Causal Inference

a summary

Contents

1 General

Causal Roadmap (Petersen and van der Laan, 2014)
systematic approach linking causality to statistical procedures
1. Specifying Knowledge. structural causal model (unifying counterfactual language, structural equations, & causal graphs): a set of possible data-generating processes, expresses background knowledge and its limits
2. Linking Data. specifying measured variables and sampling specifics (latter can be incorporated into the model)
3. Specifying Target. define hypothetical experiment: decide
1. variables to intervene on: one (point treatment), multiple (longitudinal, censoring/missing, (in)direct effects)
2. intervention scheme: static, dynamic, stochastic
3. counterfactual summary of interest: absolute or relative, marginal structural models, interaction, effect modification
4. population of interest: whole, subset, different population

4. Assessing Identifiability. are knowledge and data sufficient to derive estimand and if not, what else is needed?
5. Select Estimand. current best answer: knowledge-based assumptions + which minimal convenience-based asspumptions (transparency) gets as close as possible
6. Estimate. choose estimator by statistical properties, nothing causal here
7. Interpret. hierarchy: statistical, counterfactual, feasible intervention, randomized trial

Notation chapter 1.1

average causal effect chapter 1.2 and 1.3 and 1.4 and 1.5

randomized experiments (target trial) 2.1 and 2.2; 3.6

Standardization plug-in (or parametric) g-formula

$$\mathrm{E}\left[Y|A=a\right]=\int \mathrm{E}\left[Y|L=l,A=a\right]dF_L\left[l\right]$$

weighted average of stratum-specific risks; unknowns can be estimated non-parametrically or modeled
no need to estimate $f_L\,[l]$ as empirical distribution can be used: estimate outcome model → predict counterfactuals → average the results (→ CI by bootstrapping)
for discrete L $\mathrm{E}\left[Y|A=a\right]=\sum_l \mathrm{E}\left[Y|L=l,A=a\right]\Pr\left[L=l\right]$

IP Weighting adjust for (surrogate) confounders L

$$\mathrm{E}\left[Y|A=a\right]=\mathrm{E}\left[\frac{I(A=a)Y}{f\left[A|L\right]}\right];W^A=\frac{1}{f\left[A|L\right]};SW^A=\frac{f(A)}{f\left[A|L\right]}$$

unknowns can be estimated non-parametrically or modeled
pseudo-population: everyone is treated & untreated ($L \not\to A$)
FRCISTG *(fully randomized causally interpreted structured graph)*: probability tree for $L \to A \to Y$, can be used to calculate/visualize simulation of values for A
for discrete A, L $f\left[a|l\right]=\Pr\left[A=a,L=l\right]$
estimators: Horvitz-Thompson; Hajek (modified version)
stabilized weights SW^A should have an average of 1 (check!)
→ pseudo-population same size → CI width ↓

Standardization and IP Weighting are equivalent,
but if modeled, different "no misspecification" assumptions:
standardization: outcome model
IP weighting: treatment model
doubly robust estimators: reduce model misspecification bias, consistent if either model is correct; ***e. g.:***
1. fit outcome regression with variable $R=\begin{cases}+W^A & \text{if } A{=}1\\ -W^A & \text{if } A{=}0\end{cases}$
2. calculate counterfactual results

identifiability conditions most of 3
positivity: p. 155, p. 162
additional conditions: chapter 13.5
exchangeability: p 172f

effect modification chapter 4

interaction chapter 5

causal diagrams chapter 6, include swigs from 7.5 and that one technical point

confounding chapter 7

selection bias chapter 8

measurement bias chapter 9

random variabilty chapter 10

2 Models

Modeling data are a sample from the target population

estimand:	quantity of interest,	e. g. $\mathrm{E}[Y\|A=a]$
estimator:	function to use,	e. g. $\widehat{\mathrm{E}}[Y\|A=a]$
estimate:	apply function to data,	e. g. 4.1

model: a priori restriction of joint distribution/dose-response curve; *assumption:* no model misspecification (usually wrong)

non-parametric estimator: no restriction (saturated model) = *Fisher consistent estimator* (entire population data → true value)

parsimonious model: few parameters estimate many quantities

bias-variance trade-off:
wiggliness ↑ → misspecification bias ↓, CI width ↑

Marginal Structural Models association is causation in the IP weighted pseudo-population

$$\text{associational model } \mathrm{E}[Y|A] = \text{ causal model } \mathrm{E}[Y^a]$$

step 1: estimate/model $f[A|L]$ (and $f[A]$) → get $(S)W^A$

step 2: estimate regression parameters for pseudo-population

effect modification variables V can be included (e. g. $\beta_0 + \beta_1 a + \beta_2 Va + \beta_3 V$; technically not marginal anymore), $SW^A(V) = \frac{f[A|V]}{f[A|L]}$ more efficient than SW^A

Censoring measuring joint effect of A and C

$$\mathrm{E}\left[Y^{a,c=0}\right] \text{ is of interest}$$

standardization $\mathrm{E}[Y|A=a] = \int \mathrm{E}[Y|L{=}l, A{=}a, C{=}0]\, dF_L[l]$

IP weights $W^{A,C} = W^A \times W^C$ (uses n) or $SW^{A,C} = SW^A \times SW^C$ (uses $n^{c=0}$)

g-estimation can only adjust for confounding, not selection bias → use IP weights

G-Methods *g*eneralized treatment contrasts: adjust for (surrogate) confounders L

- **standardization**
- **IP weighting**
- **g-estimation:** not needed unless longitudinal

G-Estimation (additive) structural nested models

$$\text{logit}\Pr\left[A=1|H(\psi^\dagger), L\right] = \alpha_0 + \alpha_1 H(\psi^\dagger) + \alpha_2 L$$

$$H(\psi^\dagger) = Y - \psi_\dagger A$$

find $\psi^\dagger$ which renders $\alpha_1 = 0$; 95 %-CI: all $\psi^\dagger$ for which $p > 0.05$
closed-form solution for linear models

derivation: $H(\psi^\dagger) = Y^{a=0}$

$$\text{logit}\Pr\left[A=1|Y^{a=0}, L\right] = \alpha_0 + \alpha_1 Y^{a=0} + \alpha_2 L$$

$Y^{a=0}$ unknown, but because of exchangeability α_1 should be zero

$$Y^{a=0} = Y^a - \psi_1 a$$

equivalent to $Y^{a=0} = Y^{a=1} - \psi_1$, but using no counterfactuals

structural nested mean model

$$\text{additive:} \quad \mathrm{E}\left[Y^a - Y^{a=0}|A=a, L\right] = \beta_1 a\,(+\beta_2 aL)$$

$$\text{multiplicative:} \quad \log\left(\frac{\mathrm{E}[Y^a|A=a, L]}{\mathrm{E}[Y^{a=0}|A=a, L]}\right) = \beta_1 a\,(+\beta_2 aL)$$

multiplicative is preferred if Y always positive, but does not extend to longitudinal case
semi-parametric: agnostic about β_0 and effect of L → robust ↑
no time-varying: no nesting; model equals marginal structural models with missing β_0, β_3 (unspecified "no treatment")
sensitivity analysis: unmeasured confounding ($\alpha_1 \neq 0$) can be examined: do procedure for different values of α_1 → plot α_1 vs. $\psi^\dagger$ → how sensitive is estimate to unmeasured confounding?
effect modification: add V in both g-estimation equations
doubly robust estimators exist

Outcome regression chapter 15

instrumental variable estimation chapter 16

causal survival analysis chapter 17

Variable Selection can induce bias if L includes:

(decendant of) collider:	*selection bias under the null*
noncollider effect of A:	*selection bias under the alternative*
mediator:	*overadjustment for mediators*

temporal ordering is not enough to conclude anything
bias amplification: e.g. by adjusting for an instrument Z (can also reduce bias)

Machine Learning L is high-dimensional
use lasso or ML for IP weighting/standardization
but: ML does not guarantee elimination of confounding and has largely unknown statistical properties
→ **doubly robust estimator:** consistent if bias $< \frac{1}{\sqrt{n}}$
sample splitting: train estimators on training sample, use resulting estimators for doubly robust method on estimation sample (CIs on estimation sample are valid, but n halved)
cross-fitting: do again the other way round, average the two estimates, get CI via bootstrapping
problems: unclear choice of algorithm, is bias small enough?

3 Longitudinal Data

Time-Varying Treatments compare 2 treatments
treatment history up to k: $\bar{A}_k = (A_0, A_1, ..., A_k)$
shorthand: always (never) treated $\bar{A} = \bar{1}$ ($\bar{0}$)
- **static strategy:** $g = [g_0(\bar{a}_{-1}), ..., g_K(\bar{a}_{K-1})]$
- **dynamic strategy:** $g = \left[g_0(\bar{l}_0), ..., g_K(\bar{l}_K)\right]$
- **stochastic strategy:** non-deterministic g

optimal strategy is where $\mathrm{E}[Y^g]$ is maximized (if high is good)

identifiability middle chapter 19

treatment-confounder feedback end chapter 19 and chapter 20

g-formula chapter 21.1

IP weighting chapter 21.2

doubly robust estimators chapter 21.3

g-estimation chapter 21.4

censoring chapter 21.5

target trial chapter 22 (does that even really fit in here, maybe push to 3rd paragraph in without models)

References

If no citation is given, the source is (Hernán and Robins, 2023)

Hernán, M. A. and Robins, J. M. (2023). *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC.

Petersen, M. L. and van der Laan, M. J. (2014). Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology (Cambridge, Mass.)*, 25(3):418–426.