Causal Inference With Outcome-Dependent Missingness And Self-Censoring (Supplementary Material)

Jacob M. Chen[1] **Daniel Malinsky**[2] **Rohit Bhattacharya**[1]

[1]Department of Computer Science , Williams College
[2]Department of Biostatistics , Columbia University

A SPECIFICS OF DATA GENERATING PROCESS

For our simulations, we generate data according to the graph shown in Figure 6 and modifications of it that violate the shadow variable or backdoor conditions. We generate the pre-treatment covariates $\mathbf{W}$ from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma} =$

$$\begin{bmatrix} 1.2 & 0 & 0 & 0 \\ 0 & 1 & 0.4 & 0.4 \\ 0 & 0.4 & 1 & 0.3 \\ 0 & 0.4 & 0.3 & 1 \end{bmatrix}.$$

The above data generating process is equivalent to a structural equation model with correlated errors due to unmeasured confounders between the pairs (W_2, W_3), (W_2, W_4), and (W_3, W_4). We generate $A, Y^{(1)}, I$, and R_Y according to structural equation models following edges in Figure 6. Note that we also clip all probabilities to be between the ranges of 0.01 and 0.99.

We generate A as a binary variable with the following probabilities:

$$p(A = 1 \mid W_1, W_2, W_3, W_4) = \text{expit}(0.52 + 2 * W_1 + 2 * W_2 + 2 * W_3 + 2 * W_4)$$
$$p(A = 0 \mid W_1, W_2, W_3, W_4) = 1 - p(A = 1 \mid W_1, W_2, W_3, W_4)$$

Next, $Y^{(1)}$ is generated similarly with the following probabilities:

$$p(Y^{(1)} = 1 \mid A, W_2, W_3, W_4) = \text{expit}(3 * A + 2 * W_2 + 2 * W_3 + 2 * W_4)$$
$$p(Y^{(1)} = 0 \mid A, W_2, W_3, W_4) = 1 - p(Y^{(1)} = 1 \mid A, W_2, W_3, W_4)$$

The variable I is simply a random normal variable with mean 0 and variance 2, i.e. $I \sim \mathcal{N}(0, 2)$.

We use the odds ratio parameterization to generate R_Y with the following two probabilities. We first specify $p(R_Y = 1 \mid Y^{(1)} = 0, \mathbf{W} \setminus \{W_1\}, I)$, which represents the probability of $R_Y = 1$ when $Y^{(1)}$ is at its chosen reference value of 0. We then use that probability to generate $p(R_Y = 1 \mid Y^{(1)}, \mathbf{W} \setminus \{W_1\}, I)$ at all values of $Y^{(1)}$.

$$p(R_Y = 1 \mid Y^{(1)} = 0, \mathbf{W} \setminus \{W_1\}, I) = \text{expit}(W_2 + W_3 + W_4 + 0.5 * I)$$

$$p(R_Y = 1 \mid Y^{(1)}, \mathbf{W} \setminus \{W_1\}, I) = \frac{p(R_Y = 1 \mid Y^{(1)} = 0, \mathbf{W} \setminus \{W_1\}, I)}{p(R_Y = 1 \mid Y^{(1)} = 0, \mathbf{W} \setminus \{W_1\}, I) + \exp(-1.5 * Y^{(1)}) \times (1 - p(R_Y = 1 \mid Y^{(1)} = 0, \mathbf{W} \setminus \{W_1\}, I))}.$$

In the case where we add $A \rightarrow R_Y$ to Figure 6, we add $1.5 * A$ in the expit function for $p(R_Y = 1 \mid Y^{(1)} = 0, \cdot)$.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

B PROOF OF THEOREM 1

We first note that the presence of R_Y in the numerator ensures that we only use observed rows of data. Further, the propensity scores in the denominator are identified: $p(A \mid \mathbf{Z})$ only depends on observed quantities, and $p(R_Y = 1 \mid A, Y^{(1)}, \mathbf{Z}) = p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z})$ is identified using S1, S2, and the completeness condition. We now prove that the proposed identifying functional is equal to the backdoor adjustment functional and counterfactual mean under the full data law.

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[\frac{R_Y \times \mathbb{I}(A=a) \times Y}{p(R_Y=1 \mid Y^{(1)}, \mathbf{Z}) \times p(A=a \mid \mathbf{Z})}\right] \\
&=^{(1)} \sum_{R_Y, Y^{(1)}, A, \mathbf{Z}, Y} p(R_Y, Y^{(1)}, A, \mathbf{Z}, Y) \times \frac{R_Y \times \mathbb{I}(A=a) \times Y}{p(R_Y=1 \mid Y^{(1)}, \mathbf{Z}) \times p(A=a \mid \mathbf{Z})} \\
&=^{(2)} \sum_{Y^{(1)}, A, \mathbf{Z}} p(R_Y=1, Y^{(1)}, A, \mathbf{Z}) \times \frac{\mathbb{I}(A=a) \times Y^{(1)}}{p(R_Y=1 \mid Y^{(1)}, \mathbf{Z}) \times p(A \mid \mathbf{Z})} \\
&=^{(3)} \sum_{Y^{(1)}, A, \mathbf{Z}} p(R_Y=1 \mid Y^{(1)}, A, \mathbf{Z}) p(Y^{(1)} \mid A, \mathbf{Z}) p(A \mid \mathbf{Z}) p(\mathbf{Z}) \times \frac{\mathbb{I}(A=a) \times Y^{(1)}}{p(R_Y=1 \mid Y^{(1)}, \mathbf{Z}) \times p(A=a \mid \mathbf{Z})} \\
&=^{(4)} \sum_{Y^{(1)}, \mathbf{Z}} p(R_Y=1 \mid Y^{(1)}, \mathbf{Z}) p(Y^{(1)} \mid A=a, \mathbf{Z}) p(A=a \mid \mathbf{Z}) p(\mathbf{Z}) \times \frac{Y^{(1)}}{p(R_Y=1 \mid Y^{(1)}, \mathbf{Z}) \times p(A=a \mid \mathbf{Z})} \\
&=^{(5)} \sum_{Y^{(1)}, \mathbf{Z}} p(Y^{(1)} \mid A=a, \mathbf{Z}) \times p(\mathbf{Z}) \times Y^{(1)} \\
&=^{(6)} \sum_{\mathbf{Z}} \mathbb{E}[Y^{(1)} \mid A=a, \mathbf{Z}] \times p(\mathbf{Z}) =^{(7)} \mathbb{E}[Y^{(a,1)}].
\end{aligned}
$$

In (1) we apply the law of the unconscious statistician; in (2) we evaluate the sum over R_Y and use missing data consistency; in (3) we apply the chain rule of probability; in (4) we evaluate the sum over A and drop A from the propensity score of R_Y due to condition S2; (5) follows from cancellation of common terms in the numerator and denominator; (6) follows from definition of expectation; the last step (7) follows from the fact that $\mathbf{Z}$ satisfies the backdoor conditions B1 and B2. ☐

C PROOF OF EQUATION 3

For completeness, we provide a proof for odds ratio parameterization of the propensity score in (3). First, from Chen [2007] we have an odds ratio factorization of the joint distribution $p(R_Y, Y^{(1)} \mid \mathbf{Z})$ as follows,

$$
p(R_Y, Y^{(1)} \mid \mathbf{Z}) = \frac{p(R_Y \mid Y^{(1)} = y_0, \mathbf{Z}) \times p(Y^{(1)} \mid R_Y = 1, \mathbf{Z}) \times \text{OR}(Y^{(1)}, R_Y \mid \mathbf{Z})}{\sum_{R_Y, Y^{(1)}} p(R_Y \mid Y^{(1)} = y_0, \mathbf{Z}) \times p(Y^{(1)} \mid R_Y = 1, \mathbf{Z}) \times \text{OR}(Y^{(1)}, R_Y \mid \mathbf{Z})}, \tag{1}
$$

where y_0 is a reference value for $Y^{(1)}$ and 1 is the reference value for R_Y, and the denominator of is a normalizing function. Let $\psi = p(R_Y \mid Y^{(1)} = y_0, \mathbf{Z}) \times p(Y^{(1)} \mid R_Y = 1, \mathbf{Z}) \times \text{OR}(Y^{(1)}, R_Y \mid \mathbf{Z})$ and $\psi_1 = \psi|_{R_Y=1}$. As in the main section of the paper, $\pi_0 \coloneqq p(R_Y = 1 \mid Y^{(1)} = y_0, \mathbf{Z})$ and $\eta(Y^{(1)}, \mathbf{Z}) \coloneqq \text{OR}(R_Y = 0, Y^{(1)} \mid \mathbf{Z})$. We present the proof and an explanation of each step below.

(1) and (2) follow from standard laws of probability. In (3), we apply the odds ratio factorization in (1) to both the numerator and the denominator. In (4), we cancel out like terms in both the numerator and denominator. In (5), we simply expand out ψ_1 and ψ according to our previous definitions of these two terms. In (6), we note that $\text{OR}(Y^{(1)}, R_Y = 1 \mid \mathbf{Z})$ has R_Y at its reference value of 1; hence, it is equal to 1. Further, we move the term $p(Y^{(1)} \mid R_Y = 1, \mathbf{Z})$ outside of the sum in the denominator because this term is not a function of R_Y. In (7), we cancel out like terms from the numerator and denominator. Finally, in (8), we explicitly write out the sum over R_Y, which has only two possible values. When $R_Y = 1$, R_Y is at its reference value in the odds ratio, so the odds ratio term disappears, and we are just left with $\pi_0(\mathbf{Z})$. When $R_Y = 0$, we know that $p(R_Y = 0 \mid Y^{(1)} = y_0, \mathbf{Z}) = 1 - \pi_0(\mathbf{Z})$ and that neither $Y^{(1)}$ nor R_Y are at their reference values in the odds ratio term. Therefore, the odds ratio term remains.

Proof.

$$
\begin{aligned}
p(R_Y = 1 \mid Y^{(1)}, \mathbf{Z}) &=^{(1)} \frac{p(R_Y = 1, Y^{(1)} \mid \mathbf{Z})}{p(Y^{(1)} \mid \mathbf{Z})} \\
&=^{(2)} \frac{p(R_Y = 1, Y^{(1)} \mid \mathbf{Z})}{\sum_{R_Y} p(R_Y, Y^{(1)} \mid \mathbf{Z})} \\
&=^{(3)} \frac{\frac{\psi_1}{\sum_{R_Y, Y^{(1)}} \psi}}{\frac{\sum_{R_Y} \psi}{\sum_{R_Y, Y^{(1)}} \psi}} \\
&=^{(4)} \frac{\psi_1}{\sum_{R_Y} \psi} \\
&=^{(5)} \frac{p(R_Y = 1 \mid Y^{(1)} = y_0, \mathbf{Z}) \times p(Y^{(1)} \mid R_Y = 1, \mathbf{Z}) \times \text{OR}(Y^{(1)}, R_Y = 1 \mid \mathbf{Z})}{\sum_{R_Y} p(R_Y \mid Y^{(1)} = y_0, \mathbf{Z}) \times p(Y^{(1)} \mid R_Y = 1, \mathbf{Z}) \times \text{OR}(Y^{(1)}, R_Y \mid \mathbf{Z})} \\
&=^{(6)} \frac{p(R_Y = 1 \mid Y^{(1)} = y_0, \mathbf{Z}) \times p(Y^{(1)} \mid R_Y = 1, \mathbf{Z})}{p(Y^{(1)} \mid R_Y = 1, \mathbf{Z}) \times \sum_{R_Y} p(R_Y \mid Y^{(1)} = y_0, \mathbf{Z}) \times \text{OR}(Y^{(1)}, R_Y \mid \mathbf{Z})} \\
&=^{(7)} \frac{p(R_Y = 1 \mid Y^{(1)} = y_0, \mathbf{Z})}{\sum_{R_Y} p(R_Y \mid Y^{(1)} = y_0, \mathbf{Z}) \times \text{OR}(Y^{(1)}, R_Y \mid \mathbf{Z})} \\
&=^{(8)} \frac{\pi_0(\mathbf{Z})}{\pi_0(\mathbf{Z}) + \eta(Y^{(1)}, \mathbf{Z})(1 - \pi_0(\mathbf{Z}))}
\end{aligned}
$$

□

D ADDITIONAL SIMULATION RESULTS

We report additional simulation results for simulations of the search algorithm described in section 7. Table 1 gives results when using $\alpha = 0.01$ to conclude dependence between two variables.

Sample Size	**Sensitivity**	**Specificity**
500	0.0	0.394
2500	0.269	0.383
5000	0.690	0.717
10000	0.828	0.952

Table 1: Tables showing the accuracy of tests for different sample sizes and $\alpha = 0.01$.

Next, Table 2 gives the results of the simulations when using $\alpha = 0.1$ to conclude dependence between two variables.

Sample Size	**Sensitivity**	**Specificity**
500	0.022	0.363
2500	0.671	0.630
5000	0.837	0.770
10000	0.949	0.857

Table 2: Tables showing the accuracy of tests for different sample sizes and $\alpha = 0.1$.

As the p-value increases, the accuracy of the tests for correctly predicting an adjustment set when one is possible increases. On the other hand, the accuracy of the tests for correctly identifying that there is no possible adjustment set when no such set exists decreases as p-value increases. For all p-values, the accuracy of the tests in general increase as the sample size increases.

Next, we report additional simulation results for estimation of the causal effect described in Section 7. Figure 1 shows the estimation results for sample size 500. When using the correct adjustment method, our practical estimation method is able to

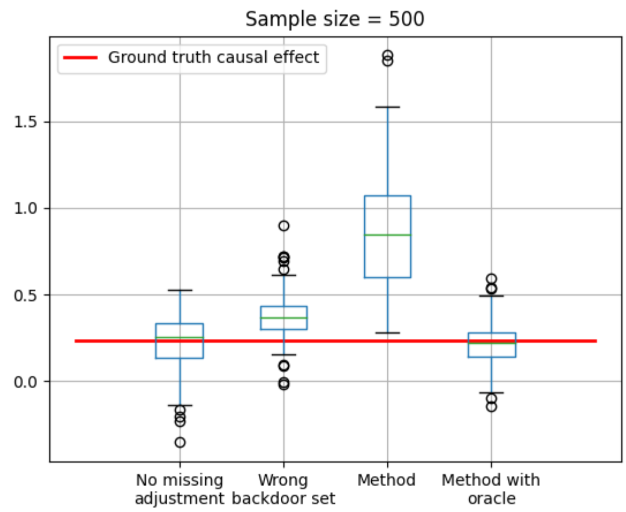

Figure 1: Estimation results for sample size 500.

accurately recover the causal effect. However, using the full pipeline of our method, the estimates are fairly inaccurate. This is to be expected, as the sensitivity of the covariate search at a small sample size is quite inaccurate regardless of p-value. In addition, due to missing data, the effective sample size of the data is roughly 300.

References

Hua Yun Chen. A semiparametric odds ratio model for measuring association. *Biometrics*, 63:413–421, 2007.