A supplemental material for "Subset Infinite Relational Models"

Katsuhiko Ishiguro **Naonori Ueda** **Hiroshi Sawada**
NTT Communication Science Laboratories, NTT Corporation. 619-0237 Kyoto, Japan.
{ishiguro.katsuhiko , ueda.naonori , sawada.hiroshi} @lab.ntt.co.jp

Abstract

This material provides additional information about the paper "Subset Infinite Relational Models" appeared in AISTATS 2012.

1 A Gibbs solution for the one-domain SIRM

1.1 The generative model

First, we review the full description of the "one-domain" SIRM model.

$$\phi|a,b \sim \text{Beta}(a,b), \tag{1}$$
$$\theta_{k,l}|c_{k,l},d_{k,l} \sim \text{Beta}(c_{k,l},d_{k,l}), \tag{2}$$
$$\lambda_i|e,f \sim \text{Beta}(e,f), \tag{3}$$
$$r_i|\lambda_i \sim \text{Bernoulli}(\lambda_i), \tag{4}$$
$$z_i|r_i=1,\alpha \sim \text{CRP}(\alpha) \tag{5}$$
$$z_i|r_i=0 \sim \mathbb{I}(z_i=0), \tag{6}$$
$$x_{i,j}|\boldsymbol{Z},\boldsymbol{R},\{\theta\},\phi \sim \text{Bernoulli}\left(\theta_{z_i,z_j}^{r_i r_j}\phi^{1-r_i r_j}\right). \tag{7}$$

Eq. (1) defines the distribution of a relation strength for irrelevant data entries ϕ. Eq. (2) defines an relation strength from the clsuter k to cluster l for relevant data entries. $\lambda_i, i=1,2,\ldots,N$ in Eq. (3) denotes the probability of a relevancy flag variable r_i being 1. $r_i=\{0,1\}$ in Eq. (4) indicates whether the object i is relevant or not.

$z_i=k\in\{1,2,\ldots\}$ indicates the clsuter assignment of the object i. z_i is also represented as a 1-of-K type vector: i.e. if $z_i=k$, then $z_{i,k}=1$ and $z_{i,k'\neq k}=0$. The relevancy variables $\boldsymbol{R}=\{r_i\}_{i=1,\ldots,N}$ affects the remaining generative process. If $r_i=1$, then z_i is chosen based on the CRP as in Eq. (5). Otherwise ($r_i=0$), then its cluster assignments is set to $z_i=0$ with a probability 1 as in Eq. (6). $\mathbb{I}(\cdot)$ denotes that the predicate always hold with a probability 1. Finally, the observed relation $x_{i,j}, 1\leq i,j\leq N$ is conditioned by $\boldsymbol{Z}$ and $\boldsymbol{R}$. Eq. (7) is slightly tricky: if the both of items i and j is assumed as relevant objects i.e. $r_i=r_j=1$, then "relevant" relation strengths θ is used as a parameter of a Bernoulli trial. Otherwise, "irrelevant" relation strength ϕ is employed.

Appearing in Proceedings of the 15[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2012, La Palma, Canary Islands. Volume XX of JMLR: W&CP XX.

1.2 Probability distributions

$$p(\phi;a,b)=\phi^{a-1}(1-\phi)^{b-1}B^{-1}(a,b). \tag{8}$$

$$p(\boldsymbol{\Theta}=\{\theta\};c,d) = \prod_{k=1}\prod_{l=1}\theta_{k,l}^{c_{k,l}-1}(1-\theta_{k,l})^{d_{k,l}-1}B^{-1}(c_{k,l},d_{k,l}). \tag{9}$$

$$p(\boldsymbol{\Lambda}=\{\lambda\};e,f)=\prod_i \lambda_i^{e-1}(1-\lambda_i)^{f-1}B^{-1}(e,f). \tag{10}$$

$$p(\boldsymbol{R}=\{r\};\boldsymbol{\Lambda})=\prod_i \lambda_i^{r_i}(1-\lambda_i)^{1-r_i}. \tag{11}$$

When the number of the clusters is K excluding the 0th cluster,

$$p(\boldsymbol{Z}=\{z\};\boldsymbol{R},\alpha)=\alpha^K\frac{\prod_{k=1}^K(m_k-1)!}{\prod_{i=1}^M(\alpha+i-1)}, \tag{12}$$

where m_k is defined later in Eq. (16), and $M=\sum_k m_k$.

$$\begin{aligned} p(\boldsymbol{X}=\{x\};\boldsymbol{Z},\boldsymbol{R},\boldsymbol{\Theta},\phi) &= \prod_{i=1}^N\prod_{j=1}^N\left(\theta_{z_i,z_j}^{r_i r_j}\phi^{1-r_i r_j}\right)^{x_{i,j}}\left(1-\theta_{z_i,z_j}^{r_i r_j}\phi^{1-r_i r_j}\right)^{(1-x_{i,j})} \\ &= \prod_i\prod_j\prod_k\prod_l\left[\theta_{k,l}^{x_{i,j}}(1-\theta_{k,l})^{(1-x_{i,j})}\right]^{r_i z_{i,k} r_j z_{j,l}} \\ &\quad\times\prod_i\prod_j\left[\phi^{x_{i,j}}(1-\phi)^{(1-x_{i,j})}\right]^{(1-r_i r_j)}, \end{aligned} \tag{13}$$

where $z_{i,k}$ and $z_{j,l}$ are the aforementioned 1-of-K vector representations.

1.3 Sampling Hidden Variables

As described in the main article paper, simultaneous sampling of r_i and z_i leads to a simpler inference for SIRM than deriving a solution for each variable independently. Therefore we explain how to simultaneously sample r_i and z_i.

Regarding the sampling of the ith object, let us denote the current number of realized clusters by K. And we divide the observations $\boldsymbol{X}$ into two parts: data entries who relates to the object i $\boldsymbol{X}^{+i} = \{x_{i,\cdot}, x_{\cdot,i}\}$, and those who does not $\boldsymbol{X}^{\setminus i} = \{\boldsymbol{X} \setminus \boldsymbol{X}^{+i}\}$. Also we define the following quantities:

$$n_{k,l} = \sum_i \sum_j r_i z_{i,k} r_j z_{j,l} x_{i,j}, \tag{14}$$

$$\bar{n}_{k,l} = \sum_i \sum_j r_i z_{i,k} r_j z_{j,l} \left(1 - x_{i,j}\right), \tag{15}$$

$$m_k = \sum_i r_i z_{i,k}, \tag{16}$$

$$q = \sum_i \sum_j \left(1 - r_i r_j\right) x_{i,j}, \tag{17}$$

$$\bar{q} = \sum_i \sum_j \left(1 - r_i r_j\right)\left(1 - x_{i,j}\right). \tag{18}$$

The superscript $\setminus i$ denotes the above statistics computed on $\boldsymbol{X}^{\setminus i}$. Also the superscript $+i0$, $+ik$ denotes that the same statistics computed on $\boldsymbol{X}^{+i}$ assuming $r_i = 0$ or $\{r_i = 1, z_i = k\}$, respectively.

We formulate the Gibbs posterior of $\{r_i, z_i\}$ as follows:

$$\begin{aligned} p\left(z_i = k, r_i | \boldsymbol{X}, \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) &\propto p\left(z_i = k, r_i | \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) \\ &\times p\left(\boldsymbol{X}^{+i} | z_i = k, r_i, \boldsymbol{X}^{\setminus i}, \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) \end{aligned} \tag{19}$$

The first term of the right hand of Eq. (19) is easy. Multiply Eq. (11), and Eq. (12) and marginalize λ_i out thanks to the conjugacy.

$$\begin{aligned} p\left(z_i = k, r_i | \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) &\propto p\left(z_i = k | r_i, \boldsymbol{Z}^{\setminus i}\right) p\left(r_i | \boldsymbol{R}^{\setminus i}\right) \\ &= \left[p\left(z_i = k | r_i = 1, \boldsymbol{Z}^{\setminus i}\right) + p\left(z_i = k | r_i = 0\right)\right] \\ &\times \int p\left(r_i | \lambda_i\right) p\left(\lambda_i | \boldsymbol{R}^{\setminus i}\right) d\lambda_i \\ \propto &\begin{cases} f + \sum_{i' \neq i} (1 - r_{i'}) & r_i = 0, z_i = 0, \\ \left(e + \sum_{i' \neq i} r_{i'}\right) \frac{m_k^{\setminus i}}{\alpha + \sum_k m_k^{\setminus i}} & r_i = 1, z_i = k \in \{1, 2, \ldots, K\}, \\ \left(e + \sum_{i' \neq i} r_{i'}\right) \frac{\alpha}{\alpha + \sum_k m_k^{\setminus i}} & r_i = 1, z_i = K + 1. \end{cases} \end{aligned} \tag{20}$$

The second term of the right hand of Eq. (19) requires some computations. To see this, we rewrite the second term in more detailed way:

$$\begin{aligned} &p\left(\boldsymbol{X}^{+i} | z_i = k, r_i, \boldsymbol{X}^{\setminus i}, \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) \\ &= \int p\left(\boldsymbol{X}^{+i} | z_i = k, r_i, \phi, \boldsymbol{\Theta}\right) p\left(\phi, \boldsymbol{\Theta} | \boldsymbol{X}^{\setminus i}, \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) d\phi d\boldsymbol{\Theta}, \end{aligned} \tag{21}$$

where $\boldsymbol{\Theta} = \{\theta_{k,l}\}$. First, we need to compute the posterior of paramters ϕ and θ excluding the information of the $i$the object. Then we compute the marginal lieklihood of $\boldsymbol{X}^{+i}$ given z_i and r_i.

Using Eq. (9), Eq. (8) and Eq. (13), the posterior of parameteres is calculated as follows:

$$\begin{aligned} p\left(\phi, \boldsymbol{\Theta} | \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}, \boldsymbol{X}^{\setminus i}\right) &\propto p\left(\boldsymbol{X}^{\setminus i} | \phi, \boldsymbol{\Theta}, \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}\right) p\left(\phi, \boldsymbol{\Theta}\right) \\ &= \mathrm{Beta}\left(\phi; a + q^{\setminus i}, b + Q^{\setminus i}\right) \\ &\times \prod_k \prod_l \mathrm{Beta}\left(\theta_{k,l}; c_{k,l} + n_{k,l}^{\setminus i}, d_{k,l} + N_{k,l}^{\setminus i}\right) \end{aligned} \tag{22}$$

As you can see in Eq. (22), the posterior is a product of Beta distributions. Since $p\left(\boldsymbol{X}^{+i} | z_i = k, r_i, \phi, \boldsymbol{\Theta}\right)$ is a product of Bernoulli distributions (Eq. (13)), again we can use conjugacy to obtain the second term of the right hand of Eq. (19). Then we have the following euqations:

$$\begin{aligned} &p\left(\boldsymbol{X}^{+i} | z_i = 0, r_i = 0, \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) \\ &= \frac{B\left(a + q^{\setminus i} + q^{+i0}, b + \bar{q}^{\setminus i} + \bar{q}^{+i0}\right)}{B\left(a + q^{\setminus i}, b + \bar{q}^{\setminus i}\right)}, \end{aligned} \tag{23}$$

and

$$\begin{aligned} &p\left(\boldsymbol{X}^{+i} | z_i = k, r_i = 1, \boldsymbol{Z}^{\setminus i}, \boldsymbol{R}^{\setminus i}\right) \\ &= \frac{B\left(a + q^{\setminus i} + q^{+i1k}, b + \bar{q}^{\setminus i} + \bar{q}^{+i1k}\right)}{B\left(a + q^{\setminus i}, b + \bar{q}^{\setminus i}\right)} \\ &\times \frac{B\left(c_{k,k} + n_{k,k}^{\setminus i} + n_{k,k}^{+i1k}, d_{k,k} + \bar{n}_{k,k}^{\setminus i} + \bar{n}_{k,k}^{+i1k}\right)}{B\left(c_{k,k} + n_{k,k}^{\setminus i}, d_{k,k} + \bar{n}_{k,k}^{\setminus i}\right)} \\ &\times \prod_{l \neq k} \frac{B\left(c_{k,l} + n_{k,l}^{\setminus i} + n_{k,l}^{+i1k}, d_{k,l} + \bar{n}_{k,l}^{\setminus i} + \bar{n}_{k,l}^{+i1k}\right)}{B\left(c_{k,l} + n_{k,l}^{\setminus i}, d_{k,l} + \bar{n}_{k,l}^{\setminus i}\right)} \\ &\times \prod_{l \neq k} \frac{B\left(c_{l,k} + n_{l,k}^{\setminus i} + n_{l,k}^{+i1k}, d_{l,k} + \bar{n}_{l,k}^{\setminus i} + \bar{n}_{l,k}^{+i1k}\right)}{B\left(c_{l,k} + n_{l,k}^{\setminus i}, d_{l,k} + \bar{n}_{l,k}^{\setminus i}\right)}. \end{aligned} \tag{24}$$

1.4 Posteriors of parameters

$$p\left(\phi | \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}\right) = \mathrm{Beta}\left(a + q, b + \bar{q}\right) \tag{25}$$

$$p\left(\theta_{k,l} | \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{R}\right) = \mathrm{Beta}\left(c_{k,l} + n_{k,l}, d_{k,l} + \bar{n}_{k,l}\right) \tag{26}$$

$$p\left(\lambda_i | \boldsymbol{R}\right) = \mathrm{Beta}\left(e + r_i, f + (1 - r_i)\right) \tag{27}$$

2 A Gibbs solution for two-domain SIRM

2.1 The genarative model

In the case of cross-domain relational data ($D_1 \times D_2 \to \{0, 1\}$), we need to augment the "two-domain" IRM model. Its extension is easy: we just double the variables of "one-domain" SIRM. The generative model for the two-domain SRIM is described as follows:

$$\phi | a, b \sim \text{Beta}(a, b), \tag{28}$$
$$\theta_{k,l} | c_{k,l}, d_{k,l} \sim \text{Beta}(c_{k,l}, d_{k,l}), \tag{29}$$
$$\lambda_{1,i} | e_1, f_1 \sim \text{Beta}(e_1, f_1), \tag{30}$$
$$\lambda_{2,j} | e_2, f_2 \sim \text{Beta}(e_2, f_2), \tag{31}$$
$$r_{1,i} | \lambda_{1,i} \sim \text{Bernoulli}(\lambda_{1,i}), \tag{32}$$
$$r_{2,j} | \lambda_{2,j} \sim \text{Bernoulli}\left(\lambda_{2,j}\right), \tag{33}$$
$$z_{1,i} | r_{1,i} = 1, \alpha_1 \sim \text{CRP}(\alpha_1), \tag{34}$$
$$z_{1,i} | r_{1,i} = 0 \sim \mathbb{I}(z_{1,i} = 0), \tag{35}$$
$$z_{2,j} | r_{2,j} = 1, \alpha_2 \sim \text{CRP}(\alpha_2), \tag{36}$$
$$z_{2,j} | r_{2,j} = 0 \sim \mathbb{I}\left(z_{2,j} = 0\right), \tag{37}$$
$$x_{i,j} | \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{R}_1, \boldsymbol{R}_2, \{\theta\}, \phi \sim \text{Bernoulli}\left(\theta_{z_{1,i},z_{2,j}}^{r_{1,i}r_{2,j}} \phi^{1-r_{1,i}r_{2,j}}\right). \tag{38}$$

Eq. (28) defines the distribution of a relation strength for irrelevant data entries ϕ. Eq. (29) defines an relation strength from the first domain clsuter k to the second domain cluster l for relevant data entries. $\lambda_{1,i}, i = 1, 2, \ldots, N_1$ in Eq. (30) denotes the probability of a relevancy flag variable r_i in the first domain being 1. $r_{1,i} = \{0, 1\}$ in Eq. (32) indicates whether the object i of the first domain is relevant or not. $z_{1,i} \in \{1, 2, \ldots\}$ indicates the clsuter assignment of the object i in the first domain. If $r_{1,i} = 1$, then $z_{1,i}$ is chosen based on the CRP as in Eq. (34). Otherwise ($r_{1,i} = 0$), then its cluster assignments is set to 0 as in Eq. (35). In a symmetric fashion, $\lambda_{2,j}, j = 1, 2, \ldots, N_2$ in Eq. (30), $r_{2,j} = \{0, 1\}$ in Eq. (33), and $z_{2,j} \in \{1, 2, \ldots\}$ are defined in the second domain.

Finally, the observed relation $x_{i,j}, 1 \leq i, j \leq N$ is conditioned by all hidden variables. If the both of items i and j is assumed as relevant objects i.e. $r_{1,i} = r_{2,j} = 1$, then "relevant" relation strengths θ is used as a parameter of a Bernoulli trial. Otherwise, "irrelevant" relation strength ϕ is employed.

2.2 Probability distributions

$$p(\phi; a, b) = \phi^{a-1}(1-\phi)^{b-1} B^{-1}(a, b) \tag{39}$$

$$p(\boldsymbol{\Theta} = \{\theta\}; c, d) = \prod_{k=1}^{K} \prod_{l=1}^{K} \theta_{k,l}^{c_{k,l}-1} (1-\theta_{k,l})^{d_{k,l}-1} B^{-1}(c_{k,l}, d_{k,l}) \tag{40}$$

$$p(\boldsymbol{\lambda}_1 = \{\lambda_{1,i}\}; e_1, f_1) = \prod_{i=1}^{N_1} \lambda_{1,i}^{e_1-1} (1-\lambda_{1,i})^{f_1-1} B^{-1}(e_1, f_1). \tag{41}$$

$$p\left(\boldsymbol{\lambda}_2 = \{\lambda_{2,j}\}; e_2, f_2\right) = \prod_{j=1}^{N_2} \lambda_{2,j}^{e_2-1} \left(1-\lambda_{2,j}\right)^{f_2-1} B^{-1}(e_2, f_2). \tag{42}$$

$$p(\boldsymbol{R}_1 = \{r_{1,i}\}; \boldsymbol{\lambda}_1) = \prod_{i=1}^{N_1} \lambda_{1,i}^{r_{1,i}} (1-\lambda_{1,i})^{1-r_{1,i}}. \tag{43}$$

$$p\left(\boldsymbol{R}_2 = \{r_{2,j}\}; \boldsymbol{\lambda}_2\right) = \prod_{j=1}^{N_2} \lambda_{2,j}^{r_{2,j}} \left(1-\lambda_{2,j}\right)^{1-r_{2,j}}. \tag{44}$$

When the number of clusters in the first domain is K_1 excluding the $k = 0$th cluster,

$$p(\boldsymbol{Z}_1 = \{z_{1,i}\}; \boldsymbol{R}_1, \alpha_1) = \alpha^{K_1} \frac{\prod_{k=1}^{K_1} (m_{1,k}-1)!}{\prod_{i=1}^{M_1} (\alpha_1 + i - 1)} \tag{45}$$

where $m_{1,k}$ is defined in Eq. (50) and $M_1 = \sum_k m_{1,k}$. Similary, when the number of clusters in the second domain is K_2 excluding the $l = 0$th cluster,

$$p\left(\boldsymbol{Z}_2 = \{z_{2,j}\}; \boldsymbol{R}_2, \alpha_2\right) = \alpha^{K_2} \frac{\prod_{l=1}^{K_2} (m_{2,l}-1)!}{\prod_{j=1}^{M_2} (\alpha_2 + j - 1)} \tag{46}$$

where $m_{2,l}$ is defined in Eq. (51) and $M_2 = \sum_l m_{2,l}$.

$$\begin{aligned} & p(\boldsymbol{X} = \{x\}; \boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{\Theta}, \phi) \\ &= \prod_{i=1}^{N_1} \prod_{j=1}^{N_2} \left(\theta_{z_{1,i},z_{2,j}}^{r_{1,i}r_{2,j}} \phi^{1-r_{1,i}r_{2,j}}\right)^{x_{i,j}} \left(1 - \theta_{z_{1,i},z_{2,j}}^{r_{1,i}r_{2,j}} \phi^{1-r_{1,i}r_{2,j}}\right)^{(1-x_{i,j})} \\ &= \prod_{i=1}^{N_1} \prod_{j=1}^{N_2} \prod_{k=1}^{K_1} \prod_{l=1}^{K_2} \left[\theta_{k,l}^{x_{i,j}} (1-\theta_{k,l})^{(1-x_{i,j})}\right]^{r_{1,i}z_{1,i,k}r_{2,j}z_{2,j,l}} \\ &\quad \times \prod_{i=1}^{N_1} \prod_{j=1}^{N_2} \left[\phi^{x_{i,j}} (1-\phi)^{(1-x_{i,j})}\right]^{1-r_{1,i}r_{2,j}} \end{aligned} \tag{47}$$

2.3 Sampling Hidden Variables

As in the case of the one-domain models, simultaneous sampling of r_i and z_i leads to a simpler inference algorithm. Further, solutions for two domains are completely symmetric. Thus, we only present the sampling scheme for $r_{1,i}$ and $z_{1,i}$.

Regarding the sampling of the ith object in the first domain, let us denote the current number of realized clusters in D_1 and D_2 by K_1 and K_2, respectively. And we divide the observations $\boldsymbol{X}$ into two parts: data entries who relates to the object i $\boldsymbol{X}^{+i} = \{x_{i,j}\}_{j=1,\ldots,N_2}$, and those who does not

$\boldsymbol{X}^{\backslash i} = \{\boldsymbol{X} \setminus \boldsymbol{X}^{+i}\}$. Also we define the following quantities:

$$n_{k,l} = \sum_i \sum_j r_{1,i} z_{1,i,k} r_{2,j} z_{2,j,l} x_{i,j}, \tag{48}$$

$$\bar{n}_{k,l} = \sum_i \sum_j r_{1,i} z_{1,i,k} r_{2,j} z_{2,j,l} \left(1 - x_{i,j}\right), \tag{49}$$

$$m_{1,k} = \sum_i r_{1,i} z_{1,i,k}, \tag{50}$$

$$m_{2,l} = \sum_j r_{2,j} z_{2,j,l}, \tag{51}$$

$$q = \sum_i \sum_j \left(1 - r_{1,i} r_{2,j}\right) x_{i,j}, \tag{52}$$

$$\bar{q} = \sum_i \sum_j \left(1 - r_{1,i} r_{2,j}\right)\left(1 - x_{i,j}\right). \tag{53}$$

The superscript $\backslash i$ denotes the above statistics computed on $\boldsymbol{X}^{\backslash i}$. Also the superscript $+i0$, $+ik$ denotes that the same statistics computed on $\boldsymbol{X}^{+i}$ assuming $r_{1,i} = 0$ or $\{r_{1,i} = 1, z_{1,i} = k\}$, respectively.

We formulate the Gibbs posterior of $\{r_i, z_i\}$ as follows:

$$\begin{aligned} p\left(z_{1,i} = k, r_{1,i} | \boldsymbol{X}, \boldsymbol{Z}_1^{\backslash i}, \boldsymbol{Z}_2, \boldsymbol{R}_1^{\backslash i}, \boldsymbol{R}_2\right) \\ \propto p\left(z_{1,i} = k, r_{1,i} | \boldsymbol{Z}_1^{\backslash i}, \boldsymbol{R}_1^{\backslash i}\right) \\ \times p\left(\boldsymbol{X}^{+i} | z_{1,i} = k, r_{1,i}, \boldsymbol{Z}_1^{\backslash i}, \boldsymbol{Z}_2, \boldsymbol{R}_1^{\backslash i}, \boldsymbol{R}_2, \boldsymbol{X}^{\backslash i}\right) \end{aligned} \tag{54}$$

We can obtain the first term of the right hand of Eq. (54) by following the computation of Eq. (20). We easily obtain the followings for the prior term:

$$p\left(z_{1,i} = k, r_{1,i} | \boldsymbol{Z}_{1,\backslash i}, \boldsymbol{R}_{1,\backslash i}\right) \propto \begin{cases} f_1 + \sum_{i' \neq i} (1 - r_{1,i'}) & r_{1,i} = 0, z_{1,i} = 0 \\ (e_1 + \sum_{i' \neq i} r_{1,i'}) \frac{m_{1,k}^{\backslash i}}{\alpha_1 + \sum_k m_{1,k}^{\backslash i}} & r_{1,i} = 1, z_{1,i} = k \in \{1, \ldots, K_1\} \\ (e_1 + \sum_{i' \neq i} r_{1,i'}) \frac{\alpha_1}{\alpha_1 + \sum_k m_{1,k}^{\backslash i}} & r_{1,i} = 1, z_{1,i} = K_1 + 1 \end{cases} \tag{55}$$

The second term of the right hand of Eq. (54) is a likelihood term. Since the domain is separated for this case, the resulting solution is much simpler thant the case of one-domain model. Again we just follow the same path with Eq. (23) and Eq. (24), we can easily compute the likelihood terms. The results are shonw below:

$$p\left(\boldsymbol{X}^{+i} | z_{1,i} = 0, r_{1,i} = 0, \boldsymbol{Z}_{1,\backslash i}, \boldsymbol{Z}_2, \boldsymbol{R}_{1,\backslash i}, \boldsymbol{R}_2, \phi, \boldsymbol{\Theta}\right) = \frac{B\left(a + q^{\backslash i} + q^{+i0}, b + \bar{q}^{\backslash i} + \bar{q}^{+i0}\right)}{B\left(a + q^{\backslash i}, b + \bar{q}^{\backslash i}\right)}, \tag{56}$$

$$\begin{aligned} p\left(\boldsymbol{X}^{+i} | r_{1,i} = 1, z_{1,i} = k, \boldsymbol{Z}_{1,\backslash i}, \boldsymbol{Z}_2, \boldsymbol{R}_{1,\backslash i}, \boldsymbol{R}_2, \phi, \boldsymbol{\Theta}\right) \\ = \frac{B\left(a + q^{\backslash i} + q^{+ik}, b + \bar{q}^{\backslash i} + \bar{q}^{+ik}\right)}{B\left(a + q^{\backslash i}, b + \bar{q}^{\backslash i}\right)} \\ \times \prod_{l=1}^{K_2} \frac{B\left(c_{k,l} + n_{k,l}^{\backslash i} + n_{k,l}^{+ik}, d_{k,l} + \bar{n}_{k,l}^{\backslash i} + \bar{n}_{k,l}^{+ik}\right)}{B\left(c_{k,l} + n_{k,l}^{\backslash i}, d_{k,l} + \bar{n}_{k,l}^{\backslash i}\right)}. \end{aligned} \tag{57}$$

2.4 Posteriors of parameters

$$p\left(\phi | \boldsymbol{X}, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{R}_1, \boldsymbol{R}_2\right) = \mathrm{Beta}\left(a + q, b + \bar{q}\right) \tag{58}$$

$$p\left(\theta_{k,l} | \boldsymbol{X}, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{\Theta}_{\backslash(k,l)}\right) = \mathrm{Beta}\left(c_{k,l} + n_{k,l}, d_{k,l} + \bar{n}_{k,l}\right) \tag{59}$$

$$p\left(\lambda_{1,i} | \boldsymbol{R}_1\right) = \mathrm{Beta}\left(e_1 + r_{1,i}, f_1 + (1 - r_{1,i})\right) \tag{60}$$