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 adjust for (surrogate) confounders L

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

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 $\mathrm{E}[Y|A=a] = \sum_l \mathrm{E}[Y|L=l, A=a]\Pr[L=l]$

IP Weighting adjust for (surrogate) confounders L

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

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 $f[a|l] = \Pr[A=a, L=l]$
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-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

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

because of exchangeability α_1 should be zero

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

holds because ...

e.g. average causal effect in certain strata E(Y1—L) - E(Y0—L)

model P(A—Y(a=0), L) = a0 + a1Y(a=0) + a2L, if exchangability holds a1=0 [Y(a=0) unknown]

model is only nested when treatment is time-varying, if no time-varying, snm equal to semiparametric marginal structural models (fp 14.1)

E(Y1—L) - E(Y0—L) = E(Y1 - Y0—L) therefore P(Ya -Y(a=0)— A=a, L) = a1a + a2aL

semi-parametric as they are agnostic about a0 and effect of L alone, therefore fewer assumptions, more robust

g-estimation can only handle confounding, not selection bias, therefore IP weights for censoring

multiplicative structural nested model (tp 14.1)

rank preservation (implausible): ranking Y in both counterfactual worlds gives same order (effect is basically the same for all; either additive or multiplicative)

additive conditional rp: Yai -Y0i = w1a + w2aLi for all individuals i

g-estimation is easier to understand for rank-preserving models

suppose goal: E(Ya -Y(a=0)— A=a, L) = b1 a

assume rank-preserving Yai -Y0i = w1a is correct

Y0 = Ya -w1a (rewritten)

Y0 = Y - w1A = H

put in formula from above

P(A—Y(a=0), L) = a0 + a1 H + a2L

find which w1 makes a1 be zero

95% CI: all w1 for which w1=0? had result of greater than 0.05

we do not require rank preservation

if there is unmeasured confounding (a1 not zero) we can use the correct a1 and still use g-estimation. If we are not sure about a1, we can try some values (sensitivity analysis)

effect modification can be added into formula

for linear mean models: closed form, otherwise search

doubly robust estimators exist

Outcome regression chapter 15

instrumental variable estimation chapter 16

causal survival analysis chapter 17

variable selection beginning of chapter 18

machine learning in CI end of chapter 18

3 Longitudinal Data

time-varying treatments beginning chapter 19

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.