CAUSAL INFERENCE WITH GRAPHICAL MODELS

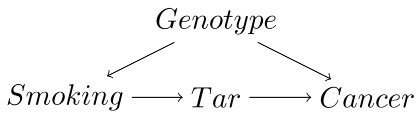
LAB

JOSE M. PEÑA
STIMA, IDA, LIU

To be done in pairs.
Submission: By e-mail to `jose.m.pena@liu.se`.
Deadline: See the course website.

---

(1) Consider the following causal model:



where $Genotype$ is unobserved. Add to the model the variables $Education$, Age and $ParentalEducation$ where the last two are unobserved. Add to the model the causal relations that you consider appropriate between the variables in each of the following pairs: (E, S), (E, C), (A, E), (A, S), (PE, E) and (PE, C). Project the resulting model onto the observed variables. Use the `R` package `causaleffect` to determine if the causal effect $p(C|do(s))$ is identifiable. If so, report the corresponding expression.

(2) Repeat the exercise 1 without the variable PE and its associated relations. Check if you obtain the same result and explain why or why not.

(3) Assume that we know that a natural phenomenon follows either the causal model

$$x = u_X$$
$$y = \alpha x + u_Y$$

with $U_X \perp U_Y | \varnothing$, or the causal model

$$x = \beta y + w_X$$
$$y = w_Y$$

with $W_X \perp W_Y | \varnothing$. Moreover, assume that U_X, U_Y, W_X and W_Y are normally distributed with unknown mean and variance each. Can observational data distinguish the two models ? If not, can interventional data do so ? If so, how many interventions (i.e., $do(x)$ or $do(y)$) are needed ?

(4) Repeat the exercise 3 assuming that $\alpha, \beta \geq 1$. Hint: Check the variances of X and Y.

(5) Repeat the exercise 3 assuming that U_X and U_Y have equal variance, and W_X and W_Y have also equal variance. Hint: Check the variances of X and Y.

(6) There is a treatment T for the disease D. However, the effectiveness of T depends on whether the patient has a rare condition C or not. Unfortunately, there is no way to know whether a patient has the condition or not. Specifically, the survival of a patient

(in years) is as follows:

$$S|T,C \sim \begin{cases} \mathcal{N}(4,1) \text{ if } T=0 \text{ and } C=0 \\ \mathcal{N}(8,1) \text{ if } T=0 \text{ and } C=1 \\ \mathcal{N}(7,1) \text{ if } T=1 \text{ and } C=0 \\ \mathcal{N}(5,1) \text{ if } T=1 \text{ and } C=1 \end{cases}$$

and $p(C=0)=0.9$.

Mr. P suffered the disease, and his doctor decided to administer him the treatment. Mr. P died after 4.5 years. Would Mr. P have lived longer (in expectation) had the doctor not administered him the treatment ? Did the doctor act optimally (in expectation) according to the knowledge available when she took the decision ?