Computationally Efficient Analysis of Randomized Trials with Non-Monotone Missing Binary Outcomes

Jaron J. R. Lee[1] (jaron.lee@jhu.edu), Daniel O. Scharfstein[2] (dscharf@jhu.edu) and Ilya Shpitser[1,2] (ilyas@cs.jhu.edu)

Department of Computer Science[1], Department of Biostatistics[2]

JOHNS HOPKINS UNIVERSITY

Summary

- Missing not at random (MNAR) models are the most realistic model class for studying non-monotone clinical trial data (Little and Rubin, 2014)
- (1) introduce an MNAR model where the probability of missingness at a visit depends on all unobserved outcomes prior to the visit, and all observed outcomes after the visit
- **Problem:** This model is intractable for large K without further assumptions
- **Idea:** Develop computational methods to efficiently estimate this model
  - Restrict to binary data
  - Introduce a Markov restriction on dependencies
  - Use directed acyclic graph (DAG) theory to derive a tractable recursive identification and estimation strategy

The Unrestricted Robins Model

For $i = 1, \ldots, K$ time points,

- $Y_i^{(1)} \in \{0, 1\}$
- $R_i \in \{0, 1\}$
- $Y_i \equiv \begin{cases} Y_i^{(1)}, & R_i = 1 \\ ?, & R_i = 0 \end{cases}$
- $O_k = (R_k, Y_k^{(1)})$
- $\overline{z}_k = (z_1, ..., z_{k-1})$ for $k = 2, \ldots, K$,
- $\underline{z}_k = (z_{k+1}, ..., z_K)$ for $k = 1, \ldots, K-1$,
- $\overline{z}_k^m = (z_{\max(1,k-m)}, ..., z_{k-1})$ for $k = 2, \ldots, K$,
- $\underline{z}_k^m = (z_{k+1}, ..., z_{\min(k+m,K)})$ for $k = 1, \ldots, K-1$

Model Assumptions:

$$\forall k \in 1, \ldots, K$$
$$p(Y_k^{(1)} \mid R_k = 0, \overline{Y^{(1)}}_k, \underline{O}_k) = p(Y_k^{(1)} \mid R_k = 1, \overline{Y^{(1)}}_k, \underline{O}_k)$$

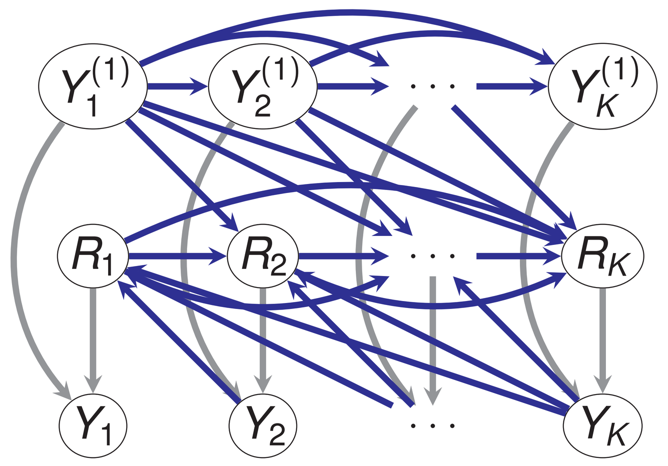

Figure 1: Unrestricted model

Identification in the Unrestricted Model

Theorem 1: $p(\overline{Y^{(1)}}_k)$ in the unrestricted model is identified.

- By induction we can show that $p(\overline{Y^{(1)}}_k, \underline{O}_k)$ is identified.
- For $k = 0$, this is $p(\underline{O}_1)$ which is observed.
- Suppose identified for $k = s$. Then,

$$p(\overline{Y^{(1)}}_s, \underline{O}_s) = \sum_{j=0}^{1} p(\overline{Y^{(1)}}_{s-1}, Y_s^{(1)}, R_s = j, \underline{O}_s)$$

  - $R_s = 1$ is identified by induction.
  - $R_s = 0$ is identified since

$$p(\overline{Y^{(1)}}_{s-1}, Y_s^{(1)}, R_s = 0, \underline{O}_s) = \overbrace{p(Y_s^{(1)} \mid R_s = 0, \overline{Y^{(1)}}_{s-1}, \underline{O}_s)}^{\text{identified by model assumption}} \underbrace{p(\overline{Y^{(1)}}_{s-1}, R_s = 0, \underline{O}_s)}_{\text{identified by induction assumption}}$$

- **Issue:** The proof relies on probability distributions over $O(2^K)$ elements (K binary variables) - inference for large K intractable!

The Restricted Robins Model

- **Solution:** Use Markov restrictions to avoid estimating high-dimensional distributions - related to efficient calculation of causal effects in (2)
- Let m denote the Markov restriction. At time t,

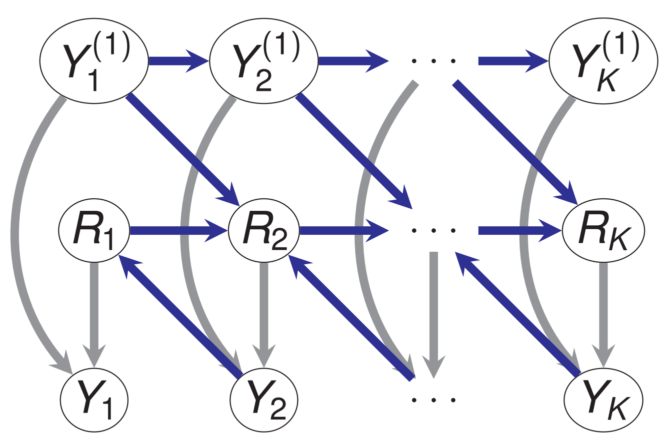

Figure 2: $m = 1$

Model Assumptions:

$$\forall k \in 1, \ldots, K$$
$$p(Y_k^{(1)} \mid R_k = 0, \overline{Y^{(1)}}_k^m, \underline{O}_k^m) = p(Y_k^{(1)} \mid R_k = 1, \overline{Y^{(1)}}_k^m, \underline{O}_k^m)$$
$$p(Y_k^{(1)} \mid \overline{Y^{(1)}}_k) = p(Y_k^{(1)} \mid \overline{Y^{(1)}}_k^m)$$

- **Remark:** Since the unrestricted model is identified, the restricted model must also be identified.
- **Remark:** $p(R_i \mid \overline{Y^{(1)}}_k^m, \underline{O}_k^m)$ is a probability distribution over $O(2^{2m})$ elements - over large K is tractable for fixed m.

Theorem 2:

$$R_k \perp\!\!\!\perp Y_k^{(1)} \mid \overline{Y^{(1)}}_k^m, \underline{O}_k^m \tag{2}$$

Estimation of the Restricted Model

The identification of the unrestricted model offers no insight into the estimation of the restricted model that is linear in K for a fixed m.

- We derive a *recursive* estimation strategy.
  - For fixed m, linear in K.

Base case:

$$p(\overline{Y^{(1)}}_2^{m+1}, \underline{O}_1^{m+1}) = p(O_{m+2} \mid Y_1^{(1)}, \underline{O}_1^m) p(Y_1^{(1)} \mid \underline{O}_1^m) p(\underline{O}_1^m)$$
$$= \underbrace{p(O_{m+2} \mid R_1 = 1, Y_1^{(1)}, \underline{O}_1^m) p(Y_1^{(1)} \mid R_1 = 1, \underline{O}_1^m)}_{\text{by Theorem 2}} p(\underline{O}_1^m)$$

Inductive case: Assume that $p(\overline{Y^{(1)}}_{k+1}^{m+1}, \underline{O}_k^{m+1})$ is identified. Then,

$$p(\overline{Y^{(1)}}_{k+1}^{m+1}, \underline{O}_k^{m+1})$$
$$= \underbrace{p(O_{k+m+1} \mid \overline{Y^{(1)}}_{k+1}^{m+1}, \underline{O}_k^m) p(Y_k^{(1)} \mid \overline{Y^{(1)}}_k^m, \underline{O}_k^m)}_{\text{by Theorem 2}} p(\overline{Y^{(1)}}_k^m, \underline{O}_k^m)$$
$$= \underbrace{\overbrace{p(O_{k+m+1} \mid R_k = 1, \overline{Y^{(1)}}_{k+1}^{m+1}, \underline{O}_k^m) p(Y_k^{(1)} \mid R_k = 1, \overline{Y^{(1)}}_k^m, \underline{O}_k^m) p(\overline{Y^{(1)}}_k^m, \underline{O}_k^m)}}_{\text{function of } p(\overline{Y^{(1)}}_2^{m+1}, \underline{O}_1^{m+1})}$$

We want to show that $p(\overline{Y^{(1)}}_m^m, \underline{O}_{k-1}^{m+2})$ is a function of $p(\overline{Y^{(1)}}_{k+1}^{m+1}, \underline{O}_k^{m+1})$.

$$p(\overline{Y^{(1)}}_k^m, \underline{O}_{k-1}^{m+1}) = \begin{cases} p(\overline{Y^{(1)}}_{k+1}^{m+1}, \underline{O}_k^{m+1}), & 2 \le k \le m+1 \\ \sum_{Y_{k-m-1}^{(1)}} p(\overline{Y^{(1)}}_k^{m+1}, \underline{O}_{k-1}^{m+1}), & m+1 < k \le K+1 \end{cases}$$

$$p(\overline{Y^{(1)}}_m^m, \underline{O}_{k-1}^{m+2}) = \begin{cases} \overbrace{p(O_{k+m+1} \mid \overline{Y^{(1)}}_k^m, \underline{O}_{k-1}^{m+1}, \overline{R}_k^m = \mathbf{1})}^{\text{by Theorem 2}} p(\overline{Y^{(1)}}_k^m, \underline{O}_{k-1}^{m+1}), & k \le K-m-1 \\ p(\overline{Y^{(1)}}_k^m, \underline{O}_{k-1}^{m+1}), & k > K-m-1 \end{cases}$$

Application: Drug Use Abatement Dataset

- **Study:** National Institute of Drug Abuse Study No. CTN-0044 - evaluating effectiveness of computer tool in outpatient substance abuse treatment. Patient dropout is non-monotone, 20-25% missing.
  - $N = 500$, $K = 24$
  - Y: negative drug test (abstinence)
- **Question:** Evaluate effectiveness of treatment vs. control
  - $T = \sum_{k=1}^{24} \mathbb{E}[Y_k^{(1)}]$
- **Estimation:** Discrete probability distributions estimated using random forests
  - Use random forests to model $p(O_K \mid \underline{O}_K^m), p(O_{K-1} \mid \underline{O}_{K-1}^m), \ldots, p(O_1)$
  - Tune random forests to avoid positivity violations

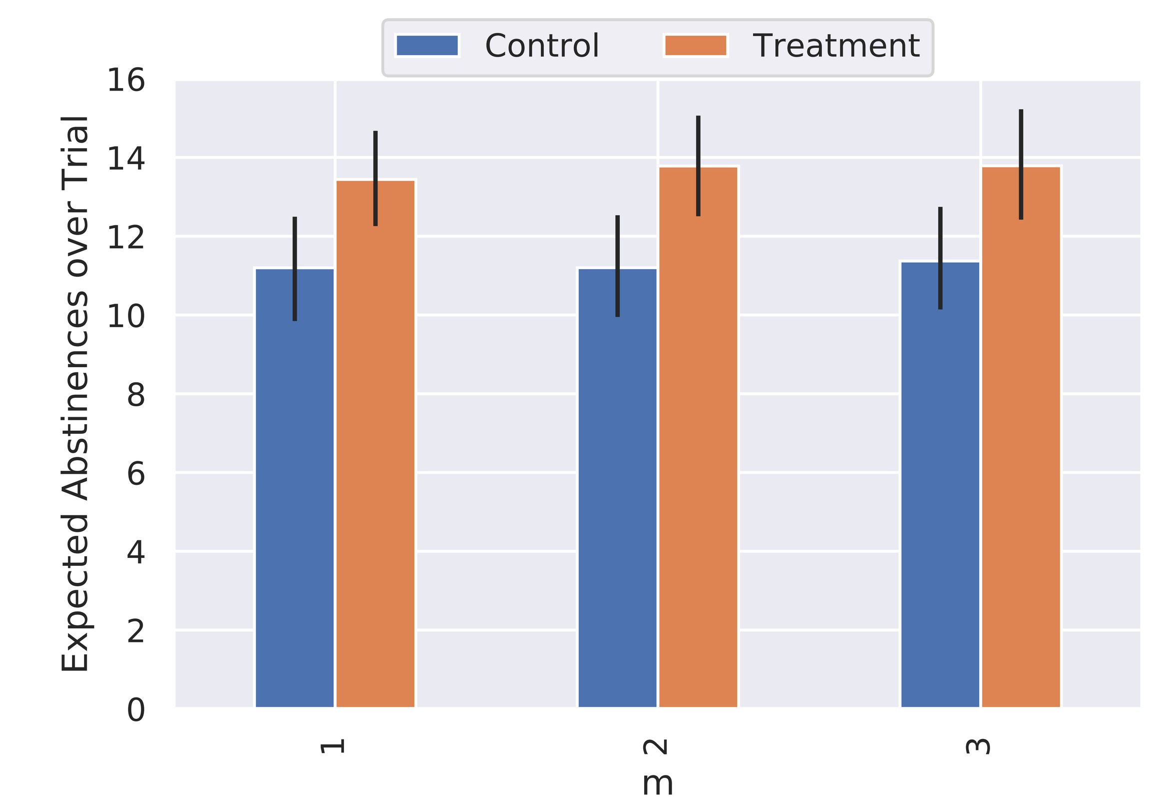

Figure 3: Expected abstinences over trial for varying m, with bootstrapped 95% CI ($N_{\text{bootstrap}} = 500$)

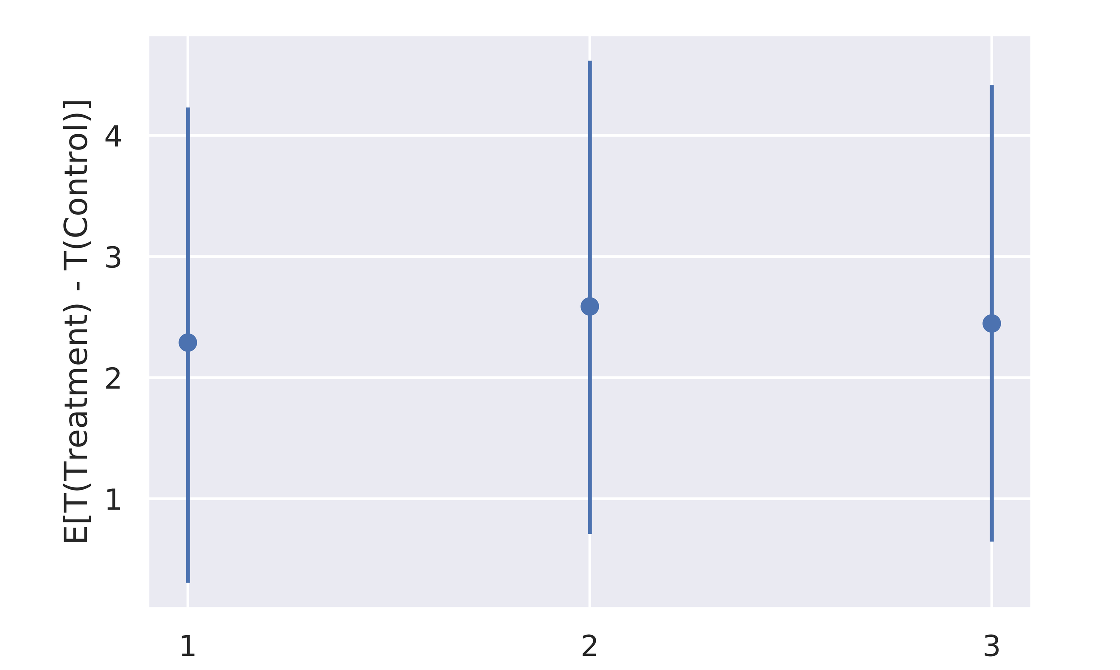

Figure 4: Expected treatment effect for varying m, with bootstrapped 95% CI ($N_{\text{bootstrap=500}}$)

References

[1] Robins, J. M. Non-Response Models for the Analysis of Non-Monotone Non-Ignorable Missing Data. *Statistics in Medicine 16*, 1 (1997), 21–37.

[2] Shpitser, I., Richardson, T. S., and Robins, J. M. An Efficient Algorithm for Computing Interventional Distributions in Latent Variable Causal Models. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence* (2011), UAI'11, AUAI Press, pp. 661–670.