Supplementary Materials for Bayesian Inference for the Causal Effect of Mediation

M. J. Daniels [*], J. Roy [†], C. Kim [‡], J.W. Hogan [§] and M.G. Perri [¶]

Step 3. of sampling algorithm: Sampling $(F_{M_1,1}, F_{M_1,0}, F_{M_0,1}, F_{M_0,0}, \pi_{1,M_1}, \pi_{0,M_0})$

1. Generate L sets (M_0, M_1). This requires sampling from $f_{M_z,M_{z'}}(m_z, m_{z'}) = f_{M_z}(m_z|m_{z'})f_{M_{z'}}(m_{z'})$. Note that in the TOURS trial, M_z is actually discrete (taking integer values 0 to 350), so we compute $f_{M_z}(m_z)$ as follows,

$$f_{M_z}(m_z) = F_{M_z}(m_z + 0.5) - F_{M_z}(m_z - 0.5); z = 0, 1$$

where $F_{M_z} = F_{M_z,1} \times \pi_{1,M_1} + F_{M_z,0} \times (1 - \pi_{1,M_1})$.

We sample $M_{z'}$ using $F_{M_{z'}}(M_{z'}) \sim Unif(0, 1)$. Then, given $M_{z'}$, we obtain M_z using the conditional CDF

$$\begin{aligned} F_{M_z}(m_z|m_{z'}) &= \sum_{t=0}^{m_z} f_{M_z}(t|m_{z'}) \\ &= \frac{\sum_{t=0}^{m_z} f_{M_z,M_{z'}}(t, m_{z'})}{f_{M_{z'}}(m_{z'})} \\ &= \frac{F_{M_z,M_{z'}}(m_z + 0.5, m_{z'} + 0.5) - F_{M_z,M_{z'}}(m_z + 0.5, m_{z'} - 0.5)}{F_{M_{z'}}(m_{z'} + 0.5) - F_{M_{z'}}(m_{z'} - 0.5)} \end{aligned}$$

using the fact $F_{M_z}(M_z|m_{z'}) \sim Unif(0, 1)$.

[*]Department of Statistics,University of Florida, Gainesville, FL 32611
[†]Department of Biostatistics, University of Pennsylvania, Philadelphia, PA
[‡]Department of Statistics,University of Florida, Gainesville, FL 32611
[§]Department of Biostatistics, Brown University, Providence, RI
[¶]Department of Clinical and Health Psychology, University of Florida, Gainesville, FL 32611

The densities $f_{M_z,M_{z'}}(m_z, m_{z'})$ can be computed in the same manner with Assumption 4.

2. Compute $f_{1,M_0}(y)$ via Monte Carlo integration using the L sets (M_0, M_1) as follows

$$\begin{aligned}
f_{1,M_0}(y) &= \int f_{1,M_0}(y|m_0,m_1) f_{M_0,M_1}(m_0,m_1) dm_0 dm_1 \\
&= \frac{1}{C(\chi,\epsilon)} \int \exp\{\text{sgn}(d)\log\chi y I(|d| \geq \epsilon)\} f_{1,M_1}(y|m_0,m_1)\ f_{M_0,M_1}(m_0,m_1)\ dm_0 dm_1 (\text{A } 2) \\
&= \frac{1}{C(\chi,\epsilon)} \int \exp\{\text{sgn}(d)\log\chi y I(|d| \geq \epsilon)\} f_{1,M_1}(y|m_0)\ f_{M_0,M_1}(m_0,m_1)\ dm_0 dm_1 (\text{A } 3),
\end{aligned}$$

where 'A' corresponds to 'Assumption' and $C(\chi, \epsilon)$ denotes a normalizing constant which can be obtained as

$$C(\chi,\epsilon) = \sum_{y=0}^{1} \int \exp\{\text{sgn}(d)\log\chi y I(|d| \geq \epsilon)\} f_{1,M_1}(y|m_0)\ f_{M_0,M_1}(m_0,m_1)\ dm_0 dm_1.$$

To compute $f_{1,M_1}(y|m_0)$,

$$f_{1,M_1}(y|m_0) = \frac{\pi_{1,M_1}^{y}(1-\pi_{1,M_1})^{1-y} f_{M_1,Y_1}(M_1 = m_0|Y_{1,M_1} = y)}{f_{M_1}(M_1 = m_0)},$$

where

$$f_{M_1,Y_1}(M_1 = m_0|Y_{1,M_1} = y) = F_{M_1,y}(m_0 + 0.5) - F_{M_1,y}(m_0 - 0.5)$$

and

$$f_{M_1}(M_1 = m_0) = f_{M_1,Y_1}(M_1 = m_0|Y_{1,M_1} = 1) \times \pi_{1,M_1} + f_{M_1,Y_1}(M_1 = m_0|Y_{1,M_1} = 0) \times (1 - \pi_{1,M_1}).$$

3. Compute the direct and indirect effects using $\pi_{1,M_0} - \pi_{0,M_0}$ and $\pi_{1,M_1} - \pi_{1,M_0}$, where $\pi_{1,M_0} = f_{1,M_0}(1)$.

Implementation of Dirichlet process priors for the TOURS data in WinBUGS (Step 2 of the sampling algorithm in Section 3)

We use the following construction of the Dirichlet process parameters for implementation in WinBUGS,

$$\gamma_i \sim \text{Beta}(1, K_z), \quad \pi_i = \gamma_i \prod_{l=1}^{i-1}(1 - \gamma_l),$$

$$\theta_i \sim W_z \times \text{Beta}_{[0,350]}(\alpha_{1z}, \beta_{1z}) + (1 - W_z) \times \text{Beta}_{[0,350]}(\alpha_{2z}, \beta_{2z}),$$

$$G_z = \sum_{i=1}^{M} \pi_i \delta_{\theta_i} \text{ and } f_{M_z,y}(m_z|Y_{z,M_z} = y) \sim G_z$$

where the precision parameter K_z has a uniform prior, $DiscUnif[1, 20]$.

Since the mediator takes values in [0,350], we specify

$$\lambda_i = Q(\theta_i) \text{ and } G_z = \sum_{i=1}^{M} \pi_i \delta_{\lambda_i}$$

where function $Q : (0, 350) \to (-0.5, 350 + 0.5)$.

We then specify

$$f_{M_z,y}(m_z|Y_{z,M_z} = y) \sim Poisson(\lambda_S),$$

where $S \sim$ categorical$(\pi_1, \pi_2, \cdots, \pi_K)$.

Comparison of posterior variances with and without Assumption 5

a) Without Assumption 5

First note that

$$E(Y_{1,M_1} Y_{1,M_0}) = \int p(Y_{11} = 1, Y_{10} = 1|m_0, m_1) f(m_0, m_1) dm_0 dm_1,$$

where $p(Y_{11} = 1, Y_{10} = 1|m_0, m_1) = p(Y_{10} = 1|Y_{11} = 1, m_0, m_1)p(Y_{11} = 1|m_0, m_1)$.

Now, assume $p(Y_{11} = 1|m_0, m_1) > 0$ and a non-negative correlation between Y_{11} and Y_{10} given m_0 and m_1. Then the correlation between Y_{11} and Y_{10} conditional on (M_1, M_0), θ, is

$$\begin{aligned}
\theta &= \frac{E(Y_{10}Y_{11}|m_0, m_1) - E(Y_{10}|m_0, m_1)E(Y_{11}|m_0, m_1)}{\sqrt{p(Y_{10} = 1|m_0, m_1)(1 - p(Y_{10} = 1|m_0, m_1))}\sqrt{p(Y_{11} = 1|m_0, m_1)(1 - p(Y_{11} = 1|m_0, m_1))}} \\
&= \frac{p(Y_{10} = 1, Y_{11} = 1|m_0, m_1) - p(Y_{10} = 1|m_0, m_1)p(Y_{11} = 1|m_0, m_1)}{\sqrt{p(Y_{10} = 1|m_0, m_1)(1 - p(Y_{10} = 1|m_0, m_1))}\sqrt{p(Y_{11} = 1|m_0, m_1)(1 - p(Y_{11} = 1|m_0, m_1))}} \\
&\geq 0
\end{aligned}$$

Note that

$$p(Y_{10}=1,Y_{11}=1|m_0,m_1)=p(Y_{10}=1|m_0,m_1)p(Y_{11}=1|m_0,m_1)+\theta\times s.d.(Y_{10})s.d.(Y_{11}),$$

where $s.d.(Y_{10})=\sqrt{p(Y_{10}=1|m_0,m_1)(1-p(Y_{10}=1|m_0,m_1))}$ and

$s.d.(Y_{11})=\sqrt{p(Y_{11}=1|m_0,m_1)(1-p(Y_{11}=1|m_0,m_1))}$.

This can be re-expressed as

$$\begin{aligned}&p(Y_{10}=1,Y_{11}=1|m_0,m_1)\\ &\quad= \exp\{sgn(d)\log\chi I(|d|\geq\epsilon)\}p(Y_{11}=1|M_1=m_0)p(Y_{11}=1|M_1=m_1)+\theta\times s.d.(Y_{10})s.d.(Y_{11})\end{aligned}$$

by Assumption 2 and 3.

Using these results, the variance of the NIE without assumption 5 is

$$\begin{aligned}\text{Var}(NIE) &= E(Y_{1,M_1}^2)-2E(Y_{1,M_1}Y_{1,M_0})+E(Y_{1,M_0}^2)-\{E(Y_{1,M_1})-E(Y_{1,M_0})\}^2\\ &= E(Y_{1,M_1}^2)+E(Y_{1,M_0}^2)-\{E(Y_{1,M_1})-E(Y_{1,M_0})\}^2-C1, \qquad (1)\end{aligned}$$

where $C1=2\int\exp\{sgn(d)\log\chi I(|d|\geq\epsilon)\}p(Y_{11}=1|M_1=m_0)p(Y_{11}=1|M_1=m_1)f(m_0,m_1)dm_0dm_1$

$+2\int\theta\times s.d.(Y_{10})s.d.(Y_{11})f(m_0,m_1)dm_0dm_1$.

b) Under Assumption 5

Note that

$$E(Y_{1,M_1}Y_{1,M_0})=\int p(Y_{11}=1,Y_{10}=1|m_0,m_1)f(m_0,m_1)dm_0dm_1,$$

where

$$\begin{aligned}p(Y_{11}=1,Y_{10}=1|m_0,m_1) &= p(Y_{10}=1|m_0,m_1)p(Y_{11}=1|m_0,m_1)\\ &= \exp\{sgn(d)\log\chi I(|d|\geq\epsilon)\}p(Y_{11}=1|M_1=m_0)p(Y_{11}=1|M_1=m_1)\end{aligned}$$

since $\theta = 0$.

Using this result, the variance of the NIE with assumption 5 is

$$\begin{aligned} \text{Var}(NIE_w) &= E(Y_{1,M_1}^2) - 2E(Y_{1,M_1}Y_{1,M_0}) + E(Y_{1,M_0}^2) - \{E(Y_{1,M_1}) - E(Y_{1,M_0})\}^2 \\ &= E(Y_{1,M_1}^2) + E(Y_{1,M_0}^2) - \{E(Y_{1,M_1}) - E(Y_{1,M_0})\}^2 - C2. \end{aligned} \tag{2}$$

where $C2 = 2\int \exp\{sgn(d)\log\chi I(|d| \geq \epsilon)\}p(Y_{11} = 1|M_1 = m_0)p(Y_{11} = 1|M_1 = m_1)f(m_0, m_1)dm_0dm_1$

c) Comparison of Var(NIE) with and without Assumption 5

Comparing (2) and (1), the difference is C1 - C2,

$$\begin{aligned} C1 &= 2\int \exp\{sgn(d)\log\chi I(|d| \geq \epsilon)\}p(Y_{11} = 1|M_1 = m_0)p(Y_{11} = 1|M_1 = m_1)f(m_0, m_1)dm_0dm_1 \\ &\quad +2\theta\int s.d.(Y_{10})s.d.(Y_{11})f(m_0, m_1)dm_0dm_1 \\ &\geq 2\int \exp\{sgn(d)\log\chi I(|d| \geq \epsilon)\}p(Y_{11} = 1|M_1 = m_0)p(Y_{11} = 1|M_1 = m_1)f(m_0, m_1)dm_0dm_1 \\ &= C2 \end{aligned}$$

which is non-negative. Thus, variance of NIE without Assumption 5 has a smaller variance;

$$Var(NIE) < Var(NIE_w),$$

where $Var(NIE)$ denotes the variance of NIE with Assumption 5 and $Var(NIE_w)$ denotes the variance of NIE without Assumption 5.

The difference in the variances, A, is given as

$$A = 2\int \theta \times s.d.(Y_{10})s.d.(Y_{11})f(m_0, m_1)dm_0dm_1,$$

where θ is bounded as below (since binary responses),

$$
\begin{aligned}
\theta &= \frac{E(Y_{10}Y_{11}|m_0,m_1) - E(Y_{10}|m_0,m_1)E(Y_{11}|m_0,m_1)}{s.d.(Y_{10})s.d.(Y_{11})} \\
&\le \frac{\sqrt{E(Y_{10}^2|m_0,m_1)E(Y_{11}^2|m_0,m_1)} - E(Y_{10}|m_0,m_1)E(Y_{11}|m_0,m_1)}{s.d.(Y_{10})s.d.(Y_{11})} \quad \text{Cauchy-Schwarz inequality} \\
&= \frac{\sqrt{p(Y_{10}=1|m_0,m_1)p(Y_{11}=1|m_0,m_1)} - p(Y_{10}=1|m_0,m_1)p(Y_{11}=1|m_0,m_1)}{s.d.(Y_{10})s.d.(Y_{11})} \\
&= \frac{\sqrt{\text{exp}\{sgn(d)\log\chi I(|d|\ge\epsilon)\}p(Y_{11}=1|M_1=m_0)p(Y_{11}=1|M_1=m_1)}}{s.d.(Y_{10})s.d.(Y_{11})} \\
&\quad - \frac{\text{exp}\{sgn(d)\log\chi I(|d|\ge\epsilon)\}p(Y_{11}=1|M_1=m_0)p(Y_{11}=1|M_1=m_1)}{s.d.(Y_{10})s.d.(Y_{11})},
\end{aligned}
$$

where the last equality is from Assumption 2 and 3.

Thus, the difference in the variances, A is bounded by

$$
A \le 2\int(\sqrt{p^*} - p^*)f(m_0,m_1)dm_0dm_1,
$$

where $p^* = \text{exp}\{sgn(d)\log\chi I(|d|\ge\epsilon)\}p(Y_{11}=1|M_1=m_0)p(Y_{11}=1|M_1=m_1)$.

		$\chi = 1$					
		$\epsilon = 50$		$\epsilon = 75$		$\epsilon = 100$	
		NIE	s.d.	NIE	s.d.	NIE	s.d.
$\beta_2 = 0$	Our Approach	0.024	(0.036)	0.033	(0.040)	0.038	(0.039)
	Truth	0.024	(0.036)	0.033	(0.040)	0.038	(0.039)
$\beta_2 = 1.28$	Our Approach	0.004	(0.044)	-0.0008	(0.045)	0.004	(0.041)
	Truth	0.001	(0.043)	-0.004	(0.045)	0.001	(0.041)
$\beta_2 = 2.56$	Our Approach	-0.017	(0.042)	-0.017	(0.033)	-0.028	(0.040)
	Truth	0.003	(0.038)	0.006	(0.031)	-0.006	(0.038)
$\beta_2 = 5.12$	Our Approach	-0.059	(0.037)	-0.065	(0.035)	-0.063	(0.033)
	Truth	0.002	(0.027)	-0.001	(0.028)	0.0007	(0.025)

		$\chi = 1.15$					
		$\epsilon = 50$		$\epsilon = 75$		$\epsilon = 100$	
		NIE	s.d.	NIE	s.d.	NIE	s.d.
$\beta_2 = 0$	Our Approach	0.014	(0.037)	0.022	(0.048)	0.018	(0.040)
	Truth	0.014	(0.037)	0.022	(0.048)	0.018	(0.040)
$\beta_2 = 1.28$	Our Approach	-0.0002	(0.038)	0.001	(0.045)	-0.0001	(0.042)
	Truth	-0.002	(0.038)	-0.001	(0.046)	-0.003	(0.041)
$\beta_2 = 2.56$	Our Approach	-0.016	(0.039)	-0.018	(0.039)	-0.010	(0.037)
	Truth	0.003	(0.036)	0.002	(0.036)	0.008	(0.036)
$\beta_2 = 5.12$	Our Approach	-0.051	(0.039)	-0.049	(0.032)	-0.059	(0.037)
	Truth	0.009	(0.029)	0.006	(0.026)	0.003	(0.028)

		$\chi = 1.3$					
		$\epsilon = 50$		$\epsilon = 75$		$\epsilon = 100$	
		NIE	s.d.	NIE	s.d.	NIE	s.d.
$\beta_2 = 0$	Our Approach	0.021	(0.038)	0.025	(0.038)	0.019	(0.041)
	Truth	0.021	(0.038)	0.025	(0.038)	0.019	(0.041)
$\beta_2 = 1.28$	Our Approach	0.004	(0.038)	0.004	(0.036)	0.002	(0.040)
	Truth	0.002	(0.037)	0.001	(0.036)	0.0003	(0.039)
$\beta_2 = 2.56$	Our Approach	-0.014	(0.037)	-0.011	(0.039)	-0.019	(0.038)
	Truth	0.0007	(0.034)	0.005	(0.037)	-0.003	(0.035)
$\beta_2 = 5.12$	Our Approach	-0.044	(0.037)	-0.050	(0.039)	-0.051	(0.036)
	Truth	0.011	(0.027)	0.006	(0.027)	0.003	(0.027)

		$\chi = 2$					
		$\epsilon = 50$		$\epsilon = 75$		$\epsilon = 100$	
		NIE	s.d.	NIE	s.d.	NIE	s.d.
$\beta_2 = 0$	Our Approach	0.012	(0.041)	0.008	(0.037)	0.005	(0.041)
	Truth	0.012	(0.041)	0.008	(0.037)	0.005	(0.041)
$\beta_2 = 1.28$	Our Approach	0.007	(0.039)	0.003	(0.042)	0.004	(0.039)
	Truth	0.006	(0.039)	0.001	(0.043)	0.002	(0.039)
$\beta_2 = 2.56$	Our Approach	0.008	(0.038)	0.008	(0.036)	-0.001	(0.046)
	Truth	0.014	(0.034)	0.016	(0.036)	0.008	(0.042)
$\beta_2 = 5.12$	Our Approach	-0.005	(0.041)	-0.010	(0.041)	-0.014	(0.038)
	Truth	0.025	(0.031)	0.026	(0.028)	0.019	(0.027)

Table 1: Simulations to assess sensitivity of estimate of NIE to violations in Assumption 3: n=120