Supplementary Materials to Functional Clustering in Nested Designs: Modeling Variability in Reproductive Epidemiology Studies

Abel Rodriguez
University of California, Santa Cruz, California, USA
David B. Dunson
Duke University, Durham, North Carolina, USA

1. Details on the computational algorithm

1.1. Sampling in the mean-based clustering model

For the purpose of sampling the parameters of the mean-based clustering model we truncate the stick-breaking construction of the random distribution G so that it has K atoms, i.e,

$$G(\cdot) = \sum_{k=1}^{K} w_k \delta_{(\boldsymbol{\theta}_k^*, \sigma_k^{*2} \boldsymbol{\Lambda}_k^*)}$$

where $w_k = u_k \prod_{s<k}(1-u_s)$ with $u_k \sim \mathsf{Beta}(a,b)$ for $k < K$ and $u_k = 1$, which ensures that the weights of all components add up to 1. Also, we introduce a sequence of latent indicator variables $\zeta_1, \ldots, \zeta_I$, where $\zeta_i = k$ if and only if subject i is assigned to cluster k, $(\boldsymbol{\theta}_i, \sigma_i^2, \boldsymbol{\Lambda}_i) = (\boldsymbol{\theta}_{\zeta_i}^*, \sigma_{\zeta_i}^{*2}, \boldsymbol{\Lambda}_{\zeta_i}^*)$. After introducing these latent indicators, the posterior distribution of the model parameters can be written as:

$$\begin{aligned} p(\{\boldsymbol{\theta}_{ij}\}, \{\boldsymbol{\theta}_k^*\}, \{\sigma_k^{*2}\}, \{\boldsymbol{\Lambda}_k^*\}, \{\zeta_i\}, \{w_k\}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, a, b, \gamma \,|\{\mathbf{y}_{ij}\}) \propto \\ p(\{\mathbf{y}_{ij}\}|\{\boldsymbol{\theta}_{ij}\}, \{\sigma_k^{*2}\}, \{\zeta_i\}) p(\{\boldsymbol{\theta}_{ij}\}|\{\boldsymbol{\theta}_k^*\}, \{\boldsymbol{\Lambda}_k^*\}, \{\sigma_k^{*2}\}, \boldsymbol{\Sigma}, \{\zeta_i\}) p(\boldsymbol{\Sigma}) \\ p(\{\boldsymbol{\theta}_k^*\}, \{\boldsymbol{\Lambda}_k^*\}, \{\sigma_k^{*2}\}|\boldsymbol{\Omega}, \nu_2, \gamma) p(\{\zeta_i\}|\{w_k\}) p(\{w_k\}|a,b) p(\boldsymbol{\Omega}) p(\nu_2) p(\gamma) p(a,b) \end{aligned}$$

The indicators $\{\zeta_i\}$ and the cluster specific parameters $\{\boldsymbol{\theta}_k^*\}$, $\{\boldsymbol{\Lambda}_k^*\}$ and $\{\sigma_k^{*2}\}$ can be sampled easily after integrating out the curve-specific parameters $\{\boldsymbol{\theta}_{ij}\}$. Specifically, ζ_i can be sampled conditionally on all other parameters in the model from a multinomial distribution where

$$\mathsf{Pr}(\zeta_i = k \mid \cdots) \propto w_k \sigma_{\zeta_i}^{-\sum_{j=1}^{n_i} T_{ij}} \times \\ \exp\left\{ -\frac{1}{2\sigma_{\zeta_i}^2} \sum_{j=1}^{n_i} \left(\mathbf{y}_{ij} - \mathbf{B}(\mathbf{x}_{ij})\boldsymbol{\Lambda}_{\zeta_i}^* \boldsymbol{\theta}_{\zeta_i}^*\right)' (\boldsymbol{\Sigma} + \mathbf{I})^{-1} \left(\mathbf{y}_{ij} - \mathbf{B}(\mathbf{x}_{ij})\boldsymbol{\Lambda}_{\zeta_i}^* \boldsymbol{\theta}_{\zeta_i}^*\right) \right\}$$

and $k = 1, \ldots, K$, while the cluster specific parameters are sampled by first integrating out $\boldsymbol{\theta}_k^*$ and σ_k^{*2} and sampling $\boldsymbol{\Lambda}_k^*$, and then sampling $\boldsymbol{\theta}_k^*$ and σ_k^{*2} conditionally on $\boldsymbol{\Lambda}_k^*$. The probability of the model associated with $\boldsymbol{\Lambda}_k^*$ is proportional to

$$\mathsf{Pr}(\boldsymbol{\Lambda}_k^* = \boldsymbol{\Lambda} \mid \cdots) \propto (2\pi)^{-s_k/2} \left|\boldsymbol{\Omega}\right|^{-1/2} \left|\boldsymbol{\Omega}^{-1} + \mathbf{E}_k(\boldsymbol{\Lambda})\right|^{1/2} \frac{\Gamma(\nu_1 + s_k/2)}{\Gamma(\nu_1)} \nu_2^{\nu_1} \\ \left\{\nu_2 + \left[A_k(\boldsymbol{\Lambda}) - \mathbf{d}_k(\boldsymbol{\Lambda})'\{\boldsymbol{\Omega}^{-1} + \mathbf{E}_k(\boldsymbol{\Lambda})\}\mathbf{d}_k(\boldsymbol{\Lambda})\right]/2\right\}^{-(\nu_1 + s_k/2)} \quad (1)$$

with

$$s_k = \sum_{\{i:\zeta_i = k\}} \sum_{j=1}^{n_i} T_{ij}$$

$$A_k(\boldsymbol{\Lambda}) = \sum_{\{i:\zeta_i = k\}} \sum_{j=1}^{n_i} \mathbf{y}_{ij}' \left[\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}} + \mathbf{I}_{p_{\boldsymbol{\Lambda}}}\right]^{-1} \mathbf{y}_{ij}$$

$$\mathbf{d}_k(\boldsymbol{\Lambda}) = \sum_{\{i:\zeta_i = k\}} \sum_{j=1}^{n_i} \mathbf{B}_{\boldsymbol{\Lambda}}(\mathbf{x}_{ij})' \left[\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}} + \mathbf{I}\right]^{-1} \mathbf{y}_{ij}$$

$$\mathbf{E}_k(\boldsymbol{\Lambda}) = \sum_{\{i:\zeta_i = k\}} \sum_{j=1}^{n_i} \mathbf{B}_{\boldsymbol{\Lambda}}(\mathbf{x}_{ij})' \left[\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}} + \mathbf{I}\right]^{-1} \mathbf{B}_{\boldsymbol{\Lambda}}(\mathbf{x}_{ij})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}$ and $\mathbf{B}_{\Lambda}(\mathbf{x}_{ij})$ correspond to the restrictions of matrices $\boldsymbol{\Sigma}$ and $\mathbf{B}(\mathbf{x}_{ij})$ to the entries where the diagonal elements of $\boldsymbol{\Lambda}$ are different from 0. In our case, since the number of possible values for $\boldsymbol{\Lambda}_k^*$ is small ($2^{13} = 8192$), we explicitly compute the full-conditional probability distribution for each of the possible models and perform exact sampling from this posterior distribution. More generally, when the number of nodes is large we can use a random-walk Metropolis-Hasting algorithm as described in George & McCulloch (1997). Now, conditionally on $\boldsymbol{\Lambda}_k^*$, we can sample σ_k^* and $\boldsymbol{\theta}_k^*$ from

$$\sigma_k^* \mid \boldsymbol{\Lambda}_k^* \sim \mathsf{IGam}\left(\nu_1 + s_k/2, \nu_2 + \left[A_k(\boldsymbol{\Lambda}_k^*) - \mathbf{d}_k(\boldsymbol{\Lambda}_k^*)'\{\boldsymbol{\Omega}^{-1} + \mathbf{E}_k(\boldsymbol{\Lambda}_k^*)\}\mathbf{d}_k(\boldsymbol{\Lambda}_k^*)\right]/2\right)$$

$$\boldsymbol{\theta}_k^* \mid \sigma_k^*, \boldsymbol{\Lambda}_k^* \sim \mathsf{N}\left(\left[\boldsymbol{\Omega}^{-1} + \mathbf{E}_k(\boldsymbol{\Lambda}_k^*)\right]^{-1} \mathbf{d}_k(\boldsymbol{\Lambda}_k^*), \sigma_k^* \left[\boldsymbol{\Omega}^{-1} + \mathbf{E}_k(\boldsymbol{\Lambda}_k^*)\right]^{-1}\right)$$

The component weights are simulated through the stick-breaking ratios, whose full conditional distribution is given by

$$u_k | \cdots \sim \mathsf{Beta}\left(a + r_k, b + \sum_{s=k+1}^{K} r_s\right)$$

where $r_k = \sum_{i=1}^{I} \mathbf{1}_{(\zeta_i = k)}$ is the number of observations assigned to component k of the mixture. Given the indicators and class specific parameters, we can sample the parameters $\{\boldsymbol{\theta}_{ij}\}$ independently from their full conditional distribution.

$$\begin{aligned}\boldsymbol{\theta}_{ij} | \cdots \sim \mathsf{N}\Big(\left[\mathbf{B}(\mathbf{x}_{ij})'\mathbf{B}(\mathbf{x}_{ij}) + \boldsymbol{\Lambda}^{*}_{\zeta_i}{}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}^{*}_{\zeta_i}\right]^{-1} \left[\mathbf{B}(\mathbf{x}_{ij})'\mathbf{y}_{ij} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}^{*}_{\zeta_i}\boldsymbol{\theta}^{*}_{\zeta_i}\right], \\ \sigma^{*2}_{\zeta_i}\left[\mathbf{B}(\mathbf{x}_{ij})'\mathbf{B}(\mathbf{x}_{ij}) + \boldsymbol{\Lambda}^{*}_{\zeta_i}{}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}^{*}_{\zeta_i}\right]\Big)\end{aligned}$$

The rest of the steps are relatively straightforward. The full conditional random effects variance $\boldsymbol{\Sigma}$ can be sampled from an inverse Wishart distribution,

$$\boldsymbol{\Sigma} | \cdots \sim \mathsf{IWis}\left(\nu_{\Sigma} + \sum_{i=1}^{I}\sum_{j=1}^{n_i} T_{ij}, \boldsymbol{\Sigma}_0 + \sum_{i=1}^{I}\sum_{j=1}^{n_i} \frac{1}{\sigma^{*2}_{\zeta_i}}(\boldsymbol{\theta}_{ij} - \boldsymbol{\theta}_{\zeta_i})(\boldsymbol{\theta}_{ij} - \boldsymbol{\theta}_{\zeta_i})'\right\},$$

and the parameters of the baseline measure $\boldsymbol{\Omega}$, ν_2 and γ are respectively sampled from another inverse Wishart, a Gamma and a beta distribution,

$$\begin{aligned}\boldsymbol{\Omega} | \cdots &\sim \mathsf{IWis}\left(\nu_{\Omega} + Kp, \boldsymbol{\Omega}_0 + \sum_{k=1}^{K} \frac{1}{\sigma^{*2}_{k}} \boldsymbol{\theta}^{*}_{k}\boldsymbol{\theta}^{*}_{k}{}'\right) \\ \nu_2 | \cdots &\sim \mathsf{Gam}\left(\rho + K\nu_1, \psi + \sum_{k=1}^{K}\sigma^{*2}_{k}\right) \\ \gamma | \cdots &\sim \mathsf{Beta}\left(\eta_1 + \sum_{k=1}^{K} \operatorname{tr}\boldsymbol{\Lambda}^{*}_{k}, \eta_2 + Kp - \sum_{k=1}^{K}\operatorname{tr}\boldsymbol{\Lambda}^{*}_{k}\right)\end{aligned}$$

where $\operatorname{tr}\mathbf{D}$ denotes the trace of the matrix $\mathbf{D}$. Finally, sampling for the precision parameters a and b is done using a random-walk Metropolis-Hastings algorithm with log normal proposals.

1.2. Sampling in the distribution-based clustering model

Again, in this case we truncate the stick-breaking constructions to generate a blocked Gibbs sampler. Hence,

$$G_i \sim \sum_{k=1}^{K} \pi_k \delta_{G_k^*} \qquad\qquad G_k^* = \sum_{l=1}^{L} \varpi_{lk} \delta_{(\boldsymbol{\theta}_{lk}^*, \sigma_{lk}^{*2}, \boldsymbol{\Lambda}_{lk}^*)}$$

Also, we introduce two sequences of latent indicator variables, $\zeta_1, \ldots, \zeta_I$ and $\{\xi_{1j}\}_{j=1}^{n_1}, \ldots, \{\xi_{Ij}\}_{j=1}^{n_I}$ such that$(\boldsymbol{\theta}_{ij}, \sigma_{ij}^2, \boldsymbol{\Lambda}_{ij}) = (\boldsymbol{\theta}^*_{\xi_{ij},\zeta_i}, \sigma^{*2}_{\xi_{ij},\zeta_i}, \boldsymbol{\Lambda}^*_{\xi_{ij},\zeta_i})$. After introducing these latent indicators, the posterior distribution of the model parameters can be written as:

$$\begin{aligned} p(\{\zeta_i\}, \{\xi_{ij}\}, \{\boldsymbol{\theta}_{lk}^*\}, \{\sigma_{lk}^{*2}\}, \{\boldsymbol{\Lambda}_{lk}^*\}, \{\pi_k\}, \{\varpi_{lk}\}, \boldsymbol{\Omega}, a_1, b_1, a_2, b_2, \gamma \,|\{\mathbf{y}_{ij}\}) \propto \\ p(\{\mathbf{y}_{ij}\}|\{\zeta_i\}, \{\xi_{ij}\}, \{\boldsymbol{\theta}_{lk}^*\}, \{\sigma_{lk}^{*2}\}, \{\boldsymbol{\Lambda}_{lk}^*\}) p(\{\boldsymbol{\theta}_{lk}^*\}, \{\sigma_{lk}^{*2}\}, \{\boldsymbol{\Lambda}_{lk}^*\}|\boldsymbol{\Omega}, \nu_2, \gamma) \\ p(\{\zeta_i\}|\{\pi_k\}) p(\{\pi_k\}|a_1, b_1) p(\{\xi_{ij}\}|\{\varpi_{lk}\}) p(\{\varpi_{lk}\}|a_2, b_2) p(\boldsymbol{\Omega}) p(\nu_2) \\ p(\gamma) p(a_1, b_1) p(a_2, b_2) \end{aligned}$$

The indicators $\{\zeta_i\}$ can be sampled from a multinomial distribution with weights

$$\begin{aligned} \mathsf{Pr}(\zeta_i = k|\cdots) \propto \pi_k \Bigg\{ \prod_{j=1}^{n_i} \Bigg[\sum_{l=1}^{L} \varpi_{lk} \left(2\pi\sigma_{lk}^{*2}\right)^{-T_{ij}/2} \times \\ \exp\left\{ -\frac{1}{2\sigma_{lk}^{*2}} \left(\mathbf{y}_{ij} - \mathbf{B}(\mathbf{x}_{ij})\boldsymbol{\Lambda}_{lk}^*\boldsymbol{\theta}_{lk}^*\right)' \left(\mathbf{y}_{ij} - \mathbf{B}(\mathbf{x}_{ij})\boldsymbol{\Lambda}_{lk}^*\boldsymbol{\theta}_{lk}^*\right) \right\} \Bigg] \Bigg\} \end{aligned}$$

for $k = 1, \ldots, K$. Similarly the indicators $\{\xi_{ij}\}$ are sampled from another multinomial where

$$\begin{aligned} \mathsf{Pr}(\xi_{ij} = l|\cdots) \propto \varpi_{l\zeta_i} \left(\sigma_{l\zeta_i}^*\right)^{-T_{ij}} \times \\ \exp\left\{ -\frac{1}{2\sigma_{l\zeta_i}^{*2}} \left(\mathbf{y}_{ij} - \mathbf{B}(\mathbf{x}_{ij})\boldsymbol{\Lambda}_{l\zeta_i}^*\boldsymbol{\theta}_{l\zeta_i}^*\right)' \left(\mathbf{y}_{ij} - \mathbf{B}(\mathbf{x}_{ij})\boldsymbol{\Lambda}_{l\zeta_i}^*\boldsymbol{\theta}_{l\zeta_i}^*\right) \right\} \end{aligned}$$

and $l = 1, \ldots, L$. The rest of the parameters are sampled almost identically as in the mean-based clustering algorithm. The cluster specific parameters are sampled by first integrating out $\boldsymbol{\theta}_{lk}^*$ and σ_{lk}^{*2} and sampling $\boldsymbol{\Lambda}_{lk}^*$, and then sampling $\boldsymbol{\theta}_{lk}^*$ and σ_{lk}^{*2} condtionally on $\boldsymbol{\Lambda}_{lk}^*$. The expression for $\mathsf{Pr}(\boldsymbol{\Lambda}_{lk}^* = \boldsymbol{\Lambda}|\cdots)$ is similar to (1), but s_k,

$A_k(\mathbf{\Lambda})$, $\mathbf{d}_k(\mathbf{\Lambda})$ and $\mathbf{E}_k(\mathbf{\Lambda})$ are replaced by

$$\begin{aligned}
s_{lk} &= \sum_{\{(i,j):\zeta_i=k,\xi_{ij}=l\}} T_{ij} \\
A_{lk}(\mathbf{\Lambda}) &= \sum_{\{(i,j):\zeta_i=k,\xi_{ij}=l\}} \mathbf{y}_{ij}'\mathbf{y}_{ij} \\
\mathbf{d}_{lk}(\mathbf{\Lambda}) &= \sum_{\{(i,j):\zeta_i=k,\xi_{ij}=l\}} \mathbf{B}_{\mathbf{\Lambda}}(\mathbf{x}_{ij})'\mathbf{y}_{ij} \\
\mathbf{E}_{lk}(\mathbf{\Lambda}) &= \sum_{\{(i,j):\zeta_i=k,\xi_{ij}=l\}} \mathbf{B}_{\mathbf{\Lambda}}(\mathbf{x}_{ij})'\mathbf{B}_{\mathbf{\Lambda}}(\mathbf{x}_{ij})
\end{aligned}$$

The mixture weights are sampled again through the stick-breaking weights $\{v_k\}$ and $\{u_{lk}\}$ so that

$$v_k \sim \mathsf{Beta}\left(a_1 + r_k, b_1 + \sum_{s=k+1}^{K} r_s\right), \qquad u_{lk} \sim \mathsf{Beta}\left(a_2 + m_{lk}, b_2 + \sum_{s=l+1}^{L} m_{ls}\right),$$

where $r_k = \sum_{i=1}^{I} \mathbf{1}_{(\zeta_i=k)}$ and $m_{lk} = \sum_{i=1}^{I}\sum_{j=1}^{n_i} \mathbf{1}_{(\zeta_i=k,\xi_{ij}=l)}$. The parameters of the baseline measure $\mathbf{\Omega}$, ν_2 and γ are sampled from their full conditional distributions,

$$\begin{aligned}
\mathbf{\Omega}|\cdots &\sim \mathsf{IWis}\left(\nu_\Omega + Kp, \mathbf{\Omega}_0 + \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{1}{\sigma_{lk}^{*2}} \boldsymbol{\theta}_{lk}^{*}\boldsymbol{\theta}_{lk}^{*\prime}\right) \\
\nu_2|\cdots &\sim \mathsf{Gam}\left(\rho + K\nu_1, \psi + \sum_{k=1}^{K}\sum_{l=1}^{L} \sigma_{lk}^{*2}\right) \\
\gamma|\cdots &\sim \mathsf{Beta}\left(\eta_1 + \sum_{k=1}^{K}\sum_{l=1}^{L} \operatorname{tr} \mathbf{\Lambda}_{lk}^{*}, \eta_2 + Kp - \sum_{k=1}^{K}\sum_{l=1}^{L} \operatorname{tr} \mathbf{\Lambda}_{lk}^{*}\right)
\end{aligned}$$

Finally, sampling for the precision parameters a_1, b_1, a_2 and b_2 is done as before using a random-walk Metropolis-Hastings algorithm with log normal proposals.

References

George, E. I. & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–373.