Estimating average treatment effect with latent variables

Iyar Lin

13 December, 2018

Contents

Intro **1**

Define model graph **1**

1) When $p(w|u)$ is known **2**

Intro

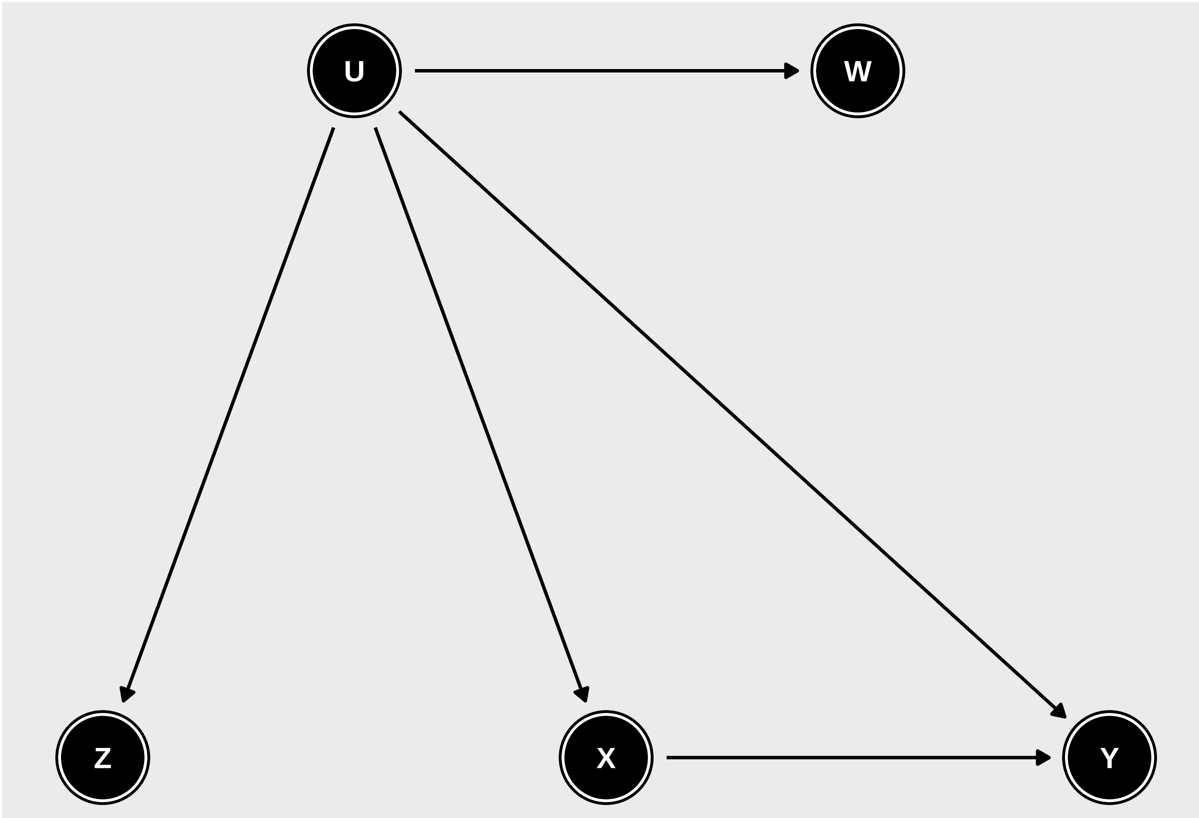
In some cases the backdoor criteria might indicate we need to condition on an unmeasured (latent) variable. Often times we'll have proxy variables (i.e. noisy measurments of the latent variable) at hand. In this script I demonstrate use cases where those proxy variables can be used to estimate the ATE (for details about the theory and background see "M. Kuroki and J. Pearl. Measurement bias and effect restoration in causal inference. Technical report, DTIC Document, 2011.")

Define model graph

We consider the following grpahical model:

```
g <- dagify(Y ~ U + X,
            X ~ U,
            W ~ U,
            Z ~ U,
            coords = data.frame(name = c("Z", "U", "X", "W", "Y"),
                                x = c(-1, 0, 1, 2, 3),
                                y = c(0, 1, 0, 1, 0)),
            exposure = "X",
            outcome = "Y")

ggdag(tidy_dagitty(g))
```



1) When $p(w|u)$ is known

We'll ignore Z in this example.

Under the following conditions the ATE can be estimated:

1. The distribution $p(w|u)$ is known
2. W and the confounder U are discrete variables with a given finite number of categories k

We thus assume the following toy model with $k = 3$:

$$U \in \{a, b, c\},\ W \in \{a, b, c\}$$

and the conditional probabilities $p(w|u)$ are given by the following matrix:

$$M = \begin{pmatrix} 0.2 & 0.1 & 0.5 \\ 0.3 & 0.9 & 0.1 \\ 0.7 & 0.6 & 0.1 \end{pmatrix}$$

where $M_{1,2} = p(W = a|U = b) = 0.3$ etc.

We also assume:

$$Y = 0.2X + 0.4I_{\{U=a\}} - 0.2I_{\{U=b\}} + 0.3I_{\{U=c\}} + \epsilon$$
$$p(U = a) = 0.3,\ p(U = b) = 0.2,\ p(U = c) = 0.5$$

$$X = 0.1I_{\{U=a\}} - 0.4I_{\{U=b\}} + 0.7I_{\{U=c\}} + \epsilon$$

Below I simulate the dataset: