

## Solution to study question 3.5.1

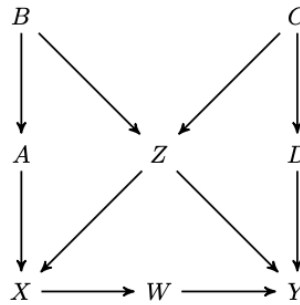
November 28, 2019

(Source: Pearl et al, 2016, solution manual)

The graph of this exercise is available at [dagitty.net/m331](http://dagitty.net/m331). Students can solve parts (a) and (b) interactively in class by forcing adjustment for single covariates (move mouse pointer over the variable and press "a" key).

An R solution of this exercise is provided at [dagitty.net/primer/3.5.1](http://dagitty.net/primer/3.5.1)

Repeating Figure 3.8 for ease of reference:



### Part (a)

By Rule 2, we must adjust for a set of variables that satisfies the backdoor criterion, conditional on  $C$ . We observe that when we condition on  $C$ , there is still an open backdoor from  $X \leftarrow Z \rightarrow Y$ , which we can block by conditioning on  $Z$ . So, we may claim that:

$$P(Y = y | do(X = x), C = c) = \sum_z P(Y = y | X = x, Z = z, C = c) P(Z = z)$$

Above, Rule 2 is applicable because the set  $\{Z, C\}$  satisfies the backdoor criterion to assess the  $c$ -specific effect of  $X$  on  $Y$ .

**Part (b)**

---

Again using Rule 2, we see that  $\{X, Y, Z, C\}$  is such a set since  $\{Z, C\}$  satisfies the backdoor criterion. We can then write:

$$P(Y = y|do(X = x), Z = z) = \sum_c P(y|x, z, c)P(c)$$

Advanced students may be challenged to show that  $\{X, Y, Z, W\}$  is also such a set, since  $W$  satisfies the front-door criterion when  $Z$  is specified.

**Part (c)**

---

Since our choice of  $X$  relies upon the value of  $Z$ , we need to adopt the conditional policy  $do(X = g(Z))$ , where:

$$g(Z) = \begin{cases} 0 & Z \leq 2 \\ 1 & Z > 2 \end{cases}$$

We then assume that  $Z \in \{1, 2, 3, 4, 5\}$ , so by Eq. (3.17) we have:

$$\begin{aligned} P(Y = y|do(X = g(Z))) &= \sum_z P(Y = y|do(X = g(z)), Z = z)P(Z = z) \\ &= P(Y = y|do(X = 0), Z = 1)P(Z = 1) \\ &\quad + P(Y = y|do(X = 0), Z = 2)P(Z = 2) \\ &\quad + P(Y = y|do(X = 1), Z = 3)P(Z = 3) \\ &\quad + P(Y = y|do(X = 1), Z = 4)P(Z = 4) \\ &\quad + P(Y = y|do(X = 1), Z = 5)P(Z = 5) \end{aligned}$$

So, for each term above in the format  $P(Y = y|do(X = x), Z = z)$ , we can substitute our findings from (b) to find an expression free of the  $do$ -operator.