**METHODS FOR CAUSAL INFERENCE: TUTORIAL 2**

* = For formative assignment (marked but not evaluated)

1. In Lecture 4, we used a dataframe consisting of 1000 individuals, together with values for their observed treatments and outcomes as well as the corresponding counterfactual values. We showed

$$\mathbb{E}[Y_1 - Y_0] \neq \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] \tag{1}$$

   and discussed the reason behind this inequality, *i.e.* , a confounding variable. Considering the usual confounding diagram, perform a simulation (in Python or R), to generate a similar table and show the above inequality.

   **Hint:** You may choose $W \sim \mathcal{N}(\mu, \sigma)$. To obtain $T$, you may add a gaussian noise to $W$ and then binarise using a cutoff. Then generate $Y$ as a linear function of $W$ and $T$, together with a guassian noise.

2. For the example above, use regression adjustment (to adjust for the confounder $W$) and obtain an estimate of the causal effect $\mathbb{E}_W[\mathbb{E}_Y[Y|T=1,W] - \mathbb{E}_Y[Y|T=0,W]]$. Compare your answer to the results obtained in Question 1.

3. * Explain the issue of positivity violation (no overlap) when performing extrapolation with no support in each of the following cases and discuss why positivity violation will lead to incorrect causal inference.


FIG. 1: Three scenarios with positivity violation

**Solution:** Positivity is violated in all 3 cases. Naive linear extrapolation in the first case may result in

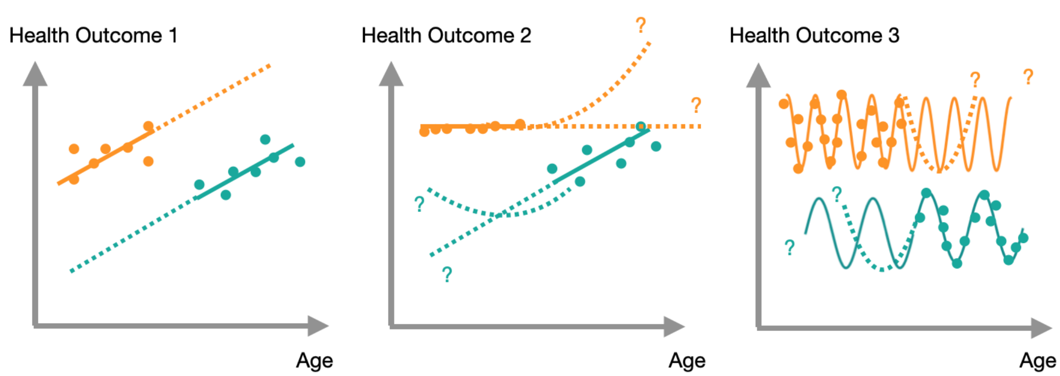


FIG. 2: Three scenarios with positivity violation

the correct answer but we immediately see that in the second and third cases this will certainly not hold. Extrapolation without support means we have no evidence to assume a particular form to fit the data. So fitting in these cases is as good as merely 'guessing'. This 'guess' may or may not turn out to be correct. Therefore, causal inference in these scenarios may well become invalid.

4. Recall the adjustment formula:

$$\mathbb{E}[Y_1 - Y_0] = \mathbb{E}_X\left[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\right]. \tag{2}$$

By expanding out as a sum both the inner expectation over X and the outer expectation over Y on the right hand side, followed by the product rule, show that violation of positivity *i.e.* , $p(T=1|X=x)=0$ leads to an inestimable ATE.

**Solution:** Write out the right hand side:

$$\mathbb{E}_X\left[\mathbb{E}[Y|T=1,X] - \mathbb{E}[Y|T=0,X]\right] = \sum_x p(X=x)\left(\sum_y y\; p(Y=y|T=1,X=x) - \sum_y y\; p(Y=y|T=0,X=x)\right) \tag{3}$$

$$= \sum_x p(X=x)\left(\sum_y y\;\frac{p(Y=y,T=1,X=x)}{p(T=1|X=x)p(X=x)} - \sum_y y\;\frac{p(Y=y,T=0,X=x)}{p(T=0|X=x)p(X=x)}\right).$$

We see when the terms $p(T=1|X=x)=0$ and $p(T=0|X=x)=0$, the fraction becomes ill-defined, as a result ATE become inestimable. Even when these probabilities are not exactly zero but very small, the ATE estimation becomes unstable.

5. Prove the following statements:

(a) For a binary variable $X$ show that $p[X=1|Y] = \mathbb{E}[X|Y]$.

(b) Law of iterated expectations, $\mathbb{E}_Z[Z] = \mathbb{E}_W[\mathbb{E}_{Z|W}[Z|W]]$.

(c) Law of iterated expectations (conditional), $\mathbb{E}[Z|W] = \mathbb{E}_{V|W}\left[\mathbb{E}[Z|W,V]\middle|W\right]$.

**Solution:**

(a) $\mathbb{E}[X|Y] = \sum_{x=0,1} x\; p(X=x|Y) = 0\times p(X=0|Y) + 1\times p(X=1|Y) = p(X=1|Y)$

(b) Write out the right hand side:

$$\begin{aligned}
\mathbb{E}_W[\mathbb{E}_{Z|W}[Z|W]] &= \sum_w\left[\sum_z z\; p(Z=z|W=w)\right]p(W=w)\\
&= \sum_{w,z} z\;\frac{p(z,w)}{p(w)}p(w)\\
&= \sum_z z\; p(z) = \mathbb{E}_Z[Z].
\end{aligned}$$

(c) Again, write out the (more complicated) right hand side and simplify:

$$\begin{aligned}
\mathbb{E}_{V|W}[\mathbb{E}[Z|W,V]\big|W] &= \sum_v\left[\sum_z z\; p(Z=z|W,V=v)\right]p(V=v|W)\\
&= \sum_{v,z} z\;\frac{p(Z=z,V=v|W)}{p(V=v|W)}p(V=v|W)\\
&= \sum_{v,z} z\; p(Z=z,V=v|W)\\
&= \sum_z z\; p(Z=z|W)\\
&= \mathbb{E}[Z|W].
\end{aligned}$$

6. *

(a) Show that in a multiple regression $Y = \alpha_0 + \beta_1 X_1 + \beta_2 X_2$, where $X_1$ and $X_2$ are **independent** of each other, one can obtain a consistent estimate for $\beta_1$ without needing to fit for $\beta_2$. In other words, fitting $Y = \alpha_0 + \beta_1 X_1$ is sufficient for obtaining $\beta_1$. To this by minimising the squared loss function.

(b) Use the above fact, and Structural Causal Equations for the following instrumental variable graph, to show that the least-squares method is justified for obtaining the causal effect $\tau$ of $T$ on $Y$.

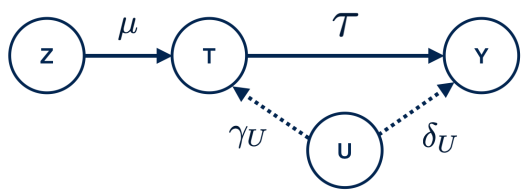


FIG. 3: Instrumental variable graph

(c) Perform a ground truth simulation based on the above SCM, fixing the values $\mu, \tau, \gamma_U$ and $\delta_U$. Then use the simulated data (blind to variable $U$), to show the 2-step least-squares procedure indeed results in the correct estimate of the causal effect $\tau$.

**Solution:**

(a) In least-squares regression we wish to minimise:

$$\text{argmin}_{\alpha_0,\beta_1\beta_2} \mathbb{E}\left[(Y - \alpha_0 - \beta_1 X_1 - \beta_2 X_2)^2\right]. \tag{4}$$

Minimising with respect to $\beta_1$ yields

$$\frac{\partial}{\partial \beta_1}\mathbb{E}\left[\beta_1^2 X_1^2 - 2\beta_1 Y X_1 + 2\alpha_0\beta_1 X_1 + 2\beta_1\beta_2 X_1 X_2 + (\text{terms independent of } \beta_1)\right] = 0, \tag{5}$$

which implies

$$\mathbb{E}\left[2\beta_1 X_1^2 - 2Y X_1 + 2\alpha_0 X_1 + 2\beta_2 X_1 X_2\right] = 0. \tag{6}$$

Rearranging gives:

$$\beta_1 = \frac{\mathbb{E}\left[X_1 Y - \alpha_0 X_1 - \beta_2 X_1 X_2\right]}{\mathbb{E}\left[X_1^2\right]} = \frac{\mathbb{E}\left[X_1 Y\right]}{\mathbb{E}\left[X_1^2\right]} - \alpha_0\frac{\mathbb{E}\left[X_1\right]}{\mathbb{E}\left[X_1^2\right]} - \beta_2\frac{\mathbb{E}\left[X_1\right]\mathbb{E}\left[X_2\right]}{\mathbb{E}\left[X_1^2\right]}, \tag{7}$$

where in the last term we have used that fact that $X_1$ and $X_2$ are independent, which implies the expectation of the product of variables factors as $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$. (Do you see why?)
Performing the minimisation over $\alpha_0$ yields,

$$\frac{\partial}{\partial \alpha_0}\mathbb{E}\left[\alpha_0^2 - 2\alpha_0 Y + 2\alpha_0\beta_1 X_1 + 2\alpha_0\beta_2 X_2 + (\text{terms independent of } \alpha_0)\right] = 0, \tag{8}$$

which implies

$$\alpha_0 = \mathbb{E}[Y] - \beta_1\mathbb{E}[X_1] - \beta_2\mathbb{E}[X_2]. \tag{9}$$

Plugging this back into Eq. **??**, we obtain

$$\beta_1 = \frac{\mathbb{E}\left[X_1 Y\right] - \mathbb{E}\left[X_1\right]\mathbb{E}\left[Y\right]}{\mathbb{E}\left[X_1^2\right]} - \beta_1\frac{\mathbb{E}\left[X_1\right]^2}{\mathbb{E}\left[X_1^2\right]} + 0. \tag{10}$$

Solving this equation for $\beta_1$ shows that

$$\beta_1 = \frac{\mathrm{Cov}[X_1, Y]}{\mathrm{Var}[X_1]}. \tag{11}$$

We conclude that to estimate $\beta_1$ we only require the variance in $X_1$ and the covariance of $X_1$ with $Y$. We do not require any information about $X_2$, because $X_1 \perp\!\!\!\perp X_2$.

(b) For the graph in Fig. **??**, we have the following structural casual equations

$$T := \mu Z + \gamma_U U \tag{12}$$
$$Y := \tau T + \delta_U U \tag{13}$$

We can perform linear regression for the top equation to consistently estimate $\mu$. This gives us the estimate $\hat{T} = \mu Z$, which is a function of $Z$ only. (Crucially, it is *not* a function of $U$.) We can then plug this into the bottom equation to get:

$$Y := \tau(\mu Z + \gamma_U U) + \delta_U U = \tau \underbrace{(\mu Z)}_{\hat{T}} + (\tau\gamma_U + \delta_U)U. \tag{14}$$

Regressing Y on the predicted value $\hat{T}$ from the first step, results in the causal effect size $\tau$.