

Stochastic variational variable selection for high-dimensional microbiome data

*Tung Dang*¹, *Kie Kumaishi*², *Erika Usui*², *Shungo Kobori*², *Takumi Sato*², *Yusuke Toda*¹,
*Yuji Yamasaki*³, *Hisashi Tsujimoto*³, *Yasunori Ichihashi*² and *Hiroyoshi Iwata*^{*,1}

¹ Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo,
 Tokyo, 113-8657, Japan

² RIKEN BioResource Research Center, Tsukuba, Ibaraki 305-0074, Japan

³ Arid Land Research Center, Tottori University, Tottori, 680-0001, Japan

* Correspondence: hiroiwata@g.ecc.u-tokyo.ac.jp

1 Mean-Field Variational Inference for DMM Model

We expand specifically the variational lower bound equation as follows:

$$\begin{aligned}
 \mathcal{L} = & \sum_{k=1}^{K_{max}} E_q [\log (p(\gamma_k | 1, \kappa))] - \sum_{k=1}^{K_{max}} E_q \left[\log \left(q \left(\gamma_k | \vartheta_k, \vartheta'_k \right) \right) \right] \\
 & + \sum_{i=1}^N \sum_{k=1}^{K_{max}} E_q [\log p(Z_i^k | \gamma_1, \gamma_2, \dots, \gamma_{K_{max}})] - \sum_{i=1}^N \sum_{k=1}^{K_{max}} E_q [\log (q(Z_i^k | r_i^k))] \\
 & + \sum_{i=1}^N \sum_{j=1}^S E_q [\log p(\phi_{ij} | \epsilon_{j1}, \epsilon_{j2})] - \sum_{i=1}^N \sum_{j=1}^S E_q [\log (q(\phi_{ij} | f_{ij}))] \\
 & + \sum_j^S E_q [\log (p(\epsilon_j | \xi))] - \sum_j^S E_q [\log (q(\epsilon_j | \xi^*))] \\
 & + \sum_{j=1}^S \sum_{k=1}^{K_{max}} E_q [\log (p(\alpha_{kj} | \lambda_{kj}))] - \sum_{j=1}^S \sum_{k=1}^{K_{max}} E_q \left[\log \left(q \left(\alpha_{kj} | \lambda_{kj}^* \right) \right) \right] \\
 & + \sum_{j=1}^S E_q [\log (p(\beta_j | \iota_j))] - \sum_{j=1}^S E_q [\log (q(\beta_j | \iota_j^*))]
 \end{aligned} \tag{1}$$

To compute the variational expectations $E_q[\cdot]$ in equation (1), we use the properties of exponential family distribution. If variational distributions for $q(\alpha_{kj} | \lambda_{kj})$ and $q(\beta_j | \iota_j)$ are Dirichlet distributions, then the exponential family representations are given by

$$\begin{aligned}
 q(\alpha_k | \lambda_k) &= \exp \left[\left(\sum_{j=1}^S (\lambda_j^k - 1) \log (\lambda_j^k) \right) + \log \Gamma \left(\sum_{j=1}^S \lambda_j^k \right) - \sum_{j=1}^S (\log \Gamma (\lambda_j^k)) \right] \\
 q(\beta | \iota) &= \exp \left[\left(\sum_{j=1}^S (\iota_j - 1) \log (\beta_j) \right) + \log \Gamma \left(\sum_{j=1}^S \iota_j \right) - \sum_{j=1}^S (\log \Gamma (\iota_j)) \right]
 \end{aligned}$$

So the natural parameters and sufficient statistics of the Dirichlet distributions for α and β are $\eta_{\alpha_{kj}} = \lambda_{kj} - 1$, $T(\alpha_{kj}) = \log(\alpha_{kj})$ and $\eta_{\beta_j} = \iota_j - 1$, $T(\beta_j) = \log(\beta_j)$, respectively. The approximation expectations are

$$\begin{aligned}
 E[\alpha_{kj}] &= \frac{\lambda_{kj}}{\sum_{j'=1}^S \lambda_{kj'}} & E[\beta_j] &= \frac{\iota_j}{\sum_{j'=1}^S \iota_{j'}} \\
 E[\log(\alpha_{kj})] &= \psi(\lambda_{kj}) - \psi \left(\sum_{j'=1}^S \lambda_{kj'} \right) & E[\log(\beta_j)] &= \psi(\iota_j) - \psi \left(\sum_{j'=1}^S \iota_{j'} \right)
 \end{aligned}$$

where $\psi(\cdot)$ is the digamma function.

With truncated stick-breaking representation of Dirichlet Mixture process, variational factors for the stick lengths $q(\gamma_k | \vartheta_k, \vartheta'_k)$ are Beta distributions and the assignment variable $Z_i^k = \mathbb{I}[Z_i = k]$ for the i^{th} sample allocation is governed by a multinomial distribution indexed by a variational parameter r_{ik} . The computation of the approximate expectation for the parameters of truncated stick-breaking representation have been considered

carefully by Blei and Jordan (2006). The results of variational expectation for both variables are as follows:

$$\begin{aligned} q(Z_i = k) &= r_i^k \\ q(Z_i > k) &= \sum_{k'=k+1}^{K_{\max}} r_i^{k'} \\ E_q[\log(\gamma_k)] &= \psi(\vartheta_k) - \psi(\vartheta_k + \vartheta'_k) \\ E_q[\log(1 - \gamma_k)] &= \psi(\vartheta'_k) - \psi(\vartheta_k + \vartheta'_k) \end{aligned}$$

2 Stochastic Optimization of the Variational Parameters

2.1 Compute the variational function parameters for the position of sample i to be reallocated to cluster k

Following the principles of the variational inference (Bishop. 2006, Blei and Jordan. 2006, Hoffman 2013), variational parameters of Z_i^k is the local parameters, we consider the derivation of the update equation for only one variable by fixing the others' distribution. The optimal coordinate updates are derived for local parameters. The log of the optimized factor to the posterior distribution of Z_i^k is

$$\begin{aligned} \log Q^*(Z) &= E_{\phi, \gamma, \epsilon, \alpha, \beta} [\log p(Z, \phi, \gamma, \epsilon, \alpha, \beta)] + \text{const} = \sum_{i=1}^N \sum_{k=1}^K Z_i^k \log(r_i^k) + \text{const} \\ \log r_i^k &= \sum_{j=1}^S E_Q[\phi_{ij}] \left(E_Q \left[\log \left(\frac{\Gamma(\sum_{j=1}^S \alpha_{kj})}{\Gamma(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \alpha_{kj})} \right) \right] + \sum_{j=1}^S E_Q \left[\log \left(\frac{\Gamma(X_{ij} + \alpha_{kj})}{\Gamma(\alpha_{kj})} \right) \right] + \log(J_i!) + \sum_{j=1}^S \log \left(\frac{1}{X_{ij}!} \right) \right) \\ &+ E_Q[\log(\lambda_k)] + \sum_{k'=1}^{k-1} E_Q[\log(1 - \lambda_{k'})] \end{aligned} \quad (2)$$

$$E_Q[Z_i^k] = r_i^k = \exp \{ \log(r_i^k) \}$$

The expected logarithm of the functions $E_Q \left[\log \left(\frac{\Gamma(\sum_{j=1}^S \alpha_{kj})}{\Gamma(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \alpha_{kj})} \right) \right]$ and $E_Q \left[\log \left(\frac{\Gamma(X_{ij} + \alpha_{kj})}{\Gamma(\alpha_{kj})} \right) \right]$ in equation (2) have not the closed form. Thus, the calculation of these equations are analytically intractable. To use standard form of the variational inference (Bishop. 2006) or the stochastic variational inference (Hoffman et al. 2013), we need a closed-form expression. In order to overcome this problem, we consider a common approach:

- By applying a first-order Taylor expansion to preserve a bound, intractable expectations are avoided. By using these results, the variational parameters are updated as follow:

$$\begin{aligned} \log \left(\frac{\Gamma(X_{ij} + \alpha_{kj})}{\Gamma(\alpha_{kj})} \right) &\geq \log \left(\frac{\Gamma(X_{ij} + \bar{\alpha}_{kj})}{\Gamma(\bar{\alpha}_{kj})} \right) + \frac{\partial F(\alpha_{kj})}{\partial \alpha_{kj}} \frac{\partial \alpha_{kj}}{\partial \log(\alpha_{kj})} \Big|_{\alpha_{kj} = \bar{\alpha}_{kj}} (\log(\alpha_{kj}) - \log(\bar{\alpha}_{kj})) \\ E_Q \left[\log \left(\frac{\Gamma(X_{ij} + \alpha_{kj})}{\Gamma(\alpha_{kj})} \right) \right] &\geq \log \left(\frac{\Gamma(X_{ij} + \bar{\alpha}_{kj})}{\Gamma(\bar{\alpha}_{kj})} \right) + \bar{\alpha}_{kj} [\Psi(\bar{\alpha}_{kj} + X_{ij}) - \Psi(\bar{\alpha}_{kj})] (E_Q[\log(\alpha_{kj})] - \log(\bar{\alpha}_{kj})) \\ &\geq \log \left(\frac{\Gamma(X_{ij} + \bar{\alpha}_{kj})}{\Gamma(\bar{\alpha}_{kj})} \right) + \bar{\alpha}_{kj} [\Psi(\bar{\alpha}_{kj} + X_{ij}) - \Psi(\bar{\alpha}_{kj})] (\Psi(\lambda_{kj}^*) - \Psi(\sum_{j=1}^S \lambda_{kj}^*) - \log(\bar{\alpha}_{kj})) \end{aligned}$$

$$\begin{aligned} E_Q \left[\log \left(\frac{\Gamma(\sum_{j=1}^S \alpha_{kj})}{\Gamma(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \alpha_{kj})} \right) \right] &\geq \log \left(\frac{\Gamma(\sum_{j=1}^S \bar{\alpha}_{kj})}{\Gamma(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\alpha}_{kj})} \right) \\ &+ \sum_{j=1}^S \left[\Psi \left(\sum_{j=1}^S \bar{\alpha}_{kj} \right) - \Psi \left(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\alpha}_{kj} \right) \right] \bar{\alpha}_{kj} (\Psi(\lambda_{kj}^*) - \Psi(\sum_{j=1}^S \lambda_{kj}^*) - \log(\bar{\alpha}_{kj})) \end{aligned}$$

Based on this framework, the variational parameters of Z_{ik} are updated as follow:

$$\begin{aligned} \log r_i^k &= \sum_{j=1}^S f_{ij} \log \left(\frac{\Gamma(\sum_{j=1}^S \bar{\alpha}_{kj})}{\Gamma(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\alpha}_{kj})} \right) \\ &+ \sum_{j=1}^S f_{ij} \left[\Psi \left(\sum_{j=1}^S \bar{\alpha}_{kj} \right) - \Psi \left(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\alpha}_{kj} \right) \right] \bar{\alpha}_{kj} (\Psi(\lambda_{kj}^*) - \Psi(\sum_{j=1}^S \lambda_{kj}^*) - \log(\bar{\alpha}_{kj})) \\ &+ \sum_{j=1}^S f_{ij} \left[\log \left(\frac{\Gamma(X_{ij} + \bar{\alpha}_{kj})}{\Gamma(\bar{\alpha}_{kj})} \right) + \bar{\alpha}_{kj} [\Psi(\bar{\alpha}_{kj} + X_{ij}) - \Psi(\bar{\alpha}_{kj})] (\Psi(\lambda_{kj}^*) - \Psi(\sum_{j=1}^S \lambda_{kj}^*) - \log(\bar{\alpha}_{kj})) \right] \\ &+ \Psi(\vartheta_k) - \Psi(\vartheta_k + \vartheta'_k) + \sum_{k'=1}^{k-1} \Psi(\vartheta'_{k'}) - \Psi(\vartheta_{k'} + \vartheta'_{k'}) \end{aligned} \quad (3)$$

2.2 Compute the variational function parameters of indicator variables for micro-biome selection

Similarity, the optimized solution to the posterior distribution of variable selection ϕ_{ij}

$$\log Q^*(\phi) = E_{Z, \gamma, \epsilon, \alpha, \beta} [\log p(Z, \phi, \gamma, \epsilon, \alpha, \beta)] + \text{const} = \sum_{i=1}^N \sum_{j=1}^S \phi_{ij} \log(f_{ij}) + \text{const}$$

$$\begin{aligned} \log f_{ij}^{\phi_{ij}} &= \sum_{k=1}^{K_{max}} r_{ik} \log \left(\frac{\Gamma(\sum_{j=1}^S \bar{\alpha}_{kj})}{\Gamma(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\alpha}_{kj})} \right) \\ &+ \sum_{k=1}^{K_{max}} r_{ik} \left[\Psi \left(\sum_{j=1}^S \bar{\alpha}_{kj} \right) - \Psi \left(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\alpha}_{kj} \right) \right] \bar{\alpha}_{kj} \left(\Psi(\lambda_{kj}^*) - \Psi \left(\sum_{j=1}^S \lambda_{kj}^* \right) - \log(\bar{\alpha}_{kj}) \right) \\ &+ \sum_{k=1}^{K_{max}} r_{ik} \left[\log \left(\frac{\Gamma(X_{ij} + \bar{\alpha}_{kj})}{\Gamma(\bar{\alpha}_{kj})} \right) + \bar{\alpha}_{kj} [\Psi(\bar{\alpha}_{kj} + X_{ij}) - \Psi(\bar{\alpha}_{kj})] \left(\Psi(\lambda_{kj}^*) - \Psi \left(\sum_{j=1}^S \lambda_{kj}^* \right) - \log(\bar{\alpha}_{kj}) \right) \right] \\ &+ [\Psi(\xi_1^*) - \Psi(\xi_1^* + \xi_2^*)] \end{aligned} \quad (4)$$

$$\begin{aligned} \log f_{ij}^{1-\phi_{ij}} &= \log \left(\frac{\Gamma(\sum_{j=1}^S \bar{\beta}_j)}{\Gamma(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\beta}_j)} \right) \\ &+ \sum_{j=1}^S \left[\Psi \left(\sum_{j=1}^S \bar{\beta}_j \right) - \Psi \left(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \bar{\beta}_j \right) \right] \bar{\beta}_j \left(\Psi(\iota_j^*) - \Psi \left(\sum_{j=1}^S \iota_j^* \right) - \log(\bar{\beta}_j) \right) \\ &+ \log \left(\frac{\Gamma(X_{ij} + \bar{\beta}_j)}{\Gamma(\bar{\beta}_j)} \right) + \bar{\beta}_j [\Psi(\bar{\beta}_j + X_{ij}) - \Psi(\bar{\beta}_j)] \left(\Psi(\iota_j^*) - \Psi \left(\sum_{j=1}^S \iota_j^* \right) - \log(\bar{\beta}_j) \right) \\ &+ [\Psi(\xi_2^*) - \Psi(\xi_1^* + \xi_2^*)] \end{aligned} \quad (5)$$

2.3 Updating variational parameters of α

Based on the principal framework of the stochastic variational inference (Hoffman et al. 2013), we consider parameters of α as global parameters. These parameters are updated by a stochastic gradient step and noisy estimations of the natural gradient of the variational objective with respect to α_{kj} . Following the computational method of the natural gradient (Hoffman et al. 2013), we compute the natural gradient of equation (1) with respect to the global variational parameters of α_{kj} .

We consider Dirichlet distribution as a prior distribution with concentration parameters λ_{kj} , so a conditional distribution of profile given the observation data has form of the Dirichlet distribution. We consider the representation of the exponential family,

$$p(\alpha|Z, \phi, \gamma, \epsilon, \beta) = h(\alpha) \exp \left(\eta(Z, \phi, \gamma, \epsilon, \beta)^T t(\alpha) - a(\eta(Z, \phi, \gamma, \epsilon, \beta)) \right)$$

where: $h(\cdot)$ is the base measure; $a(\cdot)$ is the log-normalize; $\eta(\cdot)$ is the natural parameter; $t(\cdot)$ is the sufficient statistics.

As above assumption of the variational parameters, we set $q(\alpha|\lambda^*)$ to be Dirichlet distribution as the complete conditional distributions. So we get

$$q(\alpha|\lambda^*) = h(\alpha) \exp \left((\lambda^*)^T t(\alpha) - a(\lambda^*) \right)$$

We consider the lower bound for only α ,

$$\begin{aligned} \mathcal{L}(\alpha) &= E_Q [\log p(\alpha|Z, \phi, \gamma, \epsilon, \beta)] - E_Q [q(\alpha|\lambda^*)] \\ &= E_Q \left[\log \{h(\alpha)\} + \eta(Z, \phi, \gamma, \epsilon, \beta)^T t(\alpha) - a(\eta(Z, \phi, \gamma, \epsilon, \beta)) - \log \{h(\alpha)\} - (\lambda^*)^T t(\alpha) + a(\lambda^*) \right] \\ &= E_Q \left[\eta(Z, \phi, \gamma, \epsilon, \beta)^T t(\alpha) - a(\eta(Z, \phi, \gamma, \epsilon, \beta)) - (\lambda^*)^T t(\alpha) + a(\lambda^*) \right] \\ &= E_Q [\eta(Z, \phi, \gamma, \epsilon, \beta)]^T [\nabla_{\lambda^*} \{a(\lambda^*)\}] - a(\eta(Z, \phi, \gamma, \epsilon, \beta)) - (\lambda^*)^T [\nabla_{\lambda^*} \{a(\lambda^*)\}] + a(\lambda^*) \end{aligned}$$

where, the expected value of the sufficient statistics is the gradient of log normalizer $E_Q [t(\alpha)] = \nabla_{\lambda^*} \{a(\lambda^*)\}$. If the classical principal of the gradient method for maximization is used directly to find a maximum of $\mathcal{L}(\alpha)$ based on taking step of size ρ in direction of the gradient, the optimized results is

$$\lambda_{(t+1)}^* = \lambda_{(t)}^* + \rho \nabla_{(\lambda^*)} \mathcal{L}(\alpha) = \lambda_{(t)}^* + \rho^{(t)} \left[\nabla_{(\lambda^*)}^2 \{a(\lambda^*)\} \right] \left\{ E_Q [\eta(Z, \phi, \gamma, \epsilon, \beta)]^T - (\lambda^*)^T \right\}$$

Following (Hoffman et al. 2013), by premultiplying the gradient by the inverse Fisher information $G(\lambda^*)$ and apply the stochastic natural gradient of the variational objective with respect to α , we get

$$\begin{aligned} G(\lambda^*) &= E_{\lambda^*} \left[(\nabla_{(\lambda^*)} \log q(\alpha|\lambda^*)) (\nabla_{(\lambda^*)} \log q(\alpha|\lambda^*))^T \right] = \nabla_{(\lambda^*)}^2 \{a(\lambda^*)\} \\ \widehat{\nabla_{(\lambda^*)} L}(\alpha) &= \{G(\lambda^*)\}^{-1} \nabla_{(\lambda^*)} L(\alpha) = \{E_Q [\eta(Z, \phi, \gamma, \epsilon, \beta)] - \lambda^*\} \end{aligned}$$

$$\begin{aligned}\lambda_{(t+1)}^* &= \lambda_{(t)}^* + \rho \widehat{\nabla_{(\lambda^*)}} L(\alpha) = \lambda_{(t)}^* + \rho \left\{ E_Q[\eta(Z, \phi, \gamma, \epsilon, \beta)] - \lambda_{(t)}^* \right\} \\ &= (1 - \rho) \lambda_{(t)}^* + \rho \{ E_Q[\eta(Z, \phi, \gamma, \epsilon, \beta)] \}\end{aligned}\quad (6)$$

Based traditional variational inference, we get the conditional distribution of α

$$\begin{aligned}\log p(\alpha|Z, \phi, \gamma, \epsilon, \beta) &= \sum_{i=1}^N \sum_{j=1}^S \sum_{k=1}^{K_{max}} \log \left(\frac{\Gamma(X_{ij} + \alpha_{kj})}{\Gamma(\alpha_{kj})} \right) \times \log \left(\frac{\Gamma(X_{ij} + \alpha_{kj})}{\Gamma(\alpha_{kj})} \right) \times E_Q[Z_{ik}] E_Q[\phi_{ij}] + \log(\alpha_{kj})(\lambda_{kj} - 1) + const \\ E_Q[\eta(Z, \phi, \gamma, \epsilon, \beta)] &= \sum_{i=1}^N E_Q[z_{ik}] E_Q[\phi_{ij}] \overline{\alpha_{kj}} \left[\Psi \left(\sum_{j=1}^S \overline{\alpha_{kj}} \right) - \Psi \left(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \overline{\alpha_{kj}} \right) + \Psi(\overline{\alpha_{kj}} + X_{ij}) - \Psi(\overline{\alpha_{kj}}) \right]\end{aligned}\quad (7)$$

By substituting equation (7) into equation (6), the final result of updated variational parameters is

$$\begin{aligned}(\lambda_{kj}^*)^{(t+1)} &= (1 - \rho^{(t)}) (\lambda_{kj}^*)^{(t)} \\ &+ \rho^{(t)} \left\{ \lambda_{kj} + \sum_{i=1}^N r_{ik} f_{ij} \overline{\alpha_{kj}} \left[\Psi \left(\sum_{j=1}^S \overline{\alpha_{kj}} \right) - \Psi \left(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \overline{\alpha_{kj}} \right) + \Psi(\overline{\alpha_{kj}} + X_{ij}) - \Psi(\overline{\alpha_{kj}}) \right] \right\}\end{aligned}$$

2.4 Updating variational parameters of β

Similarity, the stochastic natural gradient of the variational objective with respect to β

$$\begin{aligned}(\iota_j^*)^{(t+1)} &= (1 - \rho^{(t)}) (\iota_j^*)^{(t)} \\ &+ \rho^{(t)} \left\{ \iota_j + \sum_{i=1}^N [1 - f_{ij}] \overline{\beta_j} \left(\left[\Psi \left(\sum_{j=1}^S \overline{\beta_j} \right) - \Psi \left(\sum_{j=1}^S X_{ij} + \sum_{j=1}^S \overline{\beta_j} \right) \right] + [\Psi(\overline{\beta_j} + X_{ij}) - \Psi(\overline{\beta_j})] \right) \right\}\end{aligned}$$

2.5 Updating stick-breaking representation

Similarity, the optimized solution to the posterior distribution of unit length sticks γ_k

$$\begin{aligned}\log Q^*(\gamma_k) &= (1 - 1) \log(\gamma_k) + (\nu - 1) \log(1 - \gamma_k) \\ &+ \sum_{i=1}^N E_Q[z_{ik}] \log(\gamma_k) + \sum_{i=1}^N \sum_{k'=k+1}^{K_{max}} E_Q[z_{ik'}] \log(1 - \gamma_k) + const\end{aligned}\quad (8)$$

which has the logarithmic form of the Beta distribution and the corresponding variational distributions of breaking proportions $q(\gamma_k)$ are considered to be the Beta distribution. Then, the standard conditions are satisfied for a closed form coordinate update for local parameters. Based on the stochastic natural gradient, the variational parameters of γ_k are updated by computing variational expectations in equation (8) (Blei and Jordan. 2006, Hoffman et al. 2013),

$$\begin{aligned}(\vartheta_k)^{(t+1)} &= (1 - \rho^{(t)}) (\vartheta_k)^{(t)} + \rho^{(t)} \left\{ 1 + \sum_{i=1}^N r_{ik} \right\} \\ (\vartheta'_k)^{(t+1)} &= (1 - \rho^{(t)}) (\vartheta'_k)^{(t)} + \rho^{(t)} \left\{ \nu + \sum_{i=1}^N \sum_{k'=k+1}^{K_{max}} r_{ik'} \right\}\end{aligned}$$

2.6 Updating variational parameters of ϵ

Similarity, the stochastic natural gradient of the variational objective with respect to ϵ

$$\begin{aligned}(\xi_{j1}^*)^{(t+1)} &= (1 - \rho^{(t)}) (\xi_{j1}^*)^{(t)} + \rho^{(t)} \left\{ \xi_{j1} + \sum_{i=1}^N f_{ij} \right\} \\ (\xi_{j2}^*)^{(t+1)} &= (1 - \rho^{(t)}) (\xi_{j2}^*)^{(t)} + \rho^{(t)} \left\{ \xi_{j2} + \sum_{i=1}^N (1 - f_{ij}) \right\}\end{aligned}$$

3 Reference

1. Bishop, C. M. 2006. Pattern Recognition and Machine Learning. Springer, 1 edition.
2. Blei, D and Jordan, M. 2006. Variational inference for Dirichlet process mixtures. *Journal of Bayesian Analysis*, 1:121–144.
3. Gopalan, P., Hao, W., Blei, D., Storey, J. 2016. Scaling probabilistic models of genetic variation to millions of humans. *Nat Genet.* 48, 1587-1590.
4. Hoffman, M., Blei, D., Wang, C. & Paisley, J. 2013. Stochastic variational inference. *J. Mach. Learn. Res.* 14, 1303–1347.
5. Jordan, M., Ghahramani, Z., Jaakkola, T. & Saul, L. 1999. Introduction to variational methods for graphical models. *Mach. Learn.* 37, 183–233.
6. Lartillot N. 2006. Conjugate Gibbs sampling for Bayesian phylogenetic models. *J. Comput. Biol.* 13:1701–1722.
7. Robbins, H. and Monro, S. 1951. A stochastic approximation method. *The Annals of Mathematical Statistics* 22, 400–407.
8. Wainwright, M. & Jordan, M. 2008. Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* 1, 1–305.