Under confounding, KLM conditions are not necessary for consistency

Johann Gaebler
Stanford University

William Cai
Stanford University

Guillaume Basse
Stanford University

Ravi Shroff
New York University

Sharad Goel
Stanford University

Jennifer Hill
New York University

June 25, 2020

CORRECTION: The example given here does not capture the $Z \to M$ dependence in the DAG below. For a modified example that satisfies this condition, see: https://5harad.com/papers/klm-example-v2.pdf.

Overview

The following example illustrates the fact that Assumptions 1–4 in Knox et al. [2020] are not necessary assumptions for the "naive estimator" (i.e., the stratified difference in means) to yield a consistent estimate of the CDE_{Ob}, even when there is unobserved confounding between the outcome, Y, and the arrest decision, M. This example is virtually identical to Case 2 in the proof of Theorem 9 in our paper [Gaebler et al., 2020].

More specifically, we give an example in which: (1) the joint distribution of random variables is captured by the following causal DAG:

$$\begin{tikzcd}
 & U \arrow[dl] \arrow[dr] & \\
Z \arrow[r] \arrow[rr, bend right] & M \arrow[r] & Y
\end{tikzcd}$$

(2) the conjunction of Assumptions 1–4 is violated; and (3) the stratified difference in means is a consistent estimator of the CDE_{Ob}.

Construction of the Example

We set $X = 1$ (i.e., X is constant), and define Z, $M(b)$, $M(w)$, $Y(z, m)$, and U to all be binary variables. The joint distribution of the variables in our example factors as follows:

$$\begin{aligned}
&\Pr(Z = z, U = u, M(b) = m_b, M(w) = m_w, Y(z, m) = y_{zm}) \\
&\quad = \Pr(Y(z, m) = y_{zm} \mid U = u) \cdot \Pr(Z = z) \cdot \Pr(U = u) \\
&\qquad \cdot \Pr(M(b) = m_b \mid Z = z, U = u) \\
&\qquad \cdot \Pr(M(w) = m_w \mid Z = z, U = u).
\end{aligned} \tag{1}$$

Here we suppress the argument u in the expressions $Y(z,m)$ and $M(z)$ for consistency with our earlier notation.

Now, we set $\Pr(Z = z) = \frac{1}{2}$ for $z \in \{w, b\}$ and $\Pr(U = u) = \frac{1}{2}$ for $u \in \{0, 1\}$. Then, the conditional distributions of the remaining variables are defined as follows:[1]

$$\Pr(M(z) = 1 \mid Z, U) = \begin{cases} \frac{1}{4} \cdot (1+U) \cdot (1 + \mathbb{1}_{Z=w}) & z = b, \\ \frac{1}{8} \cdot (1+U) \cdot (1 + \mathbb{1}_{Z=w}) & z = w. \end{cases} \tag{2}$$

Likewise,[2]

$$\Pr(Y(z,m) = 1 \mid U) = \begin{cases} \frac{1}{2}(1+U) & z = b,\ m = 1 \\ \frac{1}{4}(1+U) & z = w,\ m = 1 \\ 0 & m = 0. \end{cases} \tag{3}$$

Together, Eqs. (1), (2), and (3), along with the aforementioned distributions of Z and U, fully define the joint probability distribution. Lastly, we set $M = M(Z, U)$ and $Y = Y(Z, M, U)$.

Analysis

"Necessary assumptions" are violated

This example does not satisfy treatment ignorability, as $M(z) \not\perp\!\!\!\perp Z$, contrary to Assumption 4(a) in Knox et al. [2020]. (Because X is constant, we need not condition on it when evaluating the treatment ignorability criterion.) Therefore, in particular, this example does not satisfy the conjunction of Assumptions 1–4.

To see this, we compare $\Pr(M(w) = 1 \mid Z = w)$ and $\Pr(M(w) = 1 \mid Z = b)$.

$$\begin{aligned} \Pr(M(w) = 1 \mid Z = w) &= \sum_{u \in \{0,1\}} \Pr(M(w) = 1 \mid Z = w, U = u) \cdot \Pr(U = u) \\ &= \left(\frac{1}{8} \cdot (1+1) \cdot (1+1) \cdot \frac{1}{2}\right) + \left(\frac{1}{8} \cdot (1+0) \cdot (1+1) \cdot \frac{1}{2}\right) \\ &= \frac{1}{4} + \frac{1}{8} \\ &= \frac{3}{8} \\ \Pr(M(w) = 1 \mid Z = b) &= \sum_{u \in \{0,1\}} \Pr(M(w) = 1 \mid Z = b, U = u) \cdot \Pr(U = u) \\ &= \left(\frac{1}{8} \cdot (1+1) \cdot (1+0) \cdot \frac{1}{2}\right) + \left(\frac{1}{8} \cdot (1+0) \cdot (1+0) \cdot \frac{1}{2}\right) \\ &= \frac{1}{8} + \frac{1}{16} \\ &= \frac{3}{16} \end{aligned}$$

Another way to see this is to note, as in Footnote 1, that $\Pr(M(z) = 1 \mid Z = w) = 2 \cdot \Pr(M(z) = 1 \mid Z = b)$ by Eq. (2).

[1] This joint distribution means, in effect, that both of the following are true: (1) $M(b)$ is twice as likely to be 1 as $M(w)$—i.e., an individual is twice as likely to be arrested if they were counterfactually Black than white—so there is discrimination in the arrest decision; and (2) $\Pr(M(z) = 1 \mid Z = w) = 2 \cdot \Pr(M(z) = 1 \mid Z = b)$. Charge probabilities are affected by the confound U.

[2] This joint distribution means that $Y(b, 1)$ is twice as likely to be 1 as $Y(w, 1)$—i.e., an individual is twice as likely to be charged if they are Black than if they are white. Again, charge probabilities are affected by the confound U.

Subset ignorability holds

However, despite the unobserved confounding between M and Y, subset ignorability still holds in this example. To see this, note that by Eq. (1):

$$\begin{aligned}
\Pr(Y(z,m) = 1, Z = z', M = 1) \\
&= \Pr(Y(z,m) = 1) \cdot \Pr(M(z') = 1 \mid Z = z') \cdot \Pr(Z = z') \\
&= \sum_{u \in \{0,1\}} \big(\Pr(Y(z,m) = 1 \mid U = u) \cdot \Pr(U = u) \\
&\qquad \cdot \Pr(M(z') = 1 \mid Z = z', U = u) \cdot \Pr(Z = z') \big)
\end{aligned} \tag{4}$$

Now, assume $m = 1$. Then, if $z = b$ and $z' = b$, Eq. (4) equals

$$\begin{aligned}
\left(\frac{1}{2}(1+1) \cdot \frac{1}{2} \cdot \frac{1}{4}(1+1)(1+0) \cdot \frac{1}{2}\right) + \left(\frac{1}{2}(1+0) \cdot \frac{1}{2} \cdot \frac{1}{4}(1+0)(1+0) \cdot \frac{1}{2}\right) &= \frac{1}{8} + \frac{1}{32} \\
&= \frac{5}{32}
\end{aligned}$$

where here we use the definitions in Eqs. (2) and (3). If $z = b$ and $z' = w$, then Eq. (4) equals

$$\begin{aligned}
\left(\frac{1}{2}(1+1) \cdot \frac{1}{2} \cdot \frac{1}{8}(1+1)(1+1) \cdot \frac{1}{2}\right) + \left(\frac{1}{2}(1+0) \cdot \frac{1}{2} \cdot \frac{1}{8}(1+0)(1+1) \cdot \frac{1}{2}\right) &= \frac{1}{8} + \frac{1}{32} \\
&= \frac{5}{32}
\end{aligned}$$

Virtually identical calculations show that if $z = w$ and $z' = b$, Eq. (4) equals $\frac{5}{64}$; and if $z = w$ and $z' = w$, then Eq. (4) equals $\frac{5}{64}$ as well. (One could also see this from the fact that $Y(b,1)$ is twice as likely to be 1 as $Y(w,1)$; see Foonote 2.)

Next, we see that

$$\begin{aligned}
\Pr(Z = b, M = 1) &= \Pr(Z = b, M(b) = 1) \\
&= \sum_{u \in \{0,1\}} \Pr(Z = b, M(b) = 1 \mid U = u) \cdot \Pr(U = u) \\
&= \sum_{u \in \{0,1\}} \Pr(M(b) = 1 \mid Z = b, U = u) \cdot \Pr(Z = b) \cdot \Pr(U = u) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} \\
&= \frac{3}{16}
\end{aligned}$$

while

$$\begin{aligned}
\Pr(Z = w, M = 1) &= \Pr(Z = w, M(w) = 1) \\
&= \sum_{u \in \{0,1\}} \Pr(Z = w, M(w) = 1 \mid U = u) \cdot \Pr(U = u) \\
&= \sum_{u \in \{0,1\}} \Pr(M(w) = 1 \mid Z = w, U = u) \cdot \Pr(Z = w) \cdot \Pr(U = u) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} \\
&= \frac{3}{16}.
\end{aligned}$$

Therefore

$$
\begin{aligned}
\Pr(Y(b,1)=1 \mid Z=b, M=1) &= \frac{\Pr(Y(b,1)=1, Z=b, M=1)}{\Pr(Z=b, M=1)} \\
&= \frac{5 \ / \ 32}{3 \ / \ 16} \\
&= \frac{5}{6}, \\
\Pr(Y(b,1)=1 \mid Z=w, M=1) &= \frac{\Pr(Y(b,1)=1, Z=w, M=1)}{\Pr(Z=w, M=1)} \\
&= \frac{5 \ / \ 32}{3 \ / \ 16} \\
&= \frac{5}{6}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
\Pr(Y(w,1)=1 \mid Z=b, M=1) &= \frac{\Pr(Y(w,1)=1, Z=b, M=1)}{\Pr(Z=b, M=1)} \\
&= \frac{5 \ / \ 64}{3 \ / \ 16} \\
&= \frac{5}{12}, \\
\Pr(Y(w,1)=1 \mid Z=w, M=1) &= \frac{\Pr(Y(w,1)=1, Z=w, M=1)}{\Pr(Z=w, M=1)} \\
&= \frac{5 \ / \ 64}{3 \ / \ 16} \\
&= \frac{5}{12}.
\end{aligned}
$$

This is equivalent to the statement that $Y(z,1)$ is independent of Z given $M=1$, i.e., $Y(z,1) \perp\!\!\!\perp Z \mid M=1$. Therefore subset ignorability is satisfied, and so Δ_n is a consistent estimator of the $\textsc{cde}_{\mathrm{Ob}}$.

References

J. Gaebler, W. Cai, G. Basse, R. Shroff, S. Goel, and J. Hill. Deconstructing claims of post-treatment bias in observational studies of discrimination. Available at: https://5harad.com/papers/post-treatment-bias.pdf, 2020.

D. Knox, W. Lowe, and J. Mummolo. Administrative records mask racially biased policing. *American Political Science Review*, 2020.