Causal Inference
CS 477-677

Introduction

Ilya Shpitser

JOHNS HOPKINS UNIVERSITY

Outline

Course Goals

- Scientific literacy: learning to be informed citizens, consumers, and patients.
- Understand bias sources in data – very important for data science.
- Pose causal questions:
  - Science (is this drug good?)
  - Optimal decision making.
- Construct causal models: experts, background knowledge or learn from data.
- Implement algorithms that use causal models and data to answer causal questions.
- Learn to deal with missing data.
- Think about validation. (Hard problem!)

Prerequisites

- Basic probability theory.
- Basic multivariate calculus (first, second derivatives, etc.)
- Basic linear algebra (determinant, matrix inverse, etc.)
- Ability to program (will use the R programming language).
- Ability to prove things (mostly usings graphs and probabilities, no "abstract nonsense").

Course Structure

- Five homeworks (8% each), midterm (20%), final exam (40%).
- Homeworks will have theory and programming portions (in R).
- Taught in Fall 2016. Student feedback:

Course Structure

- Five homeworks (8% each), midterm (20%), final exam (40%).
- Homeworks will have theory and programming portions (in R).
- Taught in Fall 2016. Student feedback:
  - I personally feel this is one of the best courses that I have taken in 4 years at JHU.

Course Structure

- Five homeworks (8% each), midterm (20%), final exam (40%).
- Homeworks will have theory and programming portions (in R).
- Taught in Fall 2016. Student feedback:
  - I personally feel this is one of the best courses that I have taken in 4 years at JHU.
  - There is very little required background for this course, but it gives you a new way of thinking about things. It is hard, but all the required material is contained in the class.

Course Structure

- Five homeworks (8% each), midterm (20%), final exam (40%).
- Homeworks will have theory and programming portions (in R).
- Taught in Fall 2016. Student feedback:
  - I personally feel this is one of the best courses that I have taken in 4 years at JHU.
  - There is very little required background for this course, but it gives you a new way of thinking about things. It is hard, but all the required material is contained in the class.
  - Be prepared that the HW and exam questions are challenging and there is no good reference material to study since this is a relatively new field. You better know probabilistic graphical model well.

Course Structure

- Five homeworks (8% each), midterm (20%), final exam (40%).
- Homeworks will have theory and programming portions (in R).
- Taught in Fall 2016. Student feedback:
  - I personally feel this is one of the best courses that I have taken in 4 years at JHU.
  - There is very little required background for this course, but it gives you a new way of thinking about things. It is hard, but all the required material is contained in the class.
  - Be prepared that the HW and exam questions are challenging and there is no good reference material to study since this is a relatively new field. You better know probabilistic graphical model well.
  - As it was taught this year, it was not very applicable to an undergraduate who wants a second course in machine learning.

Course Structure

- Five homeworks (8% each), midterm (20%), final exam (40%).
- Homeworks will have theory and programming portions (in R).
- Taught in Fall 2016. Student feedback:
    - I personally feel this is one of the best courses that I have taken in 4 years at JHU.
    - There is very little required background for this course, but it gives you a new way of thinking about things. It is hard, but all the required material is contained in the class.
    - Be prepared that the HW and exam questions are challenging and there is no good reference material to study since this is a relatively new field. You better know probabilistic graphical model well.
    - As it was taught this year, it was not very applicable to an undergraduate who wants a second course in machine learning.
    - There are [. . .] some open-ended, more philosophical and/or conceptual aspects to the course that get you thinking about what statistics and machine learning are really about. Causal Inference brings a valuable perspective to studying these topics.

Course Structure

- Announcements, Q&A, course info on piazza:

  `https://piazza.com/jhu/fall2018/cs600477677/home`

- Signup:

  `https://piazza.com/jhu/fall2018/cs600477677`

- Homeworks will be collected via gradescope (access code MK3JPN):

  `https://www.gradescope.com/courses/24555/`

- Generally no late submissions.
- Midterm in class, final take home.

Practical Advice

- Highly suggest brushing up on prerequisite material.
- No single textbook, will assign readings.
- Emphasis on conceptual understanding (most points for that).
- Theory probably harder than programming in this class.
- Start homeworks early.
- Homeworks are hard, but I grade generously.
- First instance of plagiarism: 0 on the assessment item.
- Second instance of plagiarism: failing grade in the class, and a way one trip to the JHU student conduct folks.

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?
- Who has taken a machine learning class?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?
- Who has taken a machine learning class?
- Who has taken a causal inference class?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?
- Who has taken a machine learning class?
- Who has taken a causal inference class?
- Who knows what a regression model is?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?
- Who has taken a machine learning class?
- Who has taken a causal inference class?
- Who knows what a regression model is?
- Who knows R?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?
- Who has taken a machine learning class?
- Who has taken a causal inference class?
- Who knows what a regression model is?
- Who knows R?
- Who knows what a Bayesian network is?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?
- Who has taken a machine learning class?
- Who has taken a causal inference class?
- Who knows what a regression model is?
- Who knows R?
- Who knows what a Bayesian network is?
- Who knows what a Markov property is?

Poll

- Who is from computer science? Pure math? Applied math/stats? Public Health? Economics? Social science? Philosophy?
- Who has taken a statistics class?
- Who has taken a machine learning class?
- Who has taken a causal inference class?
- Who knows what a regression model is?
- Who knows R?
- Who knows what a Bayesian network is?
- Who knows what a Markov property is?
- Who knows what a potential outcome is?

Association vs Causation

- Most scientific inquiry/data analyses have one of two goals:
  - Association/prediction, i.e., determine predictors or variables associated with the outcome of interest.
  - Causality, i.e., understand factors that cause or have an effect on the outcome of interest.
- We are often told that association is not causation.
- Spurious correlations are very common!

Association vs Causation

- Most scientific inquiry/data analyses have one of two goals:
  - Association/prediction, i.e., determine predictors or variables associated with the outcome of interest.
  - Causality, i.e., understand factors that cause or have an effect on the outcome of interest.
- We are often told that association is not causation.
- Spurious correlations are very common!
- When can we claim an association is causation?
- This is what causal inference is about.

Why Causal Inference?

Example (Pearl et al):
We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown below:

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Total	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Why Causal Inference?

Example (Pearl et al):
We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown below:

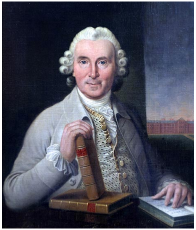
	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Total	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Is the drug helpful?

Two Quotes on Causality from the 1740s

We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second, ... where, if the first object had not been the second never had existed.

David Hume (1748)

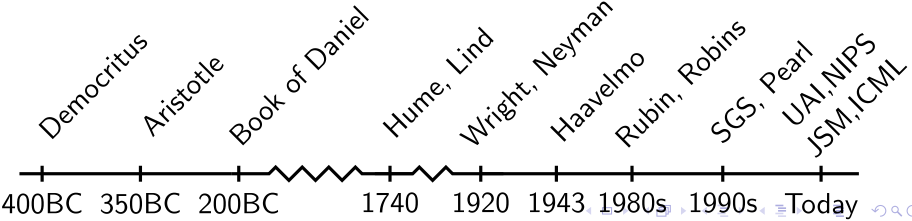


Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of the knees. They lay together in one place, being a proper apartment for the sick in the forehold...

James Lind (1747)

Timeline

- "I would rather learn one causal law than be King of Persia."
- Aristotle's four causes (material, formal, efficient, final).
- Book of Daniel: earliest recorded mention of a comparison trial.
- Hume's definition, Lindt's scurvy trial.
- Wright's path analysis, Neyman's potential outcomes.
- Haavelmo's structural equations.
- Modern methods: Rubin,Robins,Pearl,Spirtes,Glymour,Schenes.
- Today: tutorials at UAI, NIPS, ICML. Conference/journal papers. Causal inference a part of FDA best practices guidelines.



View Of Causality For This Course

- Philosophers, scientists thought about causality for a long time.
- Lots to say, lots of approaches (Hume alone had 3 definitions, Aristotle had 4).
- Will take a particular view of causality most relevant for scientific inquiry.
- Will only talk about causality we can implement.
- Example "causal effect": a difference in outcomes between hypothetical experiments we may do.

Multiple Views of Causal Inference

- Grew out of statistical analysis of experimental data to cases where data is not experimental.
- Grew out of doing linear regressions linking causes to effects.
- "How to do statistics/machine learning when your data is biased?"
- Mathematical formalization of David Hume's definition of causality.
- "In classical associational analysis, association is not causation. Why not? When is association the same as causation?"

Causal Inference vs Machine Learning/Statistics

Machine Learning:

- Supervised learning:
  - Learn relationship of "outcome" Y to "features" $\vec{X}$.
  - Classification: image has a horse? a car? a person?
  - Regression: how favorable is this chess position?
- Unsupervised learning:
  - Group emails by topic
  - Learn a model from "scratch" using data
- Reinforcement learning:
  - Agents acts in world, and senses environment,
  - Received rewards periodically, has to learn "optimal" mapping from current state to action.
  - Partly supervised, partly active learning
- Take CS475 to learn a lot more about Machine Learning!
- In ML some methods are probabilistic, some are not.

Causal Inference vs Machine Learning/Statistics

Causal Inference:

- (Partly) supervised learning:
  - Assess causal effects: is this drug effective?
  - Assess mediation: is smoking bad for you because of smoke or because of nicotine?
  - Missing data: estimate the mean of the whole population from a biased sample.
- Unsupervised learning:
  - Learn "does A cause B?" from data.
  - Connections to learning graphical models from data.
- Decision making:
  - Agent learns to maximize causal effect of actions from data.
  - Connections to reinforcement learning.
- Will always use probabilistic models in this class.
- Models defined on factual variables (what we see), and counterfactual variables (what we don't see, but are interested in).

Course Outline

What is a causal model? (Will give increasingly general answers to this).

- Probability, and linear algebra review.
- Simple causal effects, and simple missing data models.
- Simple mediation analysis.
- General graphical causal models.
- Causal inference with hidden variables.
- Midterm
- Causal decision theory.
- Learning causal models from data.
- Advanced topics if there is time: sensitivity analysis, interference, connections to machine learning.
- Final Exam.

Statistical vs Causal Models vs Missing Data Models

On the board

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.
- Define a causal model (expert, background knowledge, learn from data).

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.
- Define a causal model (expert, background knowledge, learn from data). Details later, for now say "gender causes drug, gender and drug both cause outcome."
- Formalize question as a target estimand in the causal model.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? ⇒ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.
- Define a causal model (expert, background knowledge, learn from data). Details later, for now say "gender causes drug, gender and drug both cause outcome."
- Formalize question as a target estimand in the causal model. Details later, average vs gender specific causal effect .

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.
- Define a causal model (expert, background knowledge, learn from data). Details later, for now say "gender causes drug, gender and drug both cause outcome."
- Formalize question as a target estimand in the causal model. Details later, average vs gender specific causal effect .
- Identify target as function of observed data. Identified in our case.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.
- Define a causal model (expert, background knowledge, learn from data). Details later, for now say "gender causes drug, gender and drug both cause outcome."
- Formalize question as a target estimand in the causal model. Details later, average vs gender specific causal effect .
- Identify target as function of observed data. Identified in our case.
- Choose an estimator using statistical models for parts of observed data.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.
- Define a causal model (expert, background knowledge, learn from data). Details later, for now say "gender causes drug, gender and drug both cause outcome."
- Formalize question as a target estimand in the causal model. Details later, average vs gender specific causal effect .
- Identify target as function of observed data. Identified in our case.
- Choose an estimator using statistical models for parts of observed data.
- Fit estimator parameters using data.

Causal Inference Workflow

- Formulate the question of interest as a hypothetical experiment we might want to do. Is the drug helpful? $\Rightarrow$ compare outcome for person i if we gave drug vs not.
- Define relevant variables. Gender, drug, outcome.
- Define a causal model (expert, background knowledge, learn from data). Details later, for now say "gender causes drug, gender and drug both cause outcome."
- Formalize question as a target estimand in the causal model. Details later, average vs gender specific causal effect .
- Identify target as function of observed data. Identified in our case.
- Choose an estimator using statistical models for parts of observed data.
- Fit estimator parameters using data.
- Assess assumptions (sensitivity analysis).

Next time: Probability, Statistical Models And Learning From Data