

Causal modelling in Medical Research

Debashis Ghosh

Department of Biostatistics and Informatics, Colorado School of Public Health

WNAR Seminar Series

Goals for today

- Introduction to Potential outcomes framework
- Exposure to the causal inference analytic pipeline

- Asks the question of how a person-specific outcome changes under different scenarios
- To understand this pop culturally,

The Red Pill/Blue Pill in 'The Matrix'

and

Helen in 'Sliding Doors'

Causal Inference

- The fundamental notion: counterfactual
- Also called potential outcome

Motivating example (cont'd.)

- Question asked by investigators: “ Is exposure to a statin in the early perioperative period associated with reduced postoperative complications after noncardiac surgery?”
- The causal effect studied in the paper: causal relative risk in 30-day mortality associated with use of statin on the day of or the day after surgery
- **Question:** Ideal way to estimate this effect?

Motivating example (cont'd.)

- Here, the investigators used data from an observational database on 180,478 patients undergoing noncardiac surgery who were admitted within 7 days of surgery and sampled by the Veterans Affairs Surgical Quality Improvement Program (VASQIP)
- Investigators used **propensity scores** for modelling the treatment effect and matching to estimate causal effects

A recipe for causal inference

- 1 Define the exposure
- 2 Define the causal parameter of interest
- 3 Model the assignment mechanism using propensity scores
- 4 Estimate the causal effect conditioning on the propensity score in some way
- 5 Get a variance estimate
- 6 Perform sensitivity analyses for unobserved confounding

A recipe for causal inference

- 1 Define the exposure
- 2 **Define the causal parameter of interest**
- 3 Model the assignment mechanism using propensity scores
- 4 Estimate the causal effect conditioning on the propensity score in some way
- 5 Get a variance estimate
- 6 Perform sensitivity analyses for unobserved confounding

Causal effects

- Let Y be the outcome in the movie
- Back to example from *The Matrix*
- Then one causal effect:

$$Y(\text{Neo takes blue pill}) - Y(\text{Neo takes red pill})$$

- From *Sliding Doors*, a causal effect is

$$Y(\text{Helen catches subway}) - Y(\text{Helen doesn't catch subway})$$

- Why are these causal effects?

Causal effect definition from potential outcomes

- Comparison of potential outcomes within the same individuals
- Requires envisioning a Y for a given person under each scenario
- The **observed data** will be Y under only one scenario
- **Fundamental Problem of Causal Inference**

Causal effects (cont'd.)

- Suppose a person can get a treatment ($T = 1$) or control ($T = 0$)
- Then the data for causal inference can be visualized as

$Y(0)$	$Y(1)$	T
?	y_1	1
y_2	?	0
?	y_3	1
?	y_4	1
\vdots	\vdots	\vdots
y_n	?	0

- The question marks define a missing data mechanism
- If we can fill in question marks with y -values, then we calculate causal effects.

Potential Outcomes

- Setup: let T denote a binary treatment that takes values 0 and 1
- - 1 Let $\{Y(0), Y(1)\}$ denote the potential outcomes for an individual
 - 2 If we want to have a probability model, then we need to formulate a **bivariate** distribution
 - 3 This is two-dimensional, in contrast to a one-dimensional distribution

Terminology

- Causality is tied to an *action* (also called manipulation, treatment or intervention)
- Determine a list of possible actions (equivalent to T so that we have a list of two actions)
- The action is applied to a *unit*
- The presumption is that at any point in time a unit could have been subject to either action (potential outcomes)
- We associate each unit-action pair with a potential outcome

“Ideal” data

- The full data for causal inference can be visualized as

Unit #	$Y(0)$	$Y(1)$
1	y_{10}	y_{11}
2	y_{20}	y_{21}
3	y_{30}	y_{31}
4	y_{40}	y_{41}
\vdots	\vdots	\vdots
n	y_{n0}	y_{n1}

Treatments

- Lots of controversy as to how to define this
- Examples: race, gender
- Rubin/Holland: “No intervention without manipulation”
- Think carefully about mechanisms that would lead to manipulation

Causal effects

- By definition, these are comparisons of potential outcomes made within a unit
- Regression models compare subpopulations and thus do not enjoy causal interpretations without further assumptions
- Example

$$E(\text{SBP}) = \beta_0 + \beta_1 I(\text{Age} > 45)$$

- “Correlation, not causation”

One causal effect definition

Unit #	$Y(0)$	$Y(1)$	$Y(1) - Y(0)$
1	y_{10}	y_{11}	$y_{11} - y_{10}$
2	y_{20}	y_{21}	$y_{21} - y_{20}$
3	y_{30}	y_{31}	$y_{31} - y_{30}$
4	y_{40}	y_{41}	$y_{41} - y_{40}$
\vdots	\vdots	\vdots	\vdots
n	y_{n0}	y_{n1}	$y_{n1} - y_{n0}$

Another causal effect definition

Unit #	$Y(0)$	$Y(1)$	$Y(1)/Y(0)$
1	y_{10}	y_{11}	y_{11}/y_{10}
2	y_{20}	y_{21}	y_{21}/y_{20}
3	y_{30}	y_{31}	y_{31}/y_{30}
4	y_{40}	y_{41}	y_{41}/y_{40}
\vdots	\vdots	\vdots	\vdots
n	y_{n0}	y_{n1}	y_{n1}/y_{n0}

Causal effects

- What were shown on the previous page were individual-level causal effects
- A natural thing to do is average the unit-level causal effects across units
- What assumption are we making when we do this???

Subgroup-specific causal effects

- We can also define causal effects conditional on subgroups
- The subgroups could be based on observed covariates or even on potential outcomes
- What assumption are we making when we do this???

Causal inference assumptions

- Strongly ignorable treatment assumption
- Stable unit treatment value assumption
- Treatment Positivity assumption
- Consistency Assumption

Causal inference assumptions

- Strongly ignorable treatment assumption
- Math:

$$\{Y(0), Y(1)\} \perp T | \mathbf{X}$$

where \mathbf{X} denotes a vector of confounders

- In words: there exist a rich set of confounders such that the potential outcomes can be effectively treated as baseline variables.

SITA implication

- Data visualization

$Y(0)$	$Y(1)$	T	X_1	\dots	X_p
?	y_{11}	1	x_{11}	\dots	x_{1p}
y_{20}	?	0	x_{21}	\dots	x_{2p}
?	y_{31}	1	x_{31}	\dots	x_{3p}
?	y_{41}	1	x_{41}	\dots	x_{4p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
y_{n0}	?	0	x_{n1}	\dots	x_{np}

- SITA also means that we can use x 's to fill in question marks

- Data visualization

$Y(0)$	$Y(1)$	T	X_1	\dots	X_p
?	y_{11}	1	x_{11}	\dots	x_{1p}
y_{20}	?	0	x_{21}	\dots	x_{2p}
?	y_{31}	1	x_{31}	\dots	x_{3p}
?	y_{41}	1	x_{41}	\dots	x_{4p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
y_{n0}	?	0	x_{n1}	\dots	x_{np}

- SUTVA states that the potential outcomes between rows are independent.

Treatment Positivity Assumption

- This means that $0 < P(T = 1|\mathbf{X}) < 1$ for all subjects and for all \mathbf{X} values
- This is needed to have well-defined counterfactuals for defining causal effects
- When is this violated???

Consistency Assumption

- We formulate $Y(0)$ and $Y(1)$ for everybody, but we only observe one of these.
- Consistency Assumption means that the potential outcome and observed outcome are the same.

Assumptions

- There are a lot of assumptions needed to identify causal effects from observational data
- These assumptions cannot be tested empirically

- Treatment: T = exposure to statin on day of or day after surgery (1 if received, 0 if not)
- Outcome: Y = death within 30 days
- Causal parameter: $Y(1)/Y(0)$

A recipe for causal inference

- 1 Define the exposure
- 2 Define the causal parameter of interest
- 3 **Model the assignment mechanism using propensity scores**
- 4 Estimate the causal effect conditioning on the propensity score in some way
- 5 Get a variance estimate
- 6 Perform sensitivity analyses for unobserved confounding

Assignment Mechanism

- We have shown how to conceptualize data for causal inference problems
- Let Y_i^{obs} denote the observed response for subject i
- Then

$$Y_i^{obs} = T_i Y_i(1) + (1 - T_i) Y_i(0)$$

- If T were the result of a coin flip, then

$$P(T = 1) = P(T = 0) = 1/2$$

- This is what a randomized clinical trial does!
- This is not true for an observational study

Assignment Mechanism (cont'd.)

- **Definition:** An *assignment mechanism* refers to how each unit came to receive the treatment level actually received
- The **propensity score** is one attempt at modelling the assignment mechanism. It is defined as

$$e(\mathbf{X}) = P(T = 1|\mathbf{X})$$

- Conditional on propensity score, one achieves “covariate balance” on observed covariates

Potential Outcomes (cont'd.)

- Use of propensity score leads to a three-step approach
 - 1 Estimate propensity score
 - 2 Check for balance and adjust accordingly
 - 3 Based on estimated propensity score, estimate the average causal effect

Variable selection for propensity score

- Recommendation is to include all possible treatment confounders and interactions and quadratics by Rubin
- One variable to avoid: **instruments**
- Intuition: want to include as many variables that are strongly predictive of the potential outcomes
- We will need to check for balance

Balance

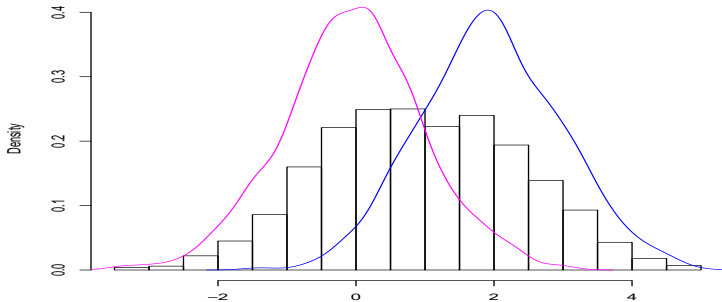


Figure: Distribution of covariate X for treatment and control groups. The blue line denotes the kernel density estimation for X in the $T = 1$ group, while the magenta line represents the kernel density estimate for X in the $T = 0$ group. The bars represent the histogram of X regardless of the treatment groups.

Recipe for causal analysis (Stuart, 2010, Stat Sci)

- 1 Decide on covariates for which balance must be achieved
- 2 Estimate the distance measure (e.g., propensity score)
- 3 Condition on the distance measure (e.g., using matching, weighting, or subclassification)
- 4 Assess balance on the covariates of interest; if poor, repeat steps 2-4
- 5 Estimate the treatment effect in the conditioned sample

Covariate balance

- Let $T \in \{0, 1\}$ denote a binary treatment
- Let \mathbf{X} denote confounders
- Ideally, we have that $\mathbf{X}|T = 1$ and $\mathbf{X}|T = 0$ are equal in distribution
- If this holds in original dataset, the dataset looks like what you might find in a randomized study

Covariate balance (cont'd.)

- Typical ways to assess: report standardized mean difference, Kolmogorov-Smirnov statistics, or t-tests
- Covariate balance corresponds to smaller values or less significant p-values

"Propensity score-matching (1:1) was conducted by matching pairs of patients with statin exposure and no exposure with a greedy matching algorithm and a caliper width of 0.2 SD of the log odds of the estimated propensity score. Covariate balance between matched pairs for continuous and dichotomous categorical variables was assessed using the standardized difference, with values equal to or less than 10% indicating minimal imbalance."

A recipe for causal inference

- 1 Define the exposure
- 2 Define the causal parameter of interest
- 3 Model the assignment mechanism using propensity scores
- 4 **Estimate the causal effect conditioning on the propensity score in some way**
- 5 Get a variance estimate
- 6 Perform sensitivity analyses for unobserved confounding

Inverse weighting estimators

- Use the observation that under SITA and treatment positivity

$$E \left[\frac{T_i Y_i}{e(\mathbf{X}_i)} \right] = E[Y_i(1)]$$

and

$$E \left[\frac{(1 - T_i) Y_i}{1 - e(\mathbf{X}_i)} \right] = E[Y_i(0)]$$

Inverse weighting estimators (cont'd.)

- The inverse-weighted estimator of average causal effect is given by

$$A\hat{C}E = n^{-1} \sum_{i=1}^n \frac{T_i Y_i}{\hat{e}(\mathbf{X}_i)} - \frac{(1 - T_i) Y_i}{1 - \hat{e}(\mathbf{X}_i)}$$

- These estimators can be sensitive to the presence of extreme weights, i.e., $\hat{e}(\mathbf{X}_i)$ is close to zero or one

One hybrid approach: double-robust estimation

- Combine inverse weighting and regression adjustment
- The double-robust estimator is given by

$$\begin{aligned} \widehat{ACE}_{DR} = & n^{-1} \sum_{i=1}^n \frac{T_i Y_i}{\hat{e}(\mathbf{X}_i)} - \frac{T_i - \hat{e}(\mathbf{X}_i)}{\hat{e}(\mathbf{X}_i)} \hat{m}_1(\mathbf{X}_i) \\ & - n^{-1} \sum_{i=1}^n \frac{(1 - T_i) Y_i}{1 - \hat{e}(\mathbf{X}_i)} + \frac{T_i - \hat{e}(\mathbf{X}_i)}{1 - \hat{e}(\mathbf{X}_i)} \hat{m}_0(\mathbf{X}_i), \end{aligned}$$

where $\hat{m}_1(\mathbf{X})$ and $\hat{m}_0(\mathbf{X})$ are the estimated functions for $E(Y|T = 1, \mathbf{X})$ and $E(Y|T = 0, \mathbf{X})$, respectively

- Asymptotic theory can give variance (or bootstrap)

Double-robust estimation

- Very popular in statistical research
- Note that $\hat{m}_t(\mathbf{X})$ ($t = 0, 1$) are the fitted values from a regression model of Y on \mathbf{X} in the $T = 0$ and $T = 1$ subgroups
- \hat{m}_t are used in the regression estimators for causal effects
- **Pro:** Double-robust estimator can be consistent if either propensity or outcome regression is correctly specified.
- **Con:** Double-robust estimator can perform poorly if both are misspecified.

Matching estimators

- Based on the estimated propensity score, for each subject with $T = 1$, try to find one (or more) subjects with similar propensity score with $T = 0$ and compare their responses; call this the matched set for subject i
- Estimate average causal effect as

$$n_t^{-1} \sum_{i:T_i=1} (Y_i - (m_i)^{-1} \sum_{j \in M_i: T_j=0} Y_j),$$

where M_i is the matched set for subject i , and n_t is the number of subjects with $T = 1$

- A more robust approach is to coarsen the propensity score using coarsening (subclassification)

- After creating a matched pair dataset using propensity scores, the authors estimated effects using matched pair methods
- Propensity score used to match observations
- Reduction in size (from $>180,000$ to 96,486)

A recipe for causal inference

- 1 Define the exposure
- 2 Define the causal parameter of interest
- 3 Model the assignment mechanism using propensity scores
- 4 Estimate the causal effect conditioning on the propensity score in some way
- 5 **Get a variance estimate**
- 6 Perform sensitivity analyses for unobserved confounding

Inference for surveys (finite-sample) versus superpopulation

- Recall earlier data structure:

$Y(0)$	$Y(1)$	T	X_1	\dots	X_p
?	y_{11}	1	x_{11}	\dots	x_{1p}
y_{20}	?	0	x_{21}	\dots	x_{2p}
?	y_{31}	1	x_{31}	\dots	x_{3p}
?	y_{41}	1	x_{41}	\dots	x_{4p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
y_{n0}	?	0	x_{n1}	\dots	x_{np}

- In this setup, the question marks occur randomly, but other than that, there is no randomness
- This is the finite-sample setup.

Finite-sample inferences

- Based on randomization distribution of T (i.e., shuffle the label of T)
- Use estimated standard deviations conditional on propensity score

From finite-sample to superpopulation

- Everything discussed so far has been a finite-sample mode of inference
- Generalize to the superpopulation
- Idea now is that the observed units are a random sample from a 'superpopulation'
- Recall $\tau_{fs} = n^{-1} \sum_{i=1}^n \{Y_i(1) - Y_i(0)\}$
- Expectations are now taken with respect to the superpopulation sampling mechanism and not T

Superpopulation inferences

- Extensive derivations can be found in the Appendix of Chapter 6 of Imbens and Rubin (2015)
- Key results:

$$E(\hat{\tau}^{dif}) = E(Y(1) - Y(0))$$

$$Var(\hat{\tau}^{dif}) = \frac{\sigma_c^2}{n_c} + \frac{\sigma_t^2}{n_t},$$

where σ_c^2 is the variance of $Y(0)$ with respect to the superpopulation sampling, and σ_t^2 is the variance of $Y(1)$ with respect to the superpopulation sampling

- After creating a matched pair dataset using propensity scores, the authors estimated variances of effects using matched pair methods
- Random effects models are superpopulation in spirit
- Conditional methods are finite-sample in spirit.

A recipe for causal inference

- 1 Define the exposure
- 2 Define the causal parameter of interest
- 3 Model the assignment mechanism using propensity scores
- 4 Estimate the causal effect conditioning on the propensity score in some way
- 5 Get a variance estimate
- 6 **Perform sensitivity analyses for unobserved confounding**

Sensitivity Analysis: basic setup

- Assume that T is **not** conditionally independent of $\{Y(0), Y(1)\}$ given \mathbf{X} (i.e., SITA violated)
- Assume that $T \perp \{Y(0), Y(1)\} | \mathbf{X}, U$, where U has a random variable
- Treat U as a 'synthetic' covariate (i.e., simulated and under the control of the user.
- Write down a model for $Y | \mathbf{X}, U$ and $T | \mathbf{X}, U$
- Let the coefficient for U vary and estimate the coefficients for \mathbf{X} in the two models.
- Hard to implement, as it requires a grid search

Sensitivity Analysis: simplification

- One simplification is due to Lin et al. (1998, Biometrics)
- Idea: use model misspecification theory
- Recall that we have invoked consistency for parameter estimates:

$$\hat{\theta} \rightarrow_p \theta,$$

or in words “theta hat” (estimate) converges in probability to θ (true value of the parameter)

- This assumes that the model for the data is correctly specified
- What happens when this is not true?

- When model being fit to data is false

$$\hat{\theta} \rightarrow_p \theta^*,$$

where θ^* is called the least false parameter

- θ^* minimizes the Kullback-Leibler divergence:

https://en.wikipedia.org/wiki/Kullback-Leibler_divergence

- Cast the violation of SITA as one of fitting the incorrect model:
- True model: Y depends on T , \mathbf{X} and U
- False model: Y depends on T and \mathbf{X}
- This approach allows for some simple approximate sensitivity analysis formulae

- Authors performed several sensitivity analyses
- Interestingly, nothing for unmeasured confounding

Cutting Edge Topic # 1: Transportability

- Goal: take effect from randomized study and see how it carries forward to an observational study
- Some initial key papers:
 - 1 Cole and Stuart, 2010, AJE
 - 2 Barenboim and Pearl, 2016, PNAS
 - 3 Peters et al., 2016, JRSS-B
 - 4 Dahabreh et al., 2020, Statistics in Medicine

Cutting Edge Topic # 2: High-dimensional mediation analysis

- Canonical approach due to Baron and Kenny (1986)
- Some initial key papers:
 - 1 Huang and Pan, 2016, Biometrics
 - 2 Chén et al., 2018, Biostatistics
 - 3 Aung et al., 2020, Nature Communications

Summary

- Causal inference aims to use observational data in order to mimic a randomized experimental design using observed covariates
- The propensity score plays the role of the coin flip (probability based on observed covariates)
- We have given a six-step pipeline for analysis
- Note that many unverifiable assumptions are needed for performing causal inference, and thus, sensitivity analysis step is a key.
- Lots of interesting new problems, e.g., transportability and high-dimensional mediation

References

- Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74**, 235 – 267.
- Imbens and Rubin. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences*
Cambridge: Cambridge University Press.
- Lin, D. Y., Psaty, B. M., and Kronmal, R. A. (1998). Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies. *Biometrics*, 54(3), 948-963.
- London, M. J., Schwartz, G. G., Hur, K. and Henderson, W. G. Association of Perioperative Statin Use With Mortality and Morbidity After Major Noncardiac Surgery. *JAMA Intern Med.* 2017 Feb 1;177(2):231-242.
- Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science* **25**,1 – 21.