STAT 591 A Sp 22: Special Topics In Statistics

Causal Inference: Identifiability and Estimation
Instructor: Ema Perković;

Office Hours
Tuesday 5:00 - 6:30 PM, If you have time conflicts with this office hour time in general, please let me know!

Course Description:
Most statistics and data science models are concerned with the task of prediction: finding the best predictors, finding the best models for prediction, etc. However, in many scientific disciplines, what we are truly interested in is predicting a system's behavior under manipulation. For such analyses, prediction is not enough. Instead, we require knowledge about the underlying causal structure of the system. In this course, we study concepts and theories behind causal inference. We will rely on graphical models to build causal intuition and cover state-of-the-art techniques and limitations for structure learning and causal effect estimation.

This course assumes that students have a decent grasp of linear algebra, the theory of linear models, probability theory, and
programming in R. Having taken a class on graphical models, or STAT 566 (Causal Modeling), is beneficial but not required. In the first part of the course, the lectures will consist of presentations by the instructor. In the second part of the course, lectures will consist of discussions of one or two papers, and students will be expected to participate in all discussions and be the lead presenters for at least two papers.

**Grading policy:**

This is a pass/fail (CR/NC) course.

- **Assigned readings/homework (30%)** - there are 3 HW in the first half of the course + the assigned readings for your presentation. The homework and presentation materials should be submitted through Gradescope.
- **Presentation (35%)** - In the second half of the quarter, each student will be a part of a pair giving a 60-minute presentation based on the results of 2+ papers.
- **Leading the discussion and class participation (35%)** - Each group will be assigned to lead the discussion on 2 other presentations. Some class discussion of the remaining presentations is also expected.

**Outline of course topics (see also the Lecture Notes page):**

**Week 1-5, Instructor presentations on the following topics:**

1. **Introduction:** Correlation does not imply causation, Simpson's paradox, Ladder of causation, Different causal "worlds"/models.
2. **Graphical models:** Directed acyclic graphs (DAGs), d-separation, faithfulness, structural linear causal models.
3. **DAGs as causal models:** do-interventions, do-calculus, total, and direct causal effects, g-formula
4. **Total effect identification/estimation with DAGs:** Covariate adjustment, IPW.
5. **Equivalence classes of DAGs and structure learning**: non-identifiability of the DAG, graph learning algorithms: constraint-based, score-based, and hybrid methods, background knowledge

6. **Causal Effects in CPDAGs and MPDAGs**: node relationships, causal effect identification issues, adjustment criterion, IDA

7. **Hidden variable models and causal estimation**: instrumental variables, Mendelian randomization, adjustment.

8. **Hidden variable models**: ambiguity and structure learning (time permitting).

For most of the topics, I will provide some R-markdown files to better understand the materials.

**Week 6-10:**

Student-run presentations and discussions of some of the following topics:

- Structure Learning,
- Identification and Enumeration of Causal Effects under Causal Sufficiency
- Efficient Estimation under Linearity and Causal Sufficiency
- Combining Observational and Experimental Data to Improve Estimates
- Experimental Design
- Types of interventions: dynamic treatment regimes, stochastic interventions
- Causal Feature Learning
- Fairness

**Textbooks/Lecture notes (green - available online)**

Our textbook officially is:

- *Elements of Causal Inference*
  by Jonas Peters, Dominik Janzing, and Bernhard Schölkopf
But I do not plan to exactly follow this book. Other textbooks that you may find useful depending on the topic and style preference:

- **Causal Inference in Statistics: A Primer (parts 1, 2, 3, and 4),** by Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell

- *Handbook of Graphical Models*
  by Marloes H. Maathuis, Mathias Drton, Steffen Lauritzen and Martin Wainwright

- *Advanced Data Analysis from an Elementary Point of View*
  by Cosma Shalizi, Chapters 20-25.

- **Brady Neal's course: Introduction to Causal Inference**

- **Causal Inference: What if?**
  by Miguel A. Hernán, James M. Robins

- **Causation, Prediction, and Search**
  by Peter Spirtes, Clark Glymour, and Richard Scheines, second Edition

- **Causality**
  by Judea Pearl, the second edition

- **Observation and Experiment: An Introduction to Causal Inference**
  by Paul R. Rosenbaum

- **Mostly Harmless Econometrics: An Empiricist's Companion**
  by Joshua D. Angrist and Jörn-Steffen Pischke

- **Graphical Models**
  by Steffen Lauritzen, [lecture notes are available.](#)

- **Causal Inference for Statistics, Social, and Biomedical Sciences**
  by Guido W. Imbens, Donald B. Rubin.

**R packages:**

I will provide some R markdown files throughout the course.
Most papers we will read have some implementation in R. I recommend using R in combination with Rstudio: [https://www.rstudio.com/products/rstudio/download/](https://www.rstudio.com/products/rstudio/download/)

Some prominent R packages are given below. A variety of these come with **vignettes** describing some of the functionality of the package.

- **pcalg**
- **dagitty** - the **primer** for dagitty, Johannes also gave a tutorial at OCIS!
- **ggdag**
- **causaleffect**
- **ggm**
- **bnlearn** - webpage.
- **dagR** - presentation paper.

- and many others, feel free to email me if you think others should be added.

**Other Software and Datasets:**

- **DAGitty** a browser-based environment for creating, editing, and analyzing causal models
- **The TETRAD project** - a desktop Java application implementing various causal inference advancements since 1998.
- **Other causal tools by CCD.**
- **Example Causal Data Sets**, collated by J. D. Ramsey.

**Other useful resources:**

- **Online Causal Inference Seminar's library of presentations**, live streams every Tuesday at 08:30 AM PT
- SimonsTV recordings of workshops on causal inference:
  - Causality bootcamp
- Learning from interventions
- Algorithmic Aspects of Causal Inference
- Quantum Physics and Statistical Causal Models

**Students with disabilities:**

If you would like to request academic accommodations due to a disability, please contact Disabled Student Services, 448 Schmitz (206) 543-8924 (V/TTY). If you have a letter from Disabled Student Services indicating you have a disability that requires academic accommodations, please present this letter to me to discuss the accommodations you might need for the class.

**Academic Integrity:**

Collaboration and discussions are allowed and encouraged in this class, but copying or letting others copy your work amounts to plagiarism. This includes copying model solutions, e.g., from prior years. Although cheating seldom occurs in graduate classes, if it does, I will take the following action: assign a grade of 0.0 for the exam/homework where the cheating occurred, and report the incident to the Graduate School Committee on Academic Conduct, which will decide upon an appropriate University course of action.

**Religious Accommodations:**

Washington state law requires that UW develop a policy for the accommodation of student absences or significant hardship due to reasons of faith or conscience or for organized religious activities. The UW’s policy, including more information about requesting accommodation, is available at Religious Accommodations Policy (https://registrar.washington.edu/staffandfaculty/religious-accommodations-policy/). Accommodations must be requested within the first two weeks
of this course using the Religious Accommodations Request form (https://registrar.washington.edu/students/religious-accommodations-request/).

**Student conduct:**

Follow the UW Student Conduct Code in your interactions with your colleagues and me in this course by respecting the many social and cultural differences among us, which may include, but are not limited to: age, cultural background, disability, ethnicity, family status, gender identity and presentation, citizenship and immigration status, national origin, race, religious and political beliefs, sex, sexual orientation, socioeconomic status, and veteran status.