Reading & Reading Exercises Meeting 2

- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. Advances in Methods and Practices in Psychological Science, 1(1), 27-42. https://doi.org/10.1177/2515245917745629
- Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. Frontiers in Genetics, 10: 524. https://doi.org/10.3389/fgene.2019.00524 Focus on sections 1 up to and including 3.1, and the recommendation sections (6 onward) toward the end of the paper. The rest you may skip.

Rohrer (2018) is written by and for psychology researcher(s), and evaluates the common practice of ``statistically controlling'' for variables from a DAG perspective. It includes an introduction to/overview of DAGs for a non-technical audience.

The paper by Glymour et al. is a review of different causal discovery/learning methods, of which we considered only one in our lecture: conditional-independence based methods (called constraint-based and score-based methods in the paper). Sections 3.2, 3.3 and 4 you may skip, or read out of interest: The algorithms described in sections 3.2 and 3.3 of the paper are not covered this week but are interesting alternative ways to pursue conditional-independence based causal discovery; In section 4 they cover Functional Causal Models (also called restricted SCMs), which is an alternative causal discovery technique not part of this course (another relatively new discovery method you could google if you are interested is 'Invariant Causal Prediction')! This paper at times goes into more technical detail than we did in the lecture. It is not necessary to understand all the technical details, the goals is you understand the general ideas.

Reading Questions

1. What are examples of different ways in which a researcher might hope to control for (i.e. condition on) a third variable?

2. If you are interested in estimating a causal effect between two variables, is it right to say you **should** control for all of the other variables you have measured? Why, or why not?

3. How do we define a *confounding* variable in the framework of graphical causal models? How is this different from a *collider* or a *mediator*?

4. Think of an example in your own field where you might want to estimate the causal effect between two variables. Can you think of a relevant a) confounder, b) collider, and c) mediator?

5. What is the difference between causal identification and causal discovery?

6. In the lecture we discussed how DAG structures imply certain marginal and conditional (in)dependence relations between variables. How can we use this information to learn about the DAG structure? Think about the chain, fork and collider DAGs. What can and can't we learn from the patterns of conditional and marginal dependencies about a causal DAG structure?

7. Why do we need the faithfulness assumption for conditional independence-based methods to work? What does it mean in practice?

Extra reading not required for the course, but for those interested:

This is non-obligatory reading, but potentially useful for the exercises: http://www.dagitty.net/learn/dsep/index.html. It is a guide to d-seperation rules, adapted from Judea Pearl's book.

You might also find it interesting to read some of the work of Judea Pearl, the main developer of DAG and SCM techniques for causal inference. Two of his more accessible works include a popular science book called *The book of why* and a primer for graduate students.

Pearl, J., & Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic books.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.