Supplemental Material to Inference for Environmental Intervention Studies using Principal Stratification

Amber J. Hackstadt, Elizabeth C. Matsui, D'Ann L. Williams,
Gregory B. Diette, Patrick N. Breysse, Arlene M. Butz, and Roger D. Peng

Estimates for the parameters of interest and the causal effects are found using the Bayesian model discussed in Section 3 of the paper. A MCMC algorithm is used with the steps summarized as follows:

1. Update the missing PM potential outcomes
2. Update the missing SFD potential outcomes
3. Update $\boldsymbol{\theta}_X = (\boldsymbol{\gamma}', \delta_1)'$
4. Update $\boldsymbol{\theta}_{\boldsymbol{Y}} = (\boldsymbol{\alpha}', \beta_1, \beta_2, \beta_3)'$
5. Update the variance components ξ^2, σ_0, and σ_1.

For the update in the first step, we sample from the full conditional distribution for X_i^{mis} given all other potential outcomes and parameters using the most recently sampled values of the other parameters and potential outcomes. If the household of individual i is assigned to the control group, $X_i(1)$ is missing and we sample the missing PM potential outcome, denoted $X_i^{\text{mis}}(1)$, from a normal distribution with variance

$$\sigma_{X1}^2 = \left[\frac{1}{\sigma_1^2(1-\rho^2)} + \frac{(\beta_2+\beta_3)^2}{\xi^2}\right]^{-1} \tag{1}$$

and mean

$$\mu_{X1} = \sigma_{X1}^2 \left[\frac{1}{\sigma_1^2(1-\rho^2)}\left(\mu_{1i} + \rho\frac{\sigma_1}{\sigma_0}\left\{X_i^{\text{obs}}(0) - \mu_{0i}\right\}\right) + \frac{\beta_2+\beta_3}{\xi^2}\left(Y_i^{\text{mis}}(1) - \boldsymbol{Z}_i'\boldsymbol{\alpha} - \beta_1\right)\right] \tag{2}$$

where μ_{0i} and μ_{1i} are as defined in Section 3.2 of the paper, $X_i^{\text{obs}}(0)$ is the observed value for $X_i(0)$, and $Y_i^{\text{mis}}(1)$ is the value for the missing SFD potential outcome, $Y_i(1)$. If the household of individual i is assigned to the treatment group, $X_i(0)$ is missing and we sample the missing PM potential outcome, denoted $X_i^{\text{mis}}(0)$, from a normal distribution with variance

$$\sigma_{X0}^2 = \left[\frac{1}{\sigma_0^2(1-\rho^2)} + \frac{\beta_2^2}{\xi^2}\right]^{-1} \tag{3}$$

and mean

$$\mu_{X0} = \sigma_{X0}^2\left[\frac{1}{\sigma_0^2(1-\rho^2)}\left(\mu_{0i} + \rho\frac{\sigma_0}{\sigma_1}\left\{X_i^{\text{obs}}(1) - \mu_{1i}\right\}\right) + \frac{\beta_2}{\xi^2}\left(Y_i^{\text{mis}}(0) - \boldsymbol{Z}_i'\boldsymbol{\alpha}\right)\right] \tag{4}$$

where μ_{0i} and μ_{1i} are as defined in Section 3.2 of the paper, $X_i^{\text{obs}}(1)$ is the observed value for $X_i(1)$, and $Y_i^{\text{mis}}(0)$ is the value for the missing SFD potential outcome, $Y_i(0)$.

For the update in Step 2, we sample from the full conditional distribution for Y_i^{mis} given all other potential outcomes and parameters using the most recently sampled values of the other parameters and potential outcomes. If the household of individual i is assigned to the control group, $Y_i(1)$ is missing and we sample $Y_i^{\text{mis}}(1)$ from a normal distribution with mean $\boldsymbol{Z}_i'\boldsymbol{\alpha}+\beta_1+(\beta_2+\beta_3)X_i^{\text{mis}}(1)$ and variance ξ^2. Likewise, if the household of individual i is assigned to the treatment group, $Y_i(0)$ is missing and we sample $Y_i^{\text{mis}}(0)$ from a normal distribution with mean $\boldsymbol{Z}_i'\boldsymbol{\alpha} + \beta_2 X_i^{\text{mis}}(0)$ and variance ξ^2.

For the update in Step 3, we sample from the full conditional distribution for $\boldsymbol{\theta}_X$ given all other parameters and the potential outcomes using the most recently sampled values of the other parameters and potential outcomes. This full conditional distribution is a $r+1$ dimensional multivariate normal with variance

$$C = \left[\Sigma_{\theta X}^{-1} + \sum_{i=1}^{n}\left(B_{Xi}'\Sigma_X^{-1}B_{Xi}\right)\right]^{-1} \tag{5}$$

and mean

$$C\sum_{i=1}^{n}\left(B_{Xi}'\Sigma_X^{-1}\boldsymbol{X}_i\right) \tag{6}$$

where $\Sigma_{\theta X}$ is the diagonal matrix with diagonal elements $\left(\sigma_{\gamma 1}^2, \ldots, \sigma_{\gamma r}^2, \sigma_\delta^2\right)$, Σ_X is the variance matrix of the PM model in (3) in Section 3.2 of the paper, $\boldsymbol{X}_i$ is as defined in Section 3.2 of the paper, and B_{Xi} is a matrix with the first row equal to $(\boldsymbol{W}_i', 0)$ and second row equal to $(\boldsymbol{W}_i', 1)$.

For the update in Step 4, we sample from the full conditional distribution for $\boldsymbol{\theta}_Y$ given all other parameters and the potential outcomes using the most recently sampled values of the other parameters and potential outcomes. This full conditional distribution is a $s+3$ dimensional multivariate normal with variance

$$D = \left[\Sigma_{\theta Y}^{-1} + B_{Y0}^{'}\Sigma_Y^{-1}B_{Y0} + B_{Y1}^{'}\Sigma_Y^{-1}B_{Y1}\right]^{-1} \tag{7}$$

and mean

$$D\left[B_{Y0}^{-1}\Sigma_Y^{-1}\boldsymbol{Y}_0 + B_{Y1}^{-1}\Sigma_Y^{-1}\boldsymbol{Y}_1\right] \tag{8}$$

where $\Sigma_{\theta Y}$ is a diagonal matrix with diagonal elements $\left(\sigma_{\alpha 1}^2, \ldots, \sigma_{\alpha s}^2, \sigma_{\beta}^2, \sigma_{\beta}^2, \sigma_{\beta}^2\right)$, Σ_Y^{-1} is a $n \times n$ diagonal matrix with diagonal elements equal to ξ^2, B_{Y0} is a $n \times (s+3)$ matrix with the ith row equal to $(\boldsymbol{Z}_i', 0, X_i\left(0\right), 0)$, B_{Y1} is a $n \times (s+3)$ matrix with the ith row equal to $(\boldsymbol{Z}_i', 1, X_i\left(1\right), X_i\left(1\right))$, $\boldsymbol{Y}_0$ is a $n \times 1$ column vector with the ith element equal to $Y_i\left(0\right)$, and $\boldsymbol{Y}_1$ is a $n \times 1$ column vector with the ith element equal to $Y_i\left(1\right)$.

To sample values for ξ^2 for the update in Step 5, we sample from the full conditional distribution for ξ^2 given all other parameters and the potential outcomes using the most recently sampled values of the other parameters and potential outcomes. This full conditional distribution is a inverse gamma distribution with shape $a = c_1 + n$ and scale

$$b = d_1 + 1/2\sum_{i=1}^{n}\left[Y_i\left(0\right) - \boldsymbol{Z}_i^{'}\boldsymbol{\alpha} - \beta_2 X_i\left(0\right)\right]^2 + 1/2\sum_{i=1}^{n}\left[Y_i\left(1\right) - \boldsymbol{Z}_i^{'}\boldsymbol{\alpha} - \beta_1 - \left(\beta_2 + \beta_3\right)X_i\left(1\right)\right]^2 \tag{9}$$

where the probability density function is of the form $b^a/\Gamma\left(a\right)\left(\xi^2\right)^{-(a+1)}\exp\left\{-b/\xi^2\right\}$ and $\Gamma\left(\cdot\right)$ is the gamma function.

To sample values for σ_0 for the update in Step 5, we sample from the full conditional distribution for σ_0 given all other parameters and the potential outcomes using the most recently sampled values of the other parameters and potential outcomes. This conditional distribution is proportional to

$$\prod_{i=1}^{n}\left[(2\pi)^{-1}\left|\Sigma_X\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\boldsymbol{X}_i-\boldsymbol{\mu}_{Xi}\right)^{'}\Sigma_X^{-1}\left(\boldsymbol{X}_i-\boldsymbol{\mu}_{Xi}\right)\right\}\right]\times \tag{10}$$
$$\frac{1}{\sigma_0\sqrt{(2\pi m_0^2)}}\exp\left\{-\frac{1}{2m_0^2}\left[\log\left(\sigma_0\right)\right]^2\right\}$$

where $\boldsymbol{\mu}_{Xi}=\left(\mu_{0i},\mu_{1i}\right)^{'}$. This distribution is proper but intractable so to sample from this distribution, we use a log transformation and a random walk in this step.

Likewise, to sample values for σ_1 for the update in Step 5, we sample from the full conditional distribution for σ_1 given all other parameters and the potential outcomes using the most recently sampled values of the other parameters and potential outcomes. This conditional distribution is proportional to

$$\prod_{i=1}^{n}\left[(2\pi)^{-1}\left|\Sigma_X\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\boldsymbol{X}_i-\boldsymbol{\mu}_{Xi}\right)^{'}\Sigma_X^{-1}\left(\boldsymbol{X}_i-\boldsymbol{\mu}_{Xi}\right)\right\}\right]\times \tag{11}$$
$$\frac{1}{\sigma_1\sqrt{(2\pi m_1^2)}}\exp\left\{-\frac{1}{2m_1^2}\left[\log\left(\sigma_1\right)\right]^2\right\}.$$

This distribution is proper but intractable so to sample from this distribution, we again use a log transformation and a random walk in this step.