L(f_{p_k}|\Lambda, f_{n_k})L(f_{n_k}|\Lambda)$$

$$
\begin{aligned}
\propto & \prod_{(P_{ij}^k)} [h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}]^{o_{ijk}} \\
& [1 - \{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}]^{1-o_{ijk}} \\
& f(f_{p_k}|\Lambda, f_{n_k}).
\end{aligned}
$$

Except for a constant, the loglikelihood is

$$
\begin{aligned}
\log L &= \sum_{ijk} O_{ijk} \log[h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}] \\
&\quad +(1-O_{ijk}) \log[1 - \{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}] + \log f(f_{p_k}|\Lambda, f_{n_k}).
\end{aligned}
$$

The first and second derivatives of the loglikelihood with respect to f_{p_k} are

$$
\begin{aligned}
\frac{\partial \log L}{\partial f_{p_k}} &= \sum_{ijk} \frac{-O_{ijk}(1-h_{ij}(\Lambda))}{h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}} + \frac{(1-O_{ijk})(1-h_{ij}(\Lambda))}{1 - \{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}}, \\
\frac{\partial^2 \log L}{\partial f_{p_k}^2} &= \sum_{ijk} \frac{-O_{ijk}(1-h_{ij}(\Lambda))^2}{[h_{ij}(\Lambda)(1-f_{n_k}) + \{1-h_{ij}(\Lambda)\}f_{p_k}]^2} + \frac{-(1-O_{ijk})(1-h_{ij}(\Lambda))^2}{[1 - \{h_{ij}(\Lambda)(1-f_{n_k}) + (1-h_{ij}(\Lambda))f_{p_k}\}]^2} \\
&< 0.
\end{aligned}
$$

Therefore, the posterior distribution of f_{p_k} is always a logconcave function regardless of $h_{ij}(\Lambda)$.

C Comparision of the Maximum Likelihood Estimator and the Bayes Estimator in a Special Scenario

In this section, we prove that the MSE of the Bayes estimator for DDI probability between two domains is smaller than that of MLE under (S.1)-(S.2) for a wide range of interaction probability described in Section 4.

Theorem 1

Let X_i represent the random variable for protein interaction of the ith protein pair and let Z denote the interaction event of domain pair in the ith protein pair. Assume that N protein pairs contain this domain pair. Let λ be the probability of $Z = 1$. Then $X_1, X_2, ..., X_N$

are iid Bernoulli(λ) and let the prior distribution on λ be Beta (α,β). Let $\widehat{\lambda}_{MLE}$ and $\widehat{\lambda}_{BAY}$ denote the MLE and the Bayes estimator of λ, respectively. Further let $MSE_{MLE}(\lambda)$ and $MSE_{BAY}(\lambda)$ represent the MSEs of MLE and Bayes estimator of λ, respectively.

Then

(1) $lim_{N\to\infty}MSE_{BAY}(\lambda) = lim_{N\to\infty}MSE_{MLE}(\lambda) = 0$

(2) If $\lambda_L < \lambda < \lambda_U$, $MSE_{BAY}\{\lambda\} < MSE_{MLE}\{\lambda\}$,

where $\lambda_L = \frac{2N(\alpha+1)(\alpha+\beta)(\alpha+\beta)^2 - \sqrt{\{2N(\alpha+1)(\alpha+\beta)+(\alpha+\beta)^2\}^2 - 4\{(N+1)(\alpha+\beta)^2+2N(\alpha+\beta)\}(\alpha^2 N)}}{2\{(N+1)(\alpha+\beta)^2+2N(\alpha+\beta)\}}$ is a nondecreasing function of N, $\lambda_U = \frac{2N(\alpha+1)(\alpha+\beta)(\alpha+\beta)^2 + \sqrt{\{2N(\alpha+1)(\alpha+\beta)+(\alpha+\beta)^2\}^2 - 4\{(N+1)(\alpha+\beta)^2+2N(\alpha+\beta)\}(\alpha^2 N)}}{2\{(N+1)(\alpha+\beta)^2+2N(\alpha+\beta)\}}$ is a nonincreasing function of N, $lim_{N\to 0}\lambda_L = 0$, $lim_{N\to 0}\lambda_U = 1$, $lim_{N\to\infty}\lambda_L = \frac{2(\alpha+1)(\alpha+\beta) - \sqrt{4(\alpha+\beta)(\alpha+\beta+2\alpha\beta)}}{2\{(\alpha+\beta)^2+2(\alpha+\beta)\}}$ and $lim_{N\to\infty}\lambda_U = \frac{2(\alpha+1)(\alpha+\beta)+\sqrt{4(\alpha+\beta)(\alpha+\beta+2\alpha\beta)}}{2\{(\alpha+\beta)^2+2(\alpha+\beta)\}}$.

Note that if $\alpha = \beta = 1$, $lim_{N\to\infty}\lambda_L = 0.14$ and $lim_{N\to\infty}\lambda_U = 0.85$.

Proof of Theorem 1:

Define $Y = \sum_i^N X_i$. The Bayes estimator of λ is the mean of the posterior distribution,

$$\begin{aligned}\hat{\lambda}_{BAY} &= \frac{y+\alpha}{\alpha+\beta+N} \\ &= (\frac{N}{\alpha+\beta+n})(\frac{y}{N}) + (\frac{\alpha+\beta}{\alpha+\beta+N})(\frac{\alpha}{\alpha+\beta}),\end{aligned}$$

which is a linear combination of the prior mean and the sample mean, with the weights determined by α, β, and N. The MLE of λ is the sample mean of X,

$$\hat{\lambda}_{MLE} = \frac{y}{N} = \bar{X}.$$

Therefore the MSEs of $\hat{\lambda}_{BAY}$ and $\hat{\lambda}_{MLE}$ are

$$\begin{aligned}E_\lambda(\hat{\lambda}_{BAY} - \lambda)^2 &= Var_\lambda(\hat{\lambda}_{BAY}) + (Bias_\lambda\hat{\lambda}_{BAY})^2 \\ &= Var_\lambda(\frac{Y+\alpha}{\alpha+\beta+N}) + (E_\lambda(\frac{Y+\alpha}{\alpha+\beta+N}) - \lambda)^2 \\ &= \frac{N\lambda(1-\lambda)}{(\alpha+\beta+N)^2} + (\frac{N\lambda+\alpha}{\alpha+\beta+N} - \lambda)^2, \\ E_\lambda(\hat{\lambda}_{MLE} - \lambda)^2 &= Var_\lambda(\bar{X}) = \frac{\lambda(1-\lambda)}{N}.\end{aligned}$$

Hence, $lim_{N\to\infty} MSE_{BAY}(\lambda) = MSE_{MLE}(\lambda) = 0$.

Let us consider when the following quadratic inequality holds as a function of λ:

$$\begin{aligned} & MSE_{BAY}(\lambda) - MSE_{MLE}(\lambda) \\ & = (\frac{N\lambda(1-\lambda)}{(\alpha+\beta+N)^2} + (\frac{N\lambda+\alpha}{\alpha+\beta+N} - \lambda)^2 - \frac{\lambda(1-\lambda)}{N}) < 0. \end{aligned}$$

This quadratic inequality can be rewritten as

$$\{(\alpha+\beta+N)^2 - N^2 + N(\alpha+\beta)^2\}\lambda^2 + (N^2 - 2\alpha(\alpha+\beta)N - (\alpha+\beta+N)^2)\lambda + \alpha^2 N < 0.$$

Define the followings:

$$\begin{aligned} a &= (\alpha+\beta+N)^2 - N^2 + N(\alpha+\beta)^2 = (N+1)(\alpha+\beta)^2 + 2N(\alpha+\beta); \\ b &= N^2 - 2\alpha(\alpha+\beta)N - (\alpha+\beta+N)^2 = -2N(\alpha+1)(\alpha+\beta) - (\alpha+\beta)^2; \\ c &= \alpha^2 N; \\ \lambda_L &= \frac{-b - \sqrt{b^2-4ac}}{2a}; \\ \lambda_U &= \frac{-b + \sqrt{b^2-4ac}}{2a}. \end{aligned}$$

The inequality is satisfied if $\lambda_L < \lambda < \lambda_U$.

We further define the followings:

$$\begin{aligned} a' &= \frac{\partial a}{\partial N} = (\alpha+\beta)^2 + 2(\alpha+\beta); \\ b' &= \frac{\partial b}{\partial N} = -2(\alpha+1)(\alpha+\beta); \\ c' &= \frac{\partial c}{\partial N} = \alpha^2. \end{aligned}$$

Then

$$\begin{aligned} \frac{\partial \lambda_L}{\partial N} &= \frac{-b'a + ba'}{2a^2} - \frac{0.5a(b^2-4ac)^{-0.5}(2bb' - 4a'c - 4ac') - a'\sqrt{b^2-4ac}}{2a^2} \geq 0, \\ \frac{\partial \lambda_U}{\partial N} &= \frac{-b'a + ba'}{2a^2} + \frac{0.5a(b^2-4ac)^{-0.5}(2bb' - 4a'c - 4ac') - a'\sqrt{b^2-4ac}}{2a^2} \leq 0. \end{aligned}$$

Therefore, the λ_L and λ_U are nondecreasing and nonincreasing functions of N. Since the constant terms with respect to N in a, b, and c are $(\alpha+\beta)^2$, $-(\alpha+\beta)$, and 0, respectively,

$lim_{N\to 0}\lambda_U = 1$ and $lim_{N\to 0}\lambda_L = 0$. Because the coefficients of N in a, b, and c are $(\alpha+\beta)^2 + 2(\alpha+\beta)$, $-2(\alpha+1)(\alpha+\beta)$, and α^2, respectively, $lim_{N\to\infty}\lambda_U = \frac{2(\alpha+1)(\alpha+\beta)+\sqrt{4(\alpha+\beta)(\alpha+\beta+2\alpha\beta)}}{2\{(\alpha+\beta)^2+2(\alpha+\beta)\}}$ and $lim_{N\to\infty}\lambda_L = \frac{2(\alpha+1)(\alpha+\beta)-\sqrt{4(\alpha+\beta)(\alpha+\beta+2\alpha\beta)}}{2\{(\alpha+\beta)^2+2(\alpha+\beta)\}}$.

If $\alpha = 1$ and $\beta = 1$, then $a = 8N + 4$, $b = -8N - 4$, $c = N$, $a' = 12$, $b' = -8$ and $c' = 1$. Therefore, we have

$$\begin{aligned}
\lambda_L &= \frac{2N + 1 - \sqrt{2N^2 + 3N + 1}}{(4N + 2)}; \\
\lambda_U &= \frac{2N + 1 + \sqrt{2N^2 + 3N + 1}}{(4N + 2)}; \\
\frac{\partial \lambda_L}{\partial N} &= \frac{1}{\sqrt{(1 + N)(1 + 2N)(4 + 8N)}} > 0; \\
\frac{\partial \lambda_U}{\partial N} &= -\frac{1}{\sqrt{(1 + N)(1 + 2N)(4 + 8N)}} < 0.
\end{aligned}$$

We also have $lim_{N\to\infty}\lambda_L = 0.14$ and $lim_{N\to\infty}\lambda_U = 0.85$.

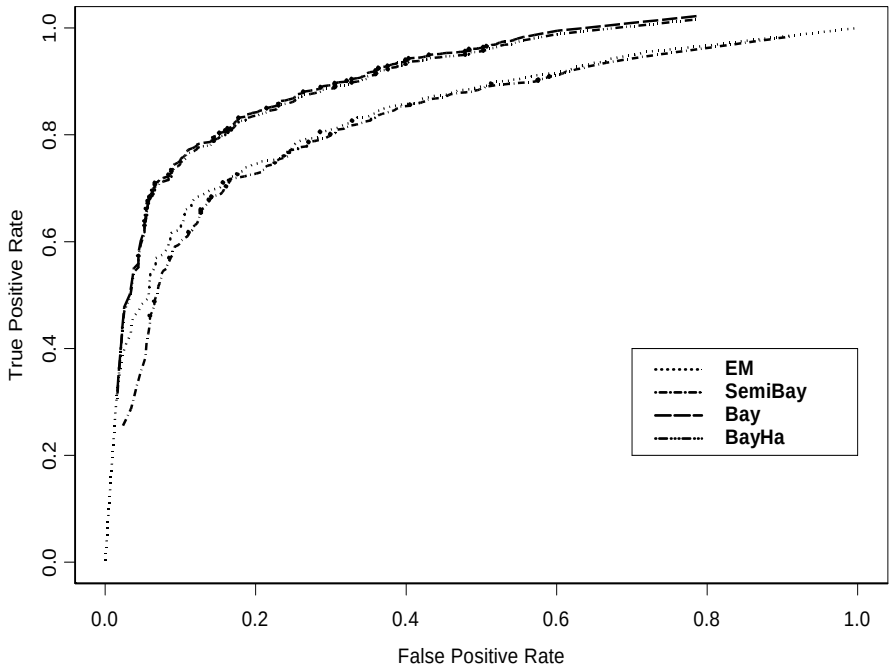

Figure 1: The average ROC curves for the likelihood based approach, the Semi-Bayesian method, and the full Bayesian methods. The number of domains is 50 and the number of proteins is 400. The distribution of DDI probability is $\lambda_{mn} \sim \text{Beta}(6, 2)$. The true PPI probabilities were generated from the model $h^1_{ij}(\Lambda) = 1 - \prod_{D^{(ij)}_{mn} \in P_{ij}}(1 - \lambda_{mn})$. EM = the likelihood based approach using true f_n and f_p, and h^1_{ij} for PPIs; SemiBay = the Semi-Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; Bay = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; BayHa = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^a_{ij} for PPIs.

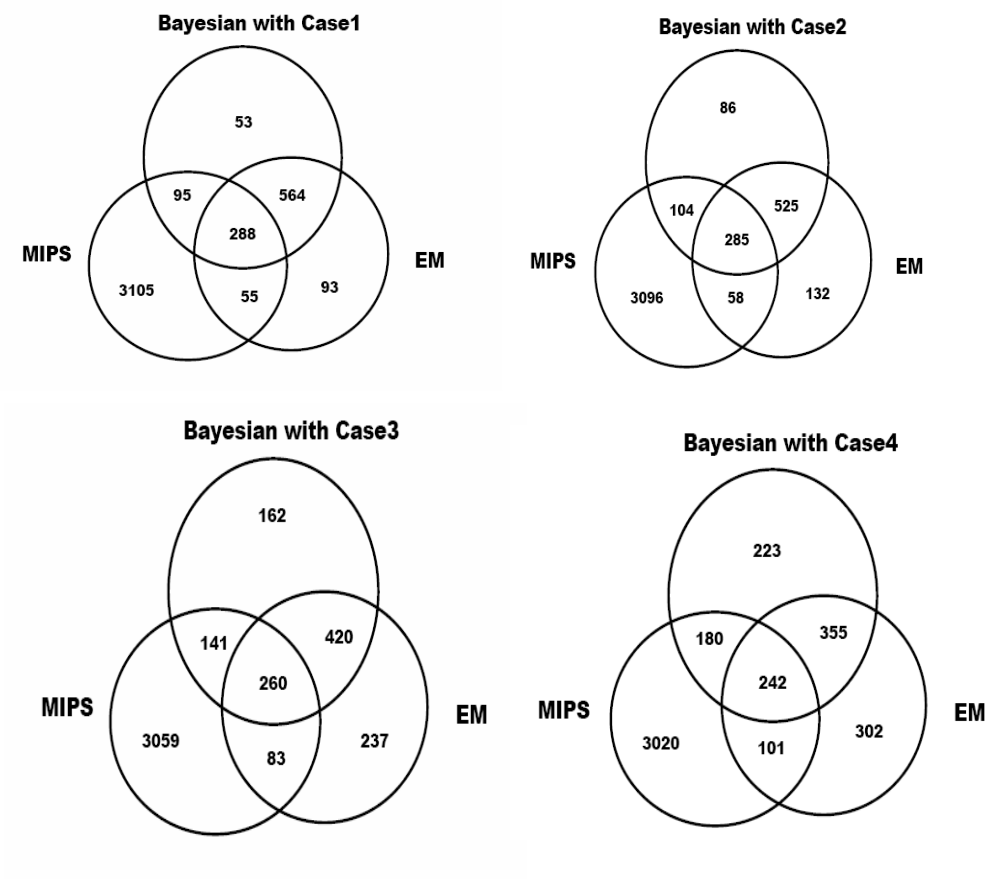

Figure 2: Overlaps among PPIs based on the MIPS dataset, the top 1,000 predicted protein pairs from the likelihood based approach, and the top 1,000 predicted protein pairs from the full Bayesian approach. MIPS = protein pairs dataset which includes experimentally verified 3,543 yeast physical interactions; EM = the likelihood based approach using $f_n = 0.8$ and $f_p = 0.0003$; Bayesian = the full Bayesian method with the estimated $\hat{f}_{n_k}$ and $\hat{f}_{p_k}$.

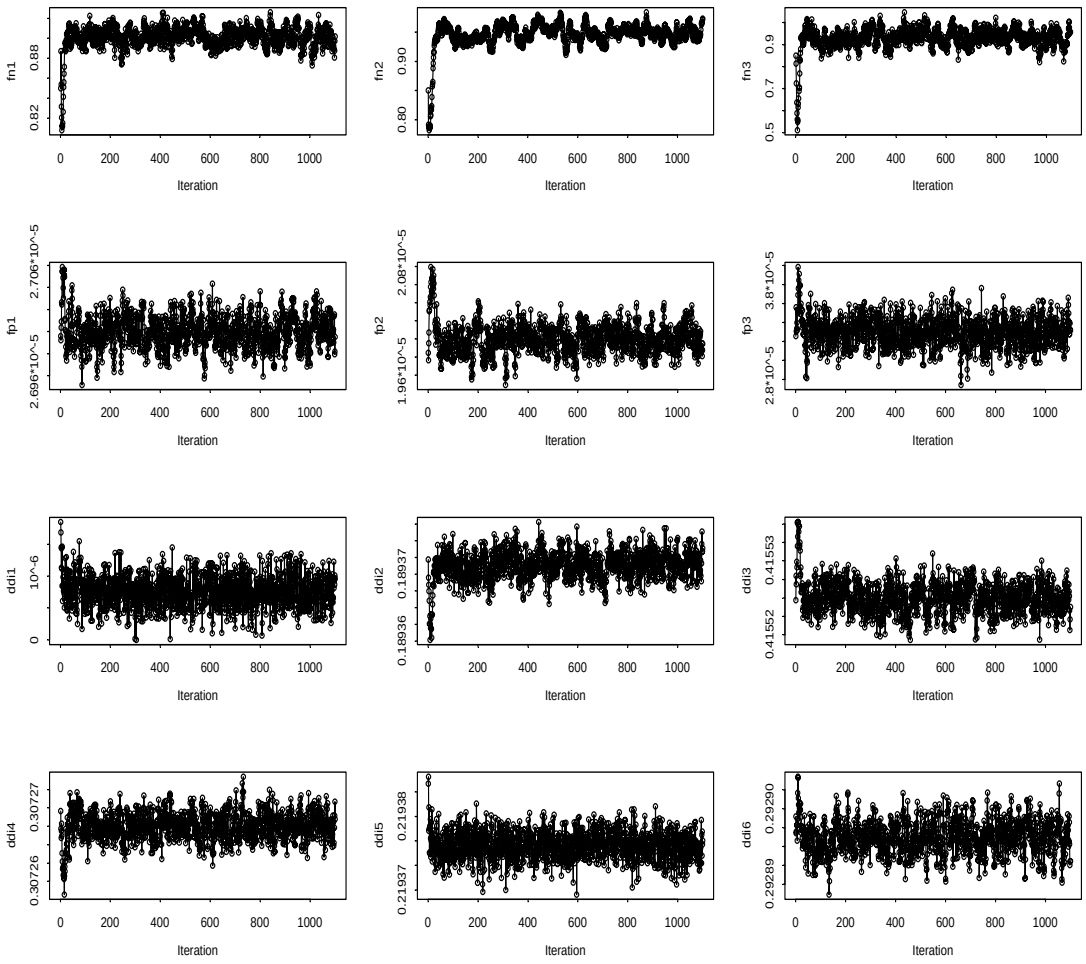

Figure 3: MCMC trace plots of the sampled false negative rates and false positive rates of three organisms, and MCMC trace plots of the samples of six randomly chosen DDI probabilities in the full Bayesian method with case 4.

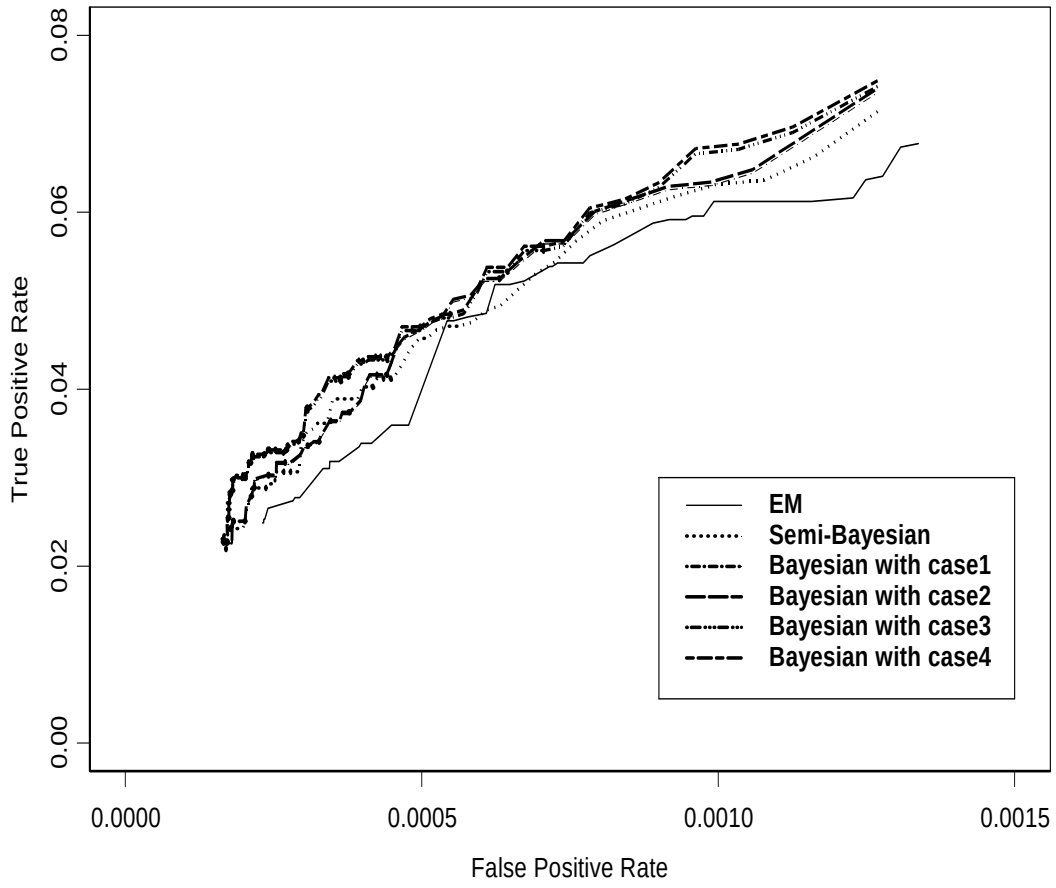

Figure 4: ROC curves for the likelihood based approach, the Semi-Bayesian method, and the full Bayesian methods. We use iPfam domain interactions as the gold standard. EM = the likelihood based approach using $f_n = 0.8$ and $f_p = 0.0003$; Semi-Bayesian = the semi-Bayesian method using the estimated f_n and f_p; Bayesian = the full Bayesian method with the estimated organism specific $\hat{f}_{n_k}$ and $\hat{f}_{p_k}$.

Table 1: The average mean square error values of DDI and PPI probabilities for each method. In this simulation, the true DDI probability (λ_{mn}) was generated from Beta$(6, 2)$. The true PPI probabilities were generated from the model $h^1_{ij}(\Lambda) = 1 - \prod_{D^{(ij)}_{mn} \in P_{ij}}(1 - \lambda_{mn})$. EM = the likelihood based approach using true f_n, f_p, and h^1_{ij} for PPIs; SemiBay = the Semi-Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; Bay = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; BayHa = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^a_{ij} for PPIs.

Num of	Num of			Domain	Level			Protein	Level	
domains	proteins		EM	Semi-Bay	Bay	BayHa	EM	Semi-Bay	Bay	BayHa
50	200	MSE	0.327	0.347	0.166	0.168	0.215	0.260	0.107	0.108
		Var	0.130	0.147	0.062	0.063	0.108	0.134	0.057	0.056
		Bias^2	0.197	0.200	0.104	0.105	0.107	0.126	0.050	0.052
	400	MSE	0.167	0.172	0.094	0.095	0.116	0.117	0.073	0.074
		Var	0.078	0.075	0.041	0.041	0.074	0.072	0.044	0.044
		Bias^2	0.089	0.097	0.053	0.054	0.042	0.045	0.029	0.030
100	200	MSE	0.539	0.561	0.269	0.271	0.330	0.340	0.120	0.121
		Var	0.179	0.170	0.096	0.097	0.169	0.176	0.069	0.068
		Bias^2	0.360	0.391	0.173	0.174	0.161	0.164	0.051	0.053
	400	MSE	0.328	0.345	0.166	0.169	0.215	0.261	0.108	0.108
		Var	0.131	0.144	0.061	0.063	0.108	0.134	0.057	0.055
		Bias^2	0.197	0.201	0.105	0.106	0.107	0.127	0.051	0.053

Table 2: The average mean square error values of f_n and f_p for each method. In this simulation, the true DDI probability (λ_{mn}) was generated from Beta$(6, 2)$. The true PPI probabilities were generated from the model $h^1_{ij}(\Lambda) = 1 - \prod_{D^{(ij)}_{mn} \in P_{ij}}(1 - \lambda_{mn})$. SemiBay = the Semi-Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; Bay = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; BayHa = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^a_{ij} for PPIs.

Num of domains	Num of proteins		SemiBay		Bay		BayHa	
			f_n	f_p	f_n	f_p	f_n	f_p
50	200	Mean	0.758	0.00042	0.767	0.00044	0.769	0.00042
		MSE	1.78e-3	2.18e-7	1.06e-3	1.91e-7	1.02e-3	1.89e-7
		Var	1.79e-5	2.04e-7	2.29e-5	1.72e-7	2.21e-5	1.73e-7
		Bias^2	1.76e-3	1.36e-8	1.04e-3	1.91e-8	1.0019e-3	1.60e-8
50	400	Mean	0.783	0.00032	0.788	0.00033	0.789	0.00031
		MSE	2.74e-4	1.58e-7	1.25e-4	1.39e-7	1.07e-3	1.23e-7
		Var	5.21e-6	1.48e-7	5.30e-6	1.28e-7	5.14e-5	1.13e-7
		Bias^2	2.69e-4	1.01e-8	1.20e-4	1.15e-8	1.02e-3	1.00e-8

Table 3: The average mean square error value of DDI probabilities for each method. In this simulation, the true DDI probability (λ_{mn}) was generated from $r_1\text{Beta}(2, 2 \times 10^7) + (1 - r_1)\text{Beta}(6, 2)$, where $r_1 = 0.5, 0.6, ..., 0.9$. The true PPI probabilities were generated from the model $h^1_{ij}(\Lambda) = 1 - \prod_{D^{(ij)}_{mn} \in P_{ij}}(1 - \lambda_{mn})$. SemiBay = the Semi-Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; Bay = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs

Dist. of λ_{mn}	Num of domains	Num of proteins		Domain EM	Level SemiBay	Bay
$0.5B(2, 2 \times 10^7)$			MSE	0.161	0.173	0.100
$+0.5B(6, 2)$	50	200	Var	0.069	0.070	0.049
			Bias^2	0.092	0.103	0.051
$0.6B(2, 2 \times 10^7)$			MSE	0.112	0.118	0.071
$+0.4B(6, 2)$	50	200	Var	0.050	0.053	0.031
			Bias^2	0.062	0.065	0.040
$0.7B(2, 2 \times 10^7)$			MSE	0.076	0.076	0.051
$+0.3B(6, 2)$	50	200	Var	0.033	0.029	0.023
			Bias^2	0.043	0.047	0.028
$0.8B(2, 2 \times 10^7)$			MSE	0.039	0.040	0.026
$+0.2B(6, 2)$	50	200	Var	0.012	0.013	0.008
			Bias^2	0.027	0.027	0.018
$0.9B(2, 2 \times 10^7)$			MSE	0.022	0.024	0.014
$+0.1B(6, 2)$	50	200	Var	0.010	0.010	0.007
			Bias^2	0.012	0.014	0.007

Table 4: The average mean square error values of DDI and PPI probabilities for each method. In this simulation, the true DDI probability (λ_{mn}) was generated from Beta$(6, 2)$ and 0.9Beta$(2, 2 \times 10^7) + 0.1$Beta$(6, 2)$, respectively. The true PPI probabilities were generated from the model $h^a_{ij}(\Lambda) = 1 - \prod_{D^{(ij)}_{mn} \in P_{ij}}(1 - \lambda^a_{mn})$, $a = \frac{\log\{1-(1/2)^{\frac{1}{M_{ij}}}\}}{\log(1/2)}$. The number of domains and the number of proteins are 50 and 400, respectively. EM = the likelihood based approach using true f_n, f_p, and using h^a_{ij} for PPIs; SemiBay = the Semi-Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; Bay = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; BayHa = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^a_{ij} for PPIs.

			Domain Level				Protein Level		
Dist. of λ_{mn}		EM	Semi-Bay	Bay	BayHa	EM	Semi-Bay	Bay	BayHa
Beta$(6, 2)$	MSE	0.396	0.380	0.198	0.103	0.236	0.224	0.158	0.087
	Var	0.090	0.075	0.048	0.045	0.086	0.075	0.083	0.044
	Bias^2	0.306	0.305	0.150	0.058	0.150	0.149	0.075	0.043
$0.9B(2, 2 \times 10^7)$	MSE	0.009	0.009	0.006	0.003	0.169	0.163	0.101	0.054
$+0.1B(6, 2)$	Var	0.002	0.002	0.002	0.001	0.152	0.147	0.091	0.048
	Bias^2	0.007	0.007	0.004	0.002	0.017	0.016	0.010	0.006

Table 5: The average mean square error values of f_n and f_p for each method. In this simulation, the true DDI probability (λ_{mn}) was generated from Beta$(6,2)$ and $0.9\text{Beta}(2, 2\times 10^7)+0.1\text{Beta}(6,2)$, respectively. The true PPI probabilities were generated from the model $h^a_{ij}(\Lambda) = 1-\prod_{D^{(ij)}_{mn}\in P_{ij}}(1-\lambda^a_{mn})$, $a = \frac{\log\{1-(1/2)^{\frac{1}{M_{ij}}}\}}{\log(1/2)}$. The number of domains and the number of proteins are 50 and 400, respectively. SemiBay = the Semi-Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; Bay = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^1_{ij} for PPIs; BayHa = the full Bayesian approach using the estimated $\hat{f}_n$, $\hat{f}_p$, and h^a_{ij} for PPIs.

Num of domains		SemiBay		Bay		BayHa	
		f_n	f_p	f_n	f_p	f_n	f_p
$B(6,2)$	Mean	0.773	0.00692	0.798	0.00120	0.781	0.00037
	MSE	7.07e-4	6.12e-5	1.13e-5	1.97e-6	1.49e-4	1.33e-7
	Var	8.00e-6	9.10e-6	5.12e-6	6.00e-7	4.20e-5	1.23e-7
	Bias^2	6.99e-4	5.21e-5	6.18e-6	1.37e-6	1.07e-4	1.02e-8
$0.9B(2,2\times 10^7)$	Mean	0.797	0.00002	0.8044	0.00003	0.7973	0.00031
$+0.1B(6,2)$	MSE	5.32e-5	9.29e-8	1.51e-4	8.11e-8	3.62e-4	3.05e-9
	Var	2.64e-5	1.30e-8	2.70e-5	1.40e-9	1.31e-4	1.39e-10
	Bias^2	2.68e-5	7.99e-8	1.24e-4	7.97e-8	2.31e-4	2.91e-9