Williams College

Causal Inference (CSCI 379)

Instructor: Rohit Bhattacharya (rb17@williams.edu)
TAs: Andrew Megalaa (anm4@williams.edu) and Jack Sullivan (jcs6@williams.edu)

Course Meeting Times And Links

Meeting Times: Monday & Thursday, 1:10pm – 2:25pm
Meeting Location: Schow 30A
Piazza: https://piazza.com/williams/spring2024/csci379 (copy & paste into your browser)
Gradescope entry code: KKW8ZD

Course Description

Seeking answers to causal (as opposed to associational) questions is a fundamental human endeavor; the answers we find can be used to support decision-making in various settings such as healthcare and public policy. This course covers core topics in causal inference including causal graphical models, unsupervised learning of the structure of these models, expression of causal quantities as functions of observed data, and robust/efficient estimation of these quantities using statistical and machine learning methods. We will study these concepts and their usage via bi-weekly case studies demonstrating the use of causal reasoning for various inferential tasks, including detecting and quantifying bias in scientific studies. The final project will be "hands on," where students will apply techniques learned in class to derive data-driven insights into a scientific question of their choosing, and write up their results.

Course Objectives

At the end of this course you should be able to

- Identify and reason about sources of bias that may lead to invalid causal inference
- Understand the semantics and limitations of various approaches to causal modeling
- Analyze real data using causal graphical models

Textbooks

There are no required textbooks for this course. Here are some textbooks for supplementing your understanding if you wish to read beyond the course materials.

- *Causal Inference: What If* by Miguel Hernán and James M. Robins (2010) – a free pdf of this book is available for download here

- *Causality* by Judea Pearl (2009)
- *Graphical Models* by Steffen L. Lauritzen (1996)

Grading Overview

This course focuses more on learning rather than assessment. In the long run, the knowledge and skills you acquire are far more important than your final grade. The grading policy and guidelines are made to reflect this course philosophy.

- Homework: HW 1: 9% and HW 2-4: 36% ($3 \times 12\%$)
- Final project: 40%
- Project proposal: 5%
- Peer review: 5% (details to be announced in class)
- Discussion session: 5% (in late April/early May, details to be announced in class)

Course Schedule

This is a rough schedule and subject to change.

- Week 1: Jan 31
  - Course overview and introduction
- Week 2: Feb 5, 8
  - Introduction to probability, association, and estimation of probabilities
  - Moving from a calculus of probability to a calculus of causation
  - **HW1 is assigned Feb 8**
- Week 3: Feb 12 and Feb 15
  - Formalizing statistical and causal models
  - Identification in randomized controlled trials
  - Identification in the presence of simple confounding
  - **HW1 is due Feb 15, 11:59pm EST**
  - **HW2 is assigned Feb 15**
- Week 4: Feb 19, Feb 22
  - Statistical and causal models of directed acyclic graphs (DAGs)
  - Causal inference with DAGs

- Week 5: Feb 26, Feb 29
  - Backdoor adjustment
  - Doubly robust and efficient estimation
  - **HW2 is due Feb 29, 11:59pm EST**
  - **HW3 is assigned Feb 29**
- Week 6: Mar 4, Mar 7
  - Introduction to unmeasured confounding and non-identifiability
  - Unmeasured confounding and acyclic directed mixed graphs (ADMGs)
- Week 7: Mar 11, 14
  - Front-door adjustment
  - Instrumental variable adjustment
  - **HW3 is due Mar 14, 11:59pm EST**
- Spring break: Mar 16 - Mar 31
- Week 8: Apr 1, 4
  - Testing assumptions in causal models
  - Tests for backdoor and front-door models using an auxiliary variable
  - Introduction to causal discovery
  - **Project proposals due Apr 4, 11:59pm EST**
  - **HW4 is assigned Apr 4**
- Week 9: Apr 8, 11
  - The PC algorithm for causal discovery
  - Extensions of PC to unmeasured confounding
- Week 10: Apr 15, 18
  - Score-based learning
  - Project expectations, scientific writing, and peer review
  - **HW4 is due Apr 18, 11:59pm EST**
- Week 11: Apr 22, 25
  - Proximal causal inference
  - Combining different causal models to gain robustness

- Week 12: Apr 29, May 2
  - Machine learning in causal inference
  - Regularization and model selection in causal inference
- Week 13: May 6, May 9
  - Missing data and causal inference
  - Algorithmic fairness and causal inference
  - Course wrap up
- **Projects due for peer review May 6, 11:59pm EST**
- **Peer review due May 9, 11:59pm EST**
- Finals week: **Final project is due May 18 at 5:00pm EST**

Homework And Final Project Schedule

There will be 4 homeworks in total, the first of which is slightly shorter and meant to familiarize you with the course and its contents. You will be given roughly two weeks to complete each subsequent homework assignment. Homeworks are due on the date assigned in the above schedule. Homework assignments must be completed individually, not in groups. For the final project, I am considering allowing students to work individually or in pairs. More details and expectations for the final project will be released as the course progresses.

Late Homework Policy

Homeworks submitted 0-24 hours late are eligible for only 90% of the maximum credit, 24-48 hours late for 80%, 48-72 hours late for 70%, and later than 72 hours for 0%. However, you have **4 late days** to be used at your discretion to turn in late assignments for full credit. Late days can be used to postpone any assignment deadline **except** those related to submitting the final project. This exception is made to allow enough time for others to peer review your final project, and prevent any delay in entering the final grades on (for which there exists a hard deadline from the college.)

If there are extenuating circumstances, such as a family or medical emergency, that prevent you from completing an assignment on time (even after having used all four late days), please let me know as soon as possible – I will do my best to accommodate your request.

Programming Language

Homework assignments will require some programming, and the final project will require real data analysis which may either require novel programming or the use of available software. All homework assignments must be completed in Python. You may use whatever programming languages and software you like for the final project.

Health And Accessibility Resources

Any student with a disability who may need accommodations in this class is encouraged to contact Dr. G. L. Wallace (Director of Accessible Education) at (413) 597-4672. Also, students experiencing mental or physical health challenges that are significantly affecting their academic work or well-being are encouraged to contact me and/or speak with a dean so we can help you find the right resources. The deans can be reached at (413) 597-4171.

Please let me know if you are unable to attend class due to sickness/other reasons. I will work with you to develop a plan that allows you to continue making progress in the course during your time away from the classroom.

On Group/Individual Work And Cheating/Plagiarism

All assignments in this course are individual, not group, assignments. You may freely discuss homework assignments with your classmates. The final solutions, however, must be written entirely on your own. This includes any programming tasks. Copying someone else's code or proofs (and subsequently making minor changes) constitutes plagiarism. A similar policy applies to the use of generative AI, except in certain special cases where I explicitly permit the usage of it. If you need to discuss programming or analytic portions of an assignment, you may discuss general strategy but should write the code/proofs by yourself. When working on your final projects, make sure to cite your sources when appropriate – we will have a class devoted to covering aspects of scientific writing, including appropriate citation etiquette, as we move towards the final project phase of the course. If in doubt, please feel free to contact me to ask whether certain modes of collaboration are allowed.