Is Rubin's Potential Outcomes Theory Well Defined?

Robert R. Tucci
tucci@ar-tiste.com

August 29, 2021

Abstract

Donald Rubin's Potential Outcomes theory makes two key assumptions that we shall call SUTVA and CIA. In this brief letter, we question whether those two assumptions can hold simultaneously.

1 Introduction

Donald Rubin's Potential Outcomes (PO) theory (a.k.a. Rubin's Causal Model) (Ref. [3]) is a popular method for doing causal inference (CI). PO theory is explained in numerous textbooks (Refs.[2, 1, 4]).

PO theory makes two key assumptions that we shall call SUTVA and CIA. In this brief letter, we question whether those two assumptions can hold simultaneously.

2 Standard PO Assumptions

Standard PO analysis considers random variables $D^\sigma \in \{0,1\}$, X^σ, Y^σ and $\vec{Y}^\sigma = (Y^\sigma(0), Y^\sigma(1))$, where index σ labels the members (individuals, units) of the population (dataset) being considered. These variables are constrained by the following 2 assumptions:

1. SUTVA
$$Y^\sigma = D^\sigma Y^\sigma(1) + (1 - D^\sigma)Y^\sigma(0) \tag{1}$$

2. Conditional Independence Assumption (CIA)
$$Y^\sigma(0), Y^\sigma(1) \perp D^\sigma | X^\sigma \tag{2}$$

By virtue of these 2 assumptions, we have, for $d \in \{0,1\}$,

$$\begin{aligned} E[Y^\sigma | D^\sigma = d, X^\sigma] &= E[Y^\sigma(d) | D^\sigma = d, X^\sigma] \quad \text{(by SUTVA)} \qquad \text{(3a)} \\ &= E[Y^\sigma(d) | X^\sigma] \quad \text{(by CIA)} \qquad \text{(3b)} \end{aligned}$$

In standard PO theory, one defines the Average Treatment Effect (ATE) by

$$ATE \overset{\text{def}}{=} E[Y^\sigma(1) - Y^\sigma(0)] \tag{4}$$

and its x stratum by

$$ATE_x \overset{\text{def}}{=} E[Y^\sigma(1) - Y^\sigma(0) | X^\sigma = x] \tag{5}$$

so that

$$ATE = \sum_x P(x) ATE_x \ . \tag{6}$$

ACE_x is defined by Eq.(5), but by virtue of Eq.3, it also equals

$$ATE_x = E[Y^\sigma | D^\sigma = 1, X^\sigma] - E[Y^\sigma | D^\sigma = 0, X^\sigma] \tag{7}$$

3 Can CIA and SUTVA be satisfied simultaneously?

Throughout the previous section, and in particular in Eqs.(3) and (7), we assumed that CIA and SUTVA can hold simultaneously. Assuming this is standard practice in PO theory. In this section, we question whether that assumption can ever hold.

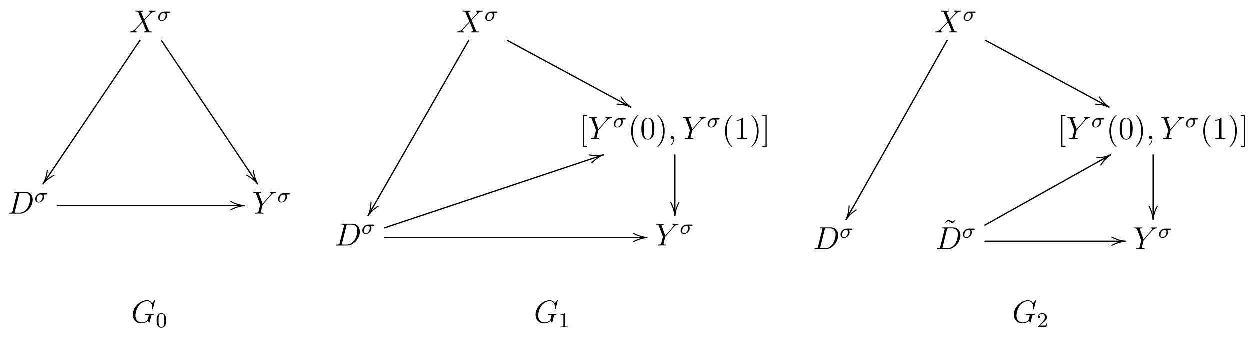

Figure 1: Three Bayesian networks (bnets) that could possibly describe PO theory.

Fig.1 shows 3 Bayesian networks[1] (bnets) labeled G_0, G_1, G_2 that could possibly describe PO theory.[2] The Transition Probability Matrices (TPMs), printed in blue, for the nodes of these 3 bnets, are as follows:

- TPMs for G_0

$$P(x^\sigma) = P_X(x^\sigma) \tag{8a}$$

$$P(d^\sigma|x^\sigma) = P_{D|X}(d^\sigma|x^\sigma) \tag{8b}$$

$$P(y^\sigma|d^\sigma, x^\sigma) = P_{Y|D,X}(y^\sigma|d^\sigma, x^\sigma) \tag{8c}$$

[1] Bayesian networks are extensively discussed by the author of this paper in his textbook Ref.[4]

[2] Remember that bnets are merely a graphical representation of the chain rule for conditional probabilities. Our using bnets in this paper does not constitute assuming anything beyond the axioms of standard probability theory.

- TPMs for G_1

$$P(x^\sigma) = P_X(x^\sigma) \tag{9a}$$

$$P(d^\sigma|x^\sigma) = P_{D|X}(d^\sigma|x^\sigma) \tag{9b}$$

$$P(y^\sigma|d^\sigma,\ \vec{y}^\sigma) = \mathbb{1}(y^\sigma = y^\sigma(d^\sigma)) \tag{9c}$$

For $c \in \{0,1\}$,

$$P(y^\sigma(c)|d^\sigma, x^\sigma) = P_{Y(c)|D,X}(y^\sigma(c)|d^\sigma, x^\sigma) \tag{9d}$$

- TPMs for G_2

$$P(x^\sigma) = P_X(x^\sigma) \tag{10a}$$

$$P(d^\sigma|x^\sigma) = P_{D|X}(d^\sigma|x^\sigma) \tag{10b}$$

$$P(y^\sigma|\tilde{d}^\sigma,\ \vec{y}^\sigma) = \mathbb{1}(y^\sigma = y^\sigma(\tilde{d}^\sigma)) \tag{10c}$$

For $c \in \{0,1\}$,

$$P(y^\sigma(c)|\tilde{d}^\sigma, x^\sigma) = P_{Y(c)|\tilde{D},X}(y^\sigma(c)|\tilde{d}^\sigma, x^\sigma) \tag{10d}$$

$$P(\tilde{d}^\sigma) = P_{\tilde{D}}(\tilde{d}^\sigma) \tag{10e}$$

Now consider Table 1. In that table,

- G_0? is NA for all 3 PO assumptions because G_0 does not contain nodes for $Y^\sigma(0)$ and $Y^\sigma(1)$ and these appear in the 3 PO assumptions.

PO assumption	G_0?	G_1?	G_2?
$E[Y^\sigma(1)\|D^\sigma=1,X^\sigma]=E[Y^\sigma(1)\|X^\sigma]$ (CIA)	NA	No	Yes
$E[Y^\sigma\|D^\sigma=1,X^\sigma]=E[Y^\sigma(1)\|D^\sigma=1,X^\sigma]$ (SUTVA)	NA	Yes (Eq.(9c))	No
$E[Y^\sigma\|\tilde{D}^\sigma=1,X^\sigma]=E[Y^\sigma(1)\|\tilde{D}^\sigma=1,X^\sigma]$ (SUTVA∼)	NA	NA	Yes (Eq.(10c))

Table 1: "NA" means not applicable. "Yes" means that the graph satisfies the PO assumption, and "No" means that it doesn't.

- The D in SUTVA is replaced by a $\tilde{D}$ in SUTVA∼.
- G_1? is NA for SUTVA∼ because G_1 doesn't have a $\tilde{D}^\sigma$ node.
- The entries for the CIA row are a consequence of Pearl's d-separation theorem.
- Two "Yes" entries are justified by referring to an equation.

As told by Table 1, G_0, G_1 and G_2 all violate either SUTVA or CIA. G_2 doesn't satisfy both CIA and SUTVA, but it does satisfy CIA and a modified version of SUTVA that we call SUTVA∼.

References

[1] Scott Cunningham. *Causal inference: The mixtape*. Yale University Press, 2021. https://mixtape.scunning.com/index.html.

[2] Matheus Facure Alves. *Causal Inference for The Brave and True*. 2021. https://matheusfacure.github.io/python-causality-handbook/landing-page.html.

[3] Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[4] Robert R. Tucci. Bayesuvius (book). https://github.com/rrtucci/Bayesuvius/raw/master/main.pdf.