Supplementary Information for Mixed Membership Matrix Factorization

Lester Mackey David Weiss Michael I. Jordan

1 Gibbs Sampling Conditionals for M^{3}F Models

1.1 The M^{3}F-TIB Model

In this section, we specify the conditional distributions used by the Gibbs sampler for the M^{3}F-TIB model.

Gibbs conditionals:

- $\Lambda^U | rest \backslash \{\mu^U\} \sim \text{Wishart}((\mathbf{W}_0^{-1} + \sum_{u=1}^{U} (\mathbf{a}_u - \bar{\mathbf{a}})(\mathbf{a}_u - \bar{\mathbf{a}})^t + \frac{\lambda_0 U}{\lambda_0 + U}(\mu_0 - \bar{\mathbf{a}})(\mu_0 - \bar{\mathbf{a}})^t)^{-1}, \nu_0 + U)$ where $\bar{\mathbf{a}} = \frac{1}{U}\sum_{u=1}^{U} \mathbf{a}_u$.
- $\Lambda^M | rest \backslash \{\mu^M\} \sim \text{Wishart}((\mathbf{W}_0^{-1} + \sum_{j=1}^{M} (\mathbf{b}_j - \bar{\mathbf{b}})(\mathbf{b}_j - \bar{\mathbf{b}})^t + \frac{\lambda_0 M}{\lambda_0 + M}(\mu_0 - \bar{\mathbf{b}})(\mu_0 - \bar{\mathbf{b}})^t)^{-1}, \nu_0 + M)$ where $\bar{\mathbf{b}} = \frac{1}{M}\sum_{j=1}^{M} \mathbf{b}_j$.
- $\mu^U | rest \sim \mathcal{N}\left(\frac{\lambda_0 \mu_0 + \sum_{u=1}^{U} \mathbf{a}_u}{\lambda_0 + U}, (\Lambda^U(\lambda_0 + U))^{-1}\right)$.
- $\mu^M | rest \sim \mathcal{N}\left(\frac{\lambda_0 \mu_0 + \sum_{j=1}^{M} \mathbf{b}_j}{\lambda_0 + M}, (\Lambda^M(\lambda_0 + M))^{-1}\right)$.
- For each u and $i \in \{1, \dots, K^M\}$,
$$c_u^i | rest \sim \mathcal{N}\left(\frac{\frac{c_0}{\sigma_0^2} + \sum_{j \in V_u} \frac{1}{\sigma^2} z_{uji}^M (r_{uj} - \chi_0 - d_j^{z_{uj}^U} - \mathbf{a}_u \cdot \mathbf{b}_j)}{\frac{1}{\sigma_0^2} + \sum_{j \in V_u} \frac{1}{\sigma^2} z_{uji}^M}, \frac{1}{\frac{1}{\sigma_0^2} + \sum_{j \in V_u} \frac{1}{\sigma^2} z_{uji}^M}\right).$$
- For each j and $i \in \{1, \dots, K^U\}$,
$$d_j^i | rest \sim \mathcal{N}\left(\frac{\frac{d_0}{\sigma_0^2} + \sum_{u: j \in V_u} \frac{1}{\sigma^2} z_{uji}^U (r_{uj} - \chi_0 - c_u^{z_{uj}^M} - \mathbf{a}_u \cdot \mathbf{b}_j)}{\frac{1}{\sigma_0^2} + \sum_{u: j \in V_u} \frac{1}{\sigma^2} z_{uji}^U}, \frac{1}{\frac{1}{\sigma_0^2} + \sum_{u: j \in V_u} \frac{1}{\sigma^2} z_{uji}^U}\right).$$

- For each u,

$$\mathbf{a}_u|rest \sim \mathcal{N}\left((\Lambda_u^{U*})^{-1}(\Lambda^U\mu^U + \sum_{j\in V_u}\frac{1}{\sigma^2}\mathbf{b}_j(r_{uj}-\chi_0-c_u^{z_{uj}^M}-d_j^{z_{uj}^U})), (\Lambda_u^{U*})^{-1}\right)$$

  where $\Lambda_u^{U*} = (\Lambda^U + \sum_{j\in V_u}\frac{1}{\sigma^2}\mathbf{b}_j(\mathbf{b}_j)^t)$.

- For each j,

$$\mathbf{b}_j|rest \sim \mathcal{N}\left((\Lambda_j^{M*})^{-1}(\Lambda^M\mu^M + \sum_{u:j\in V_u}\frac{1}{\sigma^2}\mathbf{a}_u(r_{uj}-\chi_0-c_u^{z_{uj}^M}-d_j^{z_{uj}^U})), (\Lambda_j^{M*})^{-1}\right)$$

  where $\Lambda_j^{M*} = (\Lambda^M + \sum_{u:j\in V_u}\frac{1}{\sigma^2}\mathbf{a}_u(\mathbf{a}_u)^t)$.

- For each u, $\theta_u^U|rest \sim Dir(\alpha/K^U + \sum_{j\in V_u} z_{uj}^U)$.
- For each j, $\theta_j^M|rest \sim Dir(\alpha/K^M + \sum_{u:j\in V_u} z_{uj}^M)$.
- For each u and $j \in V_u$, $z_{uj}^U|rest \sim \text{Multi}(1, \theta_{uj}^{U*})$ where

$$\theta_{uji}^{U*} \propto \theta_{ui}^U \text{exp}\left(-\frac{(r_{uj}-\chi_0-c_u^{z_{uj}^M}-d_j^i-\mathbf{a}_u\cdot\mathbf{b}_j)^2}{2\sigma^2}\right)$$

- For each j and $u : j \in V_u$, $z_{uj}^M|rest \sim \text{Multi}(1, \theta_{uj}^{M*})$ where

$$\theta_{uji}^{M*} \propto \theta_{ji}^M \,\text{exp}\left(-\frac{(r_{uj}-\chi_0-c_u^i-d_j^{z_{uj}^U}-\mathbf{a}_u\cdot\mathbf{b}_j)^2}{2\sigma^2}\right)$$

1.2 The M^{3}F-TIF Model

In this section, we specify the conditional distributions used by the Gibbs sampler for the M^{3}F-TIF model.

Gibbs conditionals:

- $\Lambda^U|rest\backslash\{\mu^U\} \sim \text{Wishart}((\mathbf{W}_0^{-1} + \sum_{u=1}^{U}(\mathbf{a}_u-\bar{\mathbf{a}})(\mathbf{a}_u-\bar{\mathbf{a}})^t + \frac{\lambda_0 U}{\lambda_0+U}(\mu_0-\bar{\mathbf{a}})(\mu_0-\bar{\mathbf{a}})^t)^{-1}, \nu_0+U)$ where $\bar{\mathbf{a}} = \frac{1}{U}\sum_{u=1}^{U}\mathbf{a}_u$.
- $\Lambda^M|rest\backslash\{\mu^M\} \sim \text{Wishart}((\mathbf{W}_0^{-1} + \sum_{j=1}^{M}(\mathbf{b}_j-\bar{\mathbf{b}})(\mathbf{b}_j-\bar{\mathbf{b}})^t + \frac{\lambda_0 M}{\lambda_0+M}(\mu_0-\bar{\mathbf{b}})(\mu_0-\bar{\mathbf{b}})^t)^{-1}, \nu_0+M)$ where $\bar{\mathbf{b}} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{b}_j$.
- $\mu^U|rest \sim \mathcal{N}\left(\frac{\lambda_0\mu_0 + \sum_{u=1}^{U}\mathbf{a}_u}{\lambda_0+U}, (\Lambda^U(\lambda_0+U))^{-1}\right)$.

- $\mu^M|rest \sim \mathcal{N}\left(\frac{\lambda_0\mu_0 + \sum_{j=1}^M \mathbf{b}_j}{\lambda_0 + M}, (\Lambda^M(\lambda_0+M))^{-1}\right)$.
- $\tilde{\Lambda}^U|rest\backslash\{\tilde{\mu}^U\} \sim \text{Wishart}((\tilde{\mathbf{W}}_0^{-1} + \sum_{u=1}^U \sum_{i=1}^{K^M} (\mathbf{c}_u^i - \bar{\mathbf{c}})(\mathbf{c}_u^i - \bar{\mathbf{c}})^t + \frac{\tilde{\lambda}_0 UK^M}{\tilde{\lambda}_0 + UK^M}(\tilde{\mu}_0 - \bar{\mathbf{c}})(\tilde{\mu}_0 - \bar{\mathbf{c}})^t)^{-1}, \tilde{\nu}_0 + UK^M)$ where $\bar{\mathbf{c}} = \frac{1}{UK^M}\sum_{u=1}^U \sum_{i=1}^{K^M} \mathbf{c}_u^i$.
- $\tilde{\Lambda}^M|rest\backslash\{\tilde{\mu}^M\} \sim \text{Wishart}((\tilde{\mathbf{W}}_0^{-1} + \sum_{j=1}^M \sum_{i=1}^{K^U} (\mathbf{d}_j^i - \bar{\mathbf{d}})(\mathbf{d}_j^i - \bar{\mathbf{d}})^t + \frac{\tilde{\lambda}_0 MK^U}{\tilde{\lambda}_0 + MK^U}(\tilde{\mu}_0 - \bar{\mathbf{d}})(\tilde{\mu}_0 - \bar{\mathbf{d}})^t)^{-1}, \tilde{\nu}_0 + MK^U)$ where $\bar{\mathbf{d}} = \frac{1}{MK^U}\sum_{j=1}^M \sum_{i=1}^{K^U} \mathbf{d}_j^i$.
- $\tilde{\mu}^U|rest \sim \mathcal{N}\left(\frac{\tilde{\lambda}_0\tilde{\mu}_0 + \sum_{u=1}^U \sum_{i=1}^{K^M} \mathbf{c}_u^i}{\tilde{\lambda}_0 + UK^M}, (\tilde{\Lambda}^U(\tilde{\lambda}_0 + UK^M))^{-1}\right)$.
- $\tilde{\mu}^M|rest \sim \mathcal{N}\left(\frac{\tilde{\lambda}_0\tilde{\mu}_0 + \sum_{j=1}^M \sum_{i=1}^{K^U} \mathbf{d}_j^i}{\tilde{\lambda}_0 + MK^U}, (\tilde{\Lambda}^M(\tilde{\lambda}_0 + MK^U))^{-1}\right)$.
- For each u,
$$\xi_u|rest \sim \mathcal{N}\left(\frac{\frac{\xi_0}{\sigma_0^2} + \sum_{j\in V_u} \frac{1}{\sigma^2}(r_{uj} - \chi_j - \mathbf{a}_u \cdot \mathbf{b}_j - \mathbf{c}_u^{z_{uj}^M} \cdot \mathbf{d}_j^{z_{uj}^U})}{\frac{1}{\sigma_0^2} + \sum_{j\in V_u}\frac{1}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \sum_{j\in V_u}\frac{1}{\sigma^2}}\right).$$
- For each j,
$$\chi_j|rest \sim \mathcal{N}\left(\frac{\frac{\chi_0}{\sigma_0^2} + \sum_{u:j\in V_u} \frac{1}{\sigma^2}(r_{uj} - \xi_u - \mathbf{a}_u \cdot \mathbf{b}_j - \mathbf{c}_u^{z_{uj}^M} \cdot \mathbf{d}_j^{z_{uj}^U})}{\frac{1}{\sigma_0^2} + \sum_{u:j\in V_u}\frac{1}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \sum_{u:j\in V_u}\frac{1}{\sigma^2}}\right).$$
- For each u,
$$\mathbf{a}_u|rest \sim \mathcal{N}\left((\Lambda_u^{U*})^{-1}(\Lambda^U\mu^U + \sum_{j\in V_u}\frac{1}{\sigma^2}\mathbf{b}_j(r_{uj} - \xi_u - \chi_j - \mathbf{c}_u^{z_{uj}^M} \cdot \mathbf{d}_j^{z_{uj}^U})), (\Lambda_u^{U*})^{-1}\right)$$
where $\Lambda_u^{U*} = (\Lambda^U + \sum_{j\in V_u}\frac{1}{\sigma^2}\mathbf{b}_j(\mathbf{b}_j)^t)$.
- For each j,
$$\mathbf{b}_j|rest \sim \mathcal{N}\left((\Lambda_j^{M*})^{-1}(\Lambda^M\mu^M + \sum_{u:j\in V_u}\frac{1}{\sigma^2}\mathbf{a}_u(r_{uj} - \xi_u - \chi_j - \mathbf{c}_u^{z_{uj}^M} \cdot \mathbf{d}_j^{z_{uj}^U})), (\Lambda_j^{M*})^{-1}\right)$$
where $\Lambda_j^{M*} = (\Lambda^M + \sum_{u:j\in V_u}\frac{1}{\sigma^2}\mathbf{a}_u(\mathbf{a}_u)^t)$.

- For each u and each $i \in 1, \ldots, K^M$,

$$\mathbf{c}_u^i | rest \sim \mathcal{N}\left((\tilde{\Lambda}_{ui}^{U*})^{-1} (\tilde{\Lambda}^U \tilde{\mu}^U + \sum_{j \in V_u} \frac{1}{\sigma^2} z_{uji}^M \mathbf{d}_j^{z_{uj}^U} (r_{uj} - \xi_u - \chi_j - \mathbf{a}_u \cdot \mathbf{b}_j)), (\tilde{\Lambda}_{ui}^{U*})^{-1} \right)$$

  where $\tilde{\Lambda}_{ui}^{U*} = (\tilde{\Lambda}^U + \sum_{j \in V_u} \frac{1}{\sigma^2} z_{uji}^M \mathbf{d}_j^{z_{uj}^U} (\mathbf{d}_j^{z_{uj}^U})^t)$.

- For each j and each $i \in 1, \ldots, K^U$,

$$\mathbf{d}_j^i | rest \sim \mathcal{N}\left((\tilde{\Lambda}_{ji}^{M*})^{-1} (\tilde{\Lambda}^M \tilde{\mu}^M + \sum_{u : j \in V_u} \frac{1}{\sigma^2} z_{uji}^U \mathbf{c}_u^{z_{uj}^M} (r_{uj} - \xi_u - \chi_j - \mathbf{a}_u \cdot \mathbf{b}_j)), (\tilde{\Lambda}_{ji}^{M*})^{-1} \right)$$

  where $\tilde{\Lambda}_{ji}^{M*} = (\tilde{\Lambda}^M + \sum_{u : j \in V_u} \frac{1}{\sigma^2} z_{uji}^U \mathbf{c}_u^{z_{uj}^M} (\mathbf{c}_u^{z_{uj}^M})^t)$.

- For each u, $\theta_u^U | rest \sim Dir(\alpha / K^U + \sum_{j \in V_u} z_{uj}^U)$.
- For each j, $\theta_j^M | rest \sim Dir(\alpha / K^M + \sum_{u : j \in V_u} z_{uj}^M)$.
- For each u and $j \in V_u$, $z_{uj}^U | rest \sim \text{Multi}(1, \theta_{uj}^{U*})$ where

$$\theta_{uji}^{U*} \propto \theta_{ui}^U \exp\left(- \frac{(r_{uj} - \xi_u - \chi_j - \mathbf{a}_u \cdot \mathbf{b}_j - \mathbf{c}_u^{z_{uj}^M} \cdot \mathbf{d}_j^i)^2}{2\sigma^2} \right)$$

- For each j and $u : j \in V_u$, $z_{uj}^M | rest \sim \text{Multi}(1, \theta_{uj}^{M*})$ where

$$\theta_{uji}^{M*} \propto \theta_{ji}^M \exp\left(- \frac{(r_{uj} - \xi_u - \chi_j - \mathbf{a}_u \cdot \mathbf{b}_j - \mathbf{c}_u^i \cdot \mathbf{d}_j^{z_{uj}^U})^2}{2\sigma^2} \right)$$