Causal Inference for Data Science

Instructor: Adam Kelleher

Syllabus and Tentative Schedule

General Information

- **Meeting time:** Wednesday, 7:00p-9:30p
- **Contact**: Adam Kelleher, ak4063@columbia.edu
- **Office Hours**: TBA
- **Textbook**: Course Notes (roughly following Morgan and Winship)
- **Grading**: Homework 40%, Mid-term 30%, Final exam 30%

Pre-requisites

- **Math:** Undergrad probability theory; Some experience with regression analysis will be useful; some knowledge of information theory will be useful, but not required. Some knowledge of bayesian networks will be useful but not required.
- **CS:** Basic knowledge of Python and R

Homework

- Homework will be due by midnight two weeks after they are assigned. There will be assignments approximately every week.
- Late assignments will be reduced as follows:
    - 0+ - 24 hours late: 25% of points deducted
    - 24 - 48 hours late: 50% of points deducted
    - More than 48 hours: no credit
- Exceptions will be made for medical emergencies or other exceptional circumstances **discussed in advance**.

Collaboration

- Collaboration is strongly encouraged. Everyone must write up their assignments on their own. Copying collaborators' work or copying work from other sources (textbooks, the internet) is prohibited.

Programming Assignments

- Programming assignments will be completed with Jupyter notebooks and R. Install and familiarize yourselves with Jupyter notebooks and R as soon as possible.

Tentative Schedule:

<u>Part 1: Experimental Causal Inference</u>

- Chapter 1: AB Testing
    - The fundamental problem of causal inference: definition of potential outcomes; definition of causal states; observing the outcome for a single unit in two states is impossible
    - Assumptions: treatment independent of potential outcomes; how it leads to identification
    - Experiment Design examples: ad effectiveness; website change testing; recsys evaluation
    - Statistical considerations in choosing outcomes: t-tests and the effect of variance on sample size; conditional effects take large samples
    - Hypothesis testing: confidence intervals and p-values; independent samples assumption; examples based on previous experiment design examples
    - dependent samples error bars: failure of indep samples means overconfidence; eckles's facebook experiment example; AA testing
    - Intent-to-treat analysis: dealing with non-compliance or instruments for treatment.
- Chapter 2: Infrastructure and experiment implementation:
    - Implementation details: Treatment assignment with hashing trick; data pipeline from frontend to database; description of required metadata; frontend logic to render the experiment
    - data structures and queries: example query for average difference in outcomes; example stratification
    - front-end implementation pitfalls: be sure to remove experiment code; watch for experiment collisions (rare nonlinearities)

    - test run-time and monitoring: check for bugs during early run; estimating run time using t-test power calculations; caution about multiple testing during test run.

<u>Unit 2: Observational Causal Inference</u>

- Chapter 3: Context: data science in business and observational data to guide policy decisions;
    - Sources of bias in causal inference (PO decomposition)
    - Introduction and examples: ad targeting; predicting the future; product structuring changes
    - A hierarchy of correlation and causation (correlative, granger causal, causal, counterfactual) and their business cases; cost often grows as we go up the hierarchy; can you afford not to pay it?
- Chapter 4: Causal Frameworks:
    - Intro to Counterfactuals and Potential Outcomes
    - endogenous treatment assignment
    - Intro to identifiability; conditional ignorability; the back-door criterion in brief
    - ATT, ATC, ATE and other CATEs; relaxed assumptions still allow identifiability; contrast pearlian and PO frameworks: weaker assumptions in PO
- Chapter 5: Intuition for the Pearlian Framework
    - causal graphs; structure implies statistics; many graphs are observationally equivalent
    - bayesian networks and factorization of the joint; statistical dependence in confounding, collider, and chain graphs. intuition for confounding, colliders, and chains; d-separation;
    - generating toy data with causal structure
- Chapter 6: Identifiability and the Pearlian Framework
    - the do(x) operation; back-door criterion and relation to conditional ignorability
    - The do operation in simulated data
    - More intuition for answering causal questions: Confounding; Berkson's Paradox; simpson's paradox; robin’s g-formula; reinforcement learning example

<u>Unit 3: Effect Estimation with Conditioning</u>

- Chapter 7: From Probabilities to Data: Intro to conditioning
    - stratification as conditioning
    - exact matching; distance matching; propensity score matching
    - inverse propensity weighting
- Chapter 8: Model-Based Conditioning:
    - OLS regression estimators of causal effects and understanding regression specifications
    - understanding the data support; model-based extrapolation outside the support
    - model misspecification error
    - variable subset selection in ML
    - Doubly robust estimators
- Chapter 9: Machine-learning Based Conditioning:
    - Beyond OLS regression estimators of causal effects
    - loss functions for building conditional effect estimators; example machine learning effect estimators
    - Caution: Induced dependence in derived features
- Chapter 10: Limitations of Conditioning:
    - the data processing inequality, information loss, and the limitations of feature engineering;
    - Unmeasured confounders
    - Bias Amplification
    - Conditioning to reduce entropy vs conditioning to remove bias: the causal statistical distinction revisited

<u>Unit 4: Effect Estimation without Conditioning</u>

- Chapter 11: Instrumental Variables*:*

    - Instrumental variables graphs; understanding the exclusion restriction; conditioning for dirty instruments
    - Ad effect example; recommender system example
    - Estimators and their properties: The wald estimator; two-stage regression estimators; machine learning estimators; binary interval estimators.
- Chapter 12: Mechanistic Inference:
    - Mechanistic graphs
    - mechanisms of action and the Front-door criterion; mechanistic estimators with ML models
    - Use cases and recent examples to pick apart routes for effects
- Chapter 13: Discontinuity Designs:
    - Regression Discontinuity Design
    - interrupted time series
    - difference in differences
    - More on choosing regression specifications
- Chapter 14: Panel Data Designs:
    - unit-level interrupted time series
    - ITS and conditioning with many units across time

Unit 5: Selection Bias Adjustment

- Chapter 15: Context:
    - data sets often aren’t generated in the context of interest.
    - Causal graphs with sampling indicators.
    - Examples: vendor data; AB tests for growth; partially observed data
- Chapter 16: Adjustment Methods:
    - post-stratification weighting (survey example)
    - causal and statistical adjustment
    - regression-based estimation
    - ML model adjustment

Unit 6: The Future

- Chapter 17: Research and Further Reading
    - Structure-inspired latent variable models
    - Machine learning estimators for potential outcomes
-

Homework Schedule
(To be Updated)