# Web-based Supplementary Material for “Learning Gene Regulatory Networks from Next Generation Sequencing Data” by Jia et al.

Bochao Jia, Suwa Xu, Guanghua Xiao, Vishal Lamba, Faming Liang*

February 13, 2017

## 1 Proof of Lemma 2.1

We first work on the posterior mean of $\alpha_i$. For any $\epsilon > 0$, the mean of the full conditional posterior distribution of $\alpha_i$ is

$$
\begin{aligned}
E[\alpha_i|\theta_{ij}, \beta_i, \boldsymbol{y}_i] &= \int_0^{\epsilon/4} \alpha_i f(\alpha_i|\theta_{ij}, \beta_i, \boldsymbol{y}_i) d\alpha_i + \int_{\epsilon/4}^{\epsilon/2} \alpha_i f(\alpha_i|\theta_{ij}, \beta_i, \boldsymbol{y}_i) d\alpha_i + \int_{\epsilon/2}^{\infty} \alpha_i f(\alpha_i|\theta_{ij}, \beta_i, \boldsymbol{y}_i) d\alpha_i \\
&\leq \epsilon/4 + (I_1) + (I_2).
\end{aligned}
$$

It is easy to see that $(I_1) \leq \epsilon/2$. To evaluate $(I_2)$, we rewrite $f(\alpha_i|\theta_{ij}, \beta_i, \boldsymbol{y}_i) = g(\alpha_i)e^{-b_1\alpha_i}$, where $g(\alpha_i)$ is an integrable function. Let $m = \min_{\alpha_i \in [\epsilon/4, \epsilon/2]} g(\alpha_i)$ and $M = \max_{\alpha_i \in [\epsilon/2, \infty)} g(\alpha_i)$,

*To whom the correspondence should be made: Faming Liang is Professor, Department of Biostatistics, University of Florida, Gainesville, FL 32611; Email: faliang@ufl.edu. B. Jia and S. Xu are Graduate Students, Department of Biostatistics, University of Florida, Gainesville, FL 32611. G. Xiao is Associate Professor, Department of Clinical Sciences, University of Texas, Southwestern Medical Center, Dallas, TX 75390. V. Lamba is Assistant Professor, Department of Pharmacotherapy and Translational Research, University of Florida, Gainesville, FL 32610.

which are known to take finite values. Then

$$\frac{I_2}{I_1} \leq \frac{M}{m} \frac{\int_{\epsilon/2}^{\infty} e^{-b_1 \alpha_i} d\alpha_i}{\int_{\epsilon/4}^{\epsilon/2} e^{-b_1 \alpha_i} d\alpha_i} = \frac{M}{m} \frac{1}{e^{b_1 \epsilon/4} - 1} \to 0,$$

as $b_1 \to \infty$. Therefore, $E[\alpha_i | \theta_{ij}, \beta_i, \boldsymbol{y}_i] \to 0$ if $b_1 \to \infty$.

Since $\beta_i | \alpha_i, \theta_{ij}, \boldsymbol{y}_i$ follows $Gamma(n\alpha_i + a_2, \sum_{j=1}^n \theta_{ij} + b_2)$, we have $E[\beta_i | \alpha_i, \theta_{ij}, \boldsymbol{y}_i] \to 0$ as $b_2 \to \infty$. With the same argument, we have

$$E[\theta_{ij} | \alpha_i, \beta_i, \boldsymbol{y}_i] = (y_{ij} + \alpha_i)/(\beta_i + 1) \to y_{ij},$$

as $b_1 \to \infty$ and $b_2 \to \infty$. By the law of iterated expectations, we have $E[\theta_{ij} | \boldsymbol{y}_i] \to y_{ij}$ as $b_1 \to \infty$ and $b_2 \to \infty$.

# 2 Proof of Lemma 2.2

## 2.1 Existing theory of adaptive MCMC

Since the prior hyperparameters are changing with iterations, the resulting posterior distribution is also changing with iterations. Hence, the proposed sampling algorithm falls into the class of adaptive MCMC algorithms. For this type of adaptive MCMC algorithms for which the target distribution changes with iterations, the ergodicity theory has been developed in Fort et al. (2011) and Liang et al. (2016). Here we adopted the theory developed by Liang et al. (2016).

To facilitate our study, we first define some notations for adaptive Markov chains. Consider a state space $(\mathbb{X}, \mathcal{F})$, where $\mathcal{F} = \mathcal{B}(\mathbb{X})$ denotes the Borel set defined on $\mathbb{X}$. Let $X_t \in \mathbb{X}$ denote the state of the Markov chain at iteration $t$, and let $P_{\gamma_t}$ denote the transition kernel at iteration $t$, where $\gamma_t$ is a realization of a $\mathbb{Y}$-valued random variable $\Gamma_t$. In simulations, $\gamma_t$ is updated according to a specified rule. Let $\mathcal{G}_t = \sigma(X_0, \ldots, X_t, \Gamma_0, \ldots, \Gamma_t)$ be the filtration generated by $\{(X_i, \Gamma_i)\}_{i=0}^t$. Thus,

$$P(X_{t+1} \in B | X_t = x, \Gamma_t = \gamma, \mathcal{G}_{t-1}) = P_\gamma(x, B), \quad x \in \mathbb{X}, \gamma \in \mathbb{Y}, B \in \mathcal{F}.$$

Let $P_\gamma^t(x, B) = P_\gamma(X_t \in B | X_0 = x)$ denote the $t$-step transition probability for the Markov chain with the fixed transition kernel $P_\gamma$ and the initial condition $X_0 = x$. Let $P^t((x, \gamma), B) = P(X_t \in B | X_0 = x, \Gamma_0 = \gamma)$, $B \in \mathcal{F}$, denote the $t$-step transition probability for the adaptive Markov chain with the initial conditions $X_0 = x$ and $\Gamma_0 = \gamma$. Let

$$T(x, \gamma, t) = \|P^t((x, \gamma), \cdot) - \pi(\cdot)\| = \sup_{B \in \mathcal{F}} |P^t((x, \gamma), B) - \pi(B)|$$

denote the total variation distance between the distribution of the adaptive Markov chain at time $t$ and the target distribution $\pi(\cdot)$. It is said the adaptive Markov chain ergodic if $\lim_{t\to\infty} T(x, \gamma, t) = 0$ for all $x \in \mathbb{X}$ and $\gamma \in \mathbb{Y}$.

For the proposed algorithm, since $\Gamma_t = (b_1^{(t)}, b_2^{(t)})$ takes values in a deterministic sequence, the ergodicity theory developed in Liang et al. (2016) can be re-stated as follows:

**Theorem 2.1** *(Ergodicity; Liang et al., 2016) Consider an adaptive Markov chain defined on the state space $(\mathbb{X}, \mathcal{F})$ with the adaption index $\Gamma_t \in \mathbb{Y}$. The adaptive Markov chain is ergodic if the following conditions are satisfied:*

*(a) (Stationarity) There exists a stationary distribution $\pi_{\gamma_t}(\cdot)$ for each transition kernel $P_{\gamma_t}$, where $\gamma_t$ denotes a realization of the random variable $\Gamma_t$.*

*(b) (Asymptotic Simultaneous Uniform Ergodicity) For any $\epsilon > 0$, there exist constants $K(\epsilon) > 0$ and $N(\epsilon) > 0$ such that*

$$\sup_{x \in \mathbb{X}} \|P_{\Gamma_t}^n(x, \cdot) - \pi(\cdot)\| \leqslant \epsilon,$$

*for all $t > K(\epsilon)$ and $n > N(\epsilon)$.*

*(c) (Diminishing Adaptation)* $\lim_{t\to 0} D_t = 0$ *in probability, where*

$$D_t = \sup_{x\in\mathbb{X}} \|P_{\Gamma_{t+1}}(x,\cdot) - P_{\Gamma_t}(x,\cdot)\|.$$

**Theorem 2.2** *(Weak Law of Large Numbers; Liang et al., 2016) Consider an adaptive Markov chain defined on the state space* $(\mathbb{X},\mathcal{F})$*. Suppose that conditions (a), (b) and (c) of Theorem* **??** *hold. Let* $\lambda(\cdot)$ *be a bounded measurable function. Then*

$$\frac{1}{n}\sum_{t=1}^{n}\lambda(X_t) \to \pi(\lambda), \quad \textit{in probability},$$

*as* $n\to\infty$*, where* $\pi(\lambda) = \int_{\mathbb{X}}\lambda(x)\pi(dx)$*.*

## 2.2 Proof of Lemma 2.2

Since the law of $\beta_i^{(t)}$ and the law of $\theta_{ij}^{(t)}$ are completely determined by the law of $\alpha_i^{(t)}$, where the supscript $t$ indicates the iteration number, our analysis concentrates on the convergence of $\alpha_i^{(t)}$. For notational simplicity, we rewrite $b_1^{(t)}$ as $\gamma_t$ and rewrite $f(\alpha_i|\theta_{ij},\beta_i,y_{ij})$ as $f_{\gamma_t}(x)$ in what follows. For the proposed algorithm, $\gamma_t$ takes values in a deterministic and monotone sequence as specified in Equation (5) of the main text.

Since the MH algorithm was used for simulating from $f_{\gamma_t}(x)$, the condition (a) holds. As shown below, for the proposed algorithm, the posterior distribution $\pi(\cdot)$ converges to a Dirac delta measure. Hence, following from Theorem **??**, the posterior mean can be obtained by setting $\lambda(x)$ to a truncated function: $\lambda(x) = x$ if $|x| < M$ and $M$ otherwise, provided that $M$ is large enough such that the interval $[-M, M]$ covers all $y_{ij}$'s. In summary, to prove Lemma 2, it suffices to verify the conditions (b) and (c).

**Verification of condition (c)** Write the target density function as

$$f_{\gamma_t}(x) = g(x)e^{-\gamma_t x},$$

where $\gamma_t$ is the adaptive parameter taking the form

$$\gamma_t = \gamma_{t-1} + \frac{c}{t^\zeta}, \quad t = 1, 2, \ldots, \qquad (*)$$

for some constants $\gamma_0 > 0$, $c > 0$ and $\zeta \in (0, 1]$. Let $q(x, y) = q(|y - x|)$ denote a random-walk proposal distribution. Define

$$s_\gamma(x, y) = q(x, y) \min \left\{ 1, \frac{g(y)e^{-\gamma y}}{g(x)e^{-\gamma x}} \frac{q(y, x)}{q(x, y)} \right\},$$

and $r_\gamma(x, y) = s_\gamma(x, y)/q(x, y)$. Then, for any Borel set $B$, the transition kernel

$$P_\gamma(x, B) = \int_B s_\gamma(x, y)dy + I(x \in B) \left[ 1 - \int_{\mathcal{X}} s_\gamma(x, z)dz \right].$$

For the derivative $ds_\gamma(x, y)/d\gamma$, we have

$$|ds_\gamma(x, y)/d\gamma| = |q(x, y)I(r_\gamma(x, y) < 1)r_\gamma(x, y)(y - x)| \leq q(x, y)|y - x|.$$

By the mean-value theorem, there exists a constant $c_1$ such that

$$|s_\gamma(x, y) - s_{\gamma'}(x, y)| \leq c_1 q(x, y)|y - x||\gamma' - \gamma|,$$

which implies that there exists a constant $c_2$ such that $\int_{\mathcal{X}} |s_\gamma(x, y) - s_{\gamma'}(x, y)|dy \leq c_2|\gamma' - \gamma|$, as the proposal is a random walk proposal. Therefore,

$$|P_\gamma(x, B) - P_{\gamma'}(x, B)| \leq 2c_2|\gamma - \gamma'|,$$

and,

$$D_t = \sup_{x \in \mathcal{X}} |P_{\gamma_{t+1}}(x, \cdot) - P_{\gamma_t}(x, \cdot)| \leq 2\frac{c_2 c_0}{(t+1)^\zeta} \to 0,$$

as $t \to \infty$.

**Verification of condition (b)** Let $P(x, B)$ denote a degenerated MH transition kernel for the Dirac delta measure $\pi(x) = \delta(x = 0)$, i.e.,

$$P(x, B) = \begin{cases} 1, & \text{if } 0 \in B, \\ 0, & \text{otherwise,} \end{cases}$$

Then it is easy to see that $\sup_x \|P_{\gamma_t}(x,B) - P(x,B)\| \to 0$ as $t \to \infty$.

For any $k \geq 1$ and any $\psi : \mathcal{X} \to [-1,1]$, we have

$$P_{\gamma_t}^k \psi(x_0) - \pi(\psi) = S_1(k) + S_2(k),$$

where $\pi(\psi) = \int \psi(x)\pi(x)dx$, and

$$S_1(k) = P^k\psi(x_0) - \pi(\psi), \qquad S_2(k) = P_{\gamma_t}^k\psi(x_0) - P^k\psi(x_0).$$

Since $P(x,B)$ is degenerated, we have $S_1(k) = 0$ for all $k \geq 1$. For the term $S_2(k)$, we can further decompose it as follows: For any $k_0$ ($1 \leq k_0 < k$),

$$\begin{aligned}
|S_2(k)| &\leq |P_{\gamma_t}^k\psi(x_0) - P_{\gamma_t}^{k_0}\psi(x_0)| + |P_{\gamma_t}^{k_0}\psi(x_0) - P^{k_0}\psi(x_0)| + |P^{k_0}\psi(x_0) - P^k\psi(x_0)| \\
&= \left|\sum_{m=0}^{k_0-1}[P^m P_{\gamma_t}^{k_0-m}\psi(x_0) - P^{m+1}P_{\gamma_t}^{k_0-(m+1)}\psi(x_0)]\right| + |P_{\gamma_t}^k\psi(x_0) - P_{\gamma_t}^{k_0}\psi(x_0)| + |P^k\psi(x_0) - P^{k_0}\psi(x_0)| \\
&= \left|\sum_{m=0}^{k_0-1} P^m(P_{\gamma_t} - P)P_{\gamma_t}^{k_0-(m+1)}\psi(x_0)\right| + |P_{\gamma_t}^k\psi(x_0) - P_{\gamma_t}^{k_0}\psi(x_0)| + |P^k\psi(x_0) - P^{k_0}\psi(x_0)|.
\end{aligned}$$

Since $\sup_x \|P_{\gamma_t}(x,B) - P(x,B)\| \to 0$ as $t \to \infty$, for any $\epsilon > 0$, there exist some $L(\epsilon)$ such that for any $t > L(\epsilon)$,

$$\begin{aligned}
|S_2(k)| &\leq 4k_0\epsilon + |P_{\gamma_t}^k\psi(\theta_0) - P_{\gamma_t}^{k_0}\psi(\theta_0)| + |P^k\psi(\theta_0) - P^{k_0}\psi(\theta_0)| \\
&= 4k_0\epsilon + S_3(t,k,k_0) + S_4(k,k_0).
\end{aligned}$$

Since $P(x,B)$ is degenerated, we have $S_4(k,k_0) = 0$ for any $k > 0$ and $k_0 > 0$. As shown in (*), $\{\gamma_t\}$ forms a monotone and deterministic sequence. With such a deterministic sequence, $P_{\gamma_t}$ converges faster and faster as $t \to \infty$. Hence, there exists some $K(\epsilon)$ and $L'(\epsilon)$ such that for any $k > k_0 \geq K(\epsilon)$, $t \geq L'(\epsilon)$,

$$S_3(t,k,k_0) \leq \epsilon.$$

Let $\tilde{L}(\epsilon) = \max\{L(\epsilon), L'(\epsilon))\}$. Furthermore, one can choose $K(\epsilon)$ such that $\epsilon K(\epsilon) \to 0$ as $\epsilon \to 0$.

Setting $\epsilon = \varepsilon/(4K(\epsilon)+1)$ and summarizing the results of $S_1(k)$ and $S_2(k)$, we conclude the following: for any $\epsilon > 0$ and any $x_0 \in \mathcal{X}$, there exists $\tilde{L}(\epsilon) \in \mathbb{N}$ and $K(\epsilon) \in \mathbb{N}$ such that for any $t > \tilde{L}(\epsilon)$ and $k > K(\epsilon)$,

$$\|P_{\gamma_t}^k(\theta_0, \cdot) - \pi(\cdot)\| \le \varepsilon.$$

Note that $\varepsilon = (4K(\epsilon)+1)\epsilon \to 0$ as $\epsilon \to 0$. Condition (b) is verified.

# 3 Definition of Precision and Recall

The precision and recall are defined by

$$\text{precision} = \frac{TP}{TP+FP}, \qquad \text{recall} = \frac{TP}{TP+FN},$$

where $TP$, $FP$ and $FN$ denote true positives, false positives and false negatives, respectively, and they are defined via a binary decision table (Table **??**).

Table 1: Outcomes of binary decision.

| | $A_{ij} = 1$ | $A_{ij} = 0$ |
|---|---|---|
| $\hat{A}_{ij} = 1$ | True Positive(TP) | False Positive(FP) |
| $\hat{A}_{ij} = 1$ | False Negative(FN) | True Negative(TN) |

# 4 Some Simulation Results

Table 2: The Posterior mean and standard deviation of $\alpha_i$, $\beta_i$ and $\theta_{ij}$ for one simulated variable as described in Section 3, where $a_1 = a_2 = a$ and $b_1^{(0)} = b_2^{(0)} = b^{(0)}$.

| a | $b^{(0)}$ | $Y_{ij}$ | $\hat{\theta}_{ij}$ | $\hat{\alpha}_i$ | $\hat{\beta}_i$ | $AUC$ |
|---|---|---|---|---|---|---|
| 1 | $10^4$ | 513.37(284.47) | 513.27(284.38) | $3.01 \times 10^{-7}(2.04 \times 10^{-6})$ | $6.58 \times 10^{-6}(6.64 \times 10^{-6})$ | 0.940 |
|  | $10^6$ | 513.37(284.47) | 513.32(284.41) | $8.58 \times 10^{-7}(5.99 \times 10^{-6})$ | $9.46 \times 10^{-7}(9.52 \times 10^{-7})$ | 0.941 |
|  | $10^{10}$ | 513.37(284.47) | 513.37(284.47) | $7.47 \times 10^{-7}(5.41 \times 10^{-6})$ | $9.87 \times 10^{-11}(9.71 \times 10^{-11})$ | 0.943 |
| 0.001 | $10^4$ | 513.37(284.47) | 513.44(284.43) | $6.54 \times 10^{-7}(5.03 \times 10^{-6})$ | $1.58 \times 10^{-8}(5.10 \times 10^{-7})$ | 0.941 |
|  | $10^6$ | 513.37(284.47) | 513.51(284.45) | $3.78 \times 10^{-7}(2.15 \times 10^{-6})$ | $1.15 \times 10^{-9}(2.87 \times 10^{-8})$ | 0.941 |
|  | $10^{10}$ | 513.37(284.47) | 513.37(284.48) | $5.75 \times 10^{-7}(3.56 \times 10^{-6})$ | $6.24 \times 10^{-14}(1.72 \times 10^{-12})$ | 0.942 |

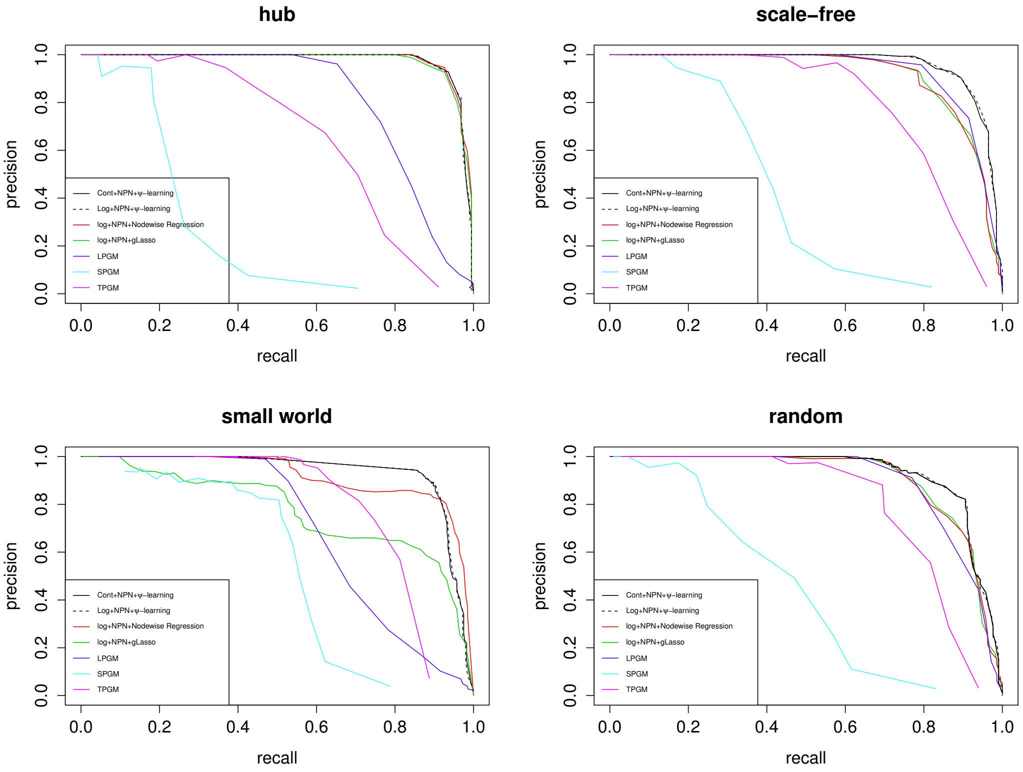

Figure 1: Precision-recall curves of each method for different type of structures with $(n, p) = (500, 200)$. Upper left: hub; upper right: scale-free; lower left: small-world; lower right: random. Refer to the legend of Figure 2 (of the main text) for the labels.

# 5 Availability of the Code and Dataset

We have attached the code for our simulation part and also the two real example datasets. These sources are available at the Biometrics website on Wiley Online Library.

## References

Fort, G., E. Moulines and P. Priouret (2011). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Annals of Statistics*, 39, 3262-3289.

Liang, F., Jin, I.H., Song, Q., and Liu, J.S. (2016). An Adaptive Exchange Algorithm for Sampling from Distribution with Intractable Normalizing Constants. *J. Amer. Statist. Assoc.*, **111**, 377-393.