Homework 11: Use and Abuse of Conditioning

36-402, Advanced Data Analysis

Due at the start of class, 26 April 2011

1. (30 points) Refer to figure 1 in Homework 10.
   (a) (5 points) Using the back door criterion, describe a way to estimate the causal effect of smoking on cancer.
   (b) (5 points) Using the front door criterion, describe a *different* way to estimate the causal effect of smoking on cancer.
   (c) (5 points) Is there a way to use instrumental variables to estimate the causal effect of smoking on cancer in this model? Explain.
   (d) (5 points) Using your back-door identification strategy and the data file from last time, estimate $\Pr(cancer = 1|do(smoking = 1.5))$.
   (e) (5 points) Repeat this using your front-door identification strategy.
   (f) (5 points) Do your two estimates of the casual effect match? Explain.
2. (25 points)Take the model in Figure 1. Suppose that $X \sim \mathcal{N}(0,1)$, $Y = \alpha X + \epsilon$ and $Z = \beta_1 X + \beta_2 Y + \eta$, where ϵ and η are mean-zero Gaussian noise with common variance σ^2. Set this up in R and regress Y twice, once on X alone and once on X and Z. Can you find any values of the parameters where the coefficient of X in the second regression is even approximately equal to α? (It's possible to solve this problem exactly through linear algebra instead.)
3. (25 points) Take the model in Figure 2 and parameterize it as follows: $U \sim \mathcal{N}(0,1)$, $X = \alpha_1 U + \epsilon$, $Z = \beta X + \eta$, $Y = \gamma Z + \alpha_2 U + \xi$, where ϵ, η, ξ are independent Gaussian noises with mean zero and common variance σ^2. If you regress Y on Z, what coefficient do you get, on average? If you regress Y on Z and X? If you do a back-door adjustment for X? (Approach this either analytically or through simulation, as you like.)
4. (20 points) Continuing in the set-up of the previous problem, what coefficient do you get for X when you regress Y on Z and X? Now compare this to the front-door adjustment for the effect of X on Y.

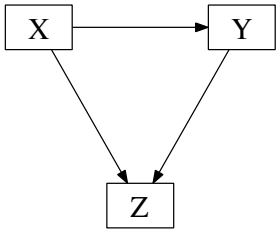

Figure 1: DAG for problem 2.

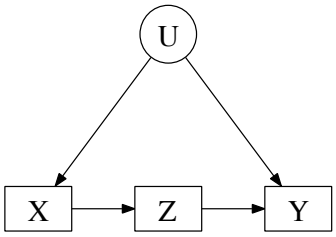

Figure 2: DAG for problems 3 and 4.