causaloptim

Introduction to the package

Michael C Sachs, Section of Biostatistics,
2024-09-03

UNIVERSITY OF COPENHAGEN

Crash course in causal inference

The causal roadmap:

1. Translate the scientific question of interest into a formal causal parameter – optional: define the ideal study
2. Specify a model for the generating mechanism of the observed data, with the causal parameter in mind – i.e., a directed acyclic graph (DAG) and some other assumptions
3. State the identifying assumptions, and, derive the statistical parameter
4. Estimate it and do inference

Point 3: under what conditions can the observed data narrow down the causal parameter to a single value?
Sometimes, under the most reasonable set of assumptions, not even an infinite amount of data can narrow down the causal parameter to a single point.

A causal parameter

Suppose we have a new treatment that we think will prolong life. How do we measure its efficacy?

I would say let X represent the treatment indicator, 0 for the standard of care, and 1 for the new treatment.

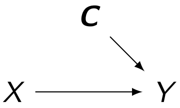
Let Y represent a bad outcome, 0 for alive at 3 years, 1 if dead within 3 years.

$Y(X = 1)$ denotes the outcome if X were intervened upon to have value 1, called a potential outcome. $Y(X = 0)$ also exists.

One way to measure the efficacy of X on Y is with the causal risk difference/average treatment effect:

$$p\{Y(X = 1) = 1\} - p\{Y(X = 0) = 1\} := \theta.$$

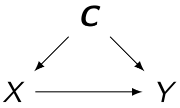
DAGs and assumptions – random treatment assignment



This is a functional causal model $\{F_V : pa(V) \to V \mid V \in \mathcal{V}\}$ for the measured variables $\mathcal{V} = \{C, X, Y\}$

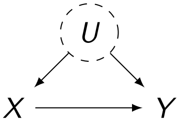
Here we assume that no variable influences X, so the observation $Y|X = x$ coincides with the potential outcome $Y(X = x)$ for $X \in \{0, 1\}$. Thus we can estimate θ easily using the observed data, i.e., it is *identifiable*.

DAGs and assumptions – confounding



Here, $Y = F_Y(X, C)$ and $X = F_X(C)$, there is confounding.
We can still estimate θ easily using the observed data, it is just a little more complicated, we have to adjust for $\boldsymbol{C}$.

DAGs and assumptions – unmeasured confounding



Here, $Y = F_Y(X, U)$ and $X = F_X(U)$, there is confounding, but we are in trouble because U is not measured (as indicated by the dashed circle).
Not even knowing the true probabilities $\boldsymbol{p} = \{p(Y = y, X = x) \text{ for } y, x \in \{0, 1\}\}$ suffices to point identify θ. Instead, we aim to get an upper and lower bound for θ:

$$L(\boldsymbol{p}) \leq \theta \leq U(\boldsymbol{p}).$$

This is called partial identification, or nonparametric causal bounding.

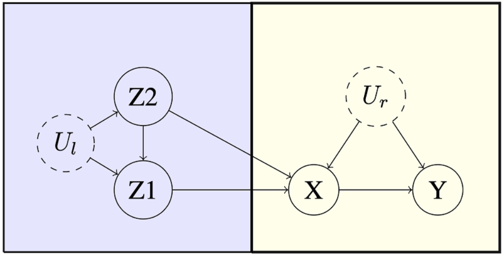
What does `causaloptim` do?

- Provides a framework for deriving $L(\boldsymbol{p}), U(\boldsymbol{p})$ in specific scenarios.
- Users specify a DAG (and/or other assumptions), and a parameter of interest
- They get out $L(\boldsymbol{p}), U(\boldsymbol{p})$ for that parameter under that DAG.
- These are *symbolic* nonparametric bounds: expressions in terms of estimable probabilities, not specific numbers.
- The method only works under certain constraints on the DAG and the parameter of interest.

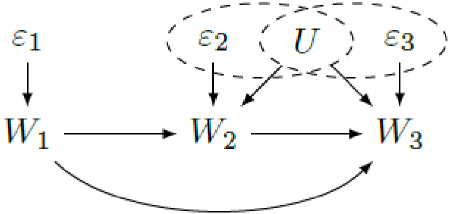
Demo

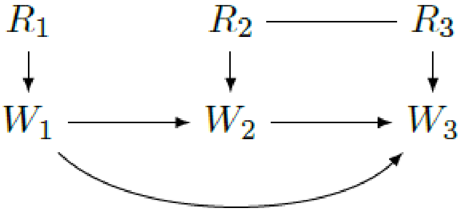
How does it work?

- Graph is split into left and right sides
- Observed variables are categorical, unobserved can be anything
- Complete confounding within each side but paths connecting sides are unconfounded
- Assume that we observe $p\{\text{right vars}|\text{left vars}\}$ for all levels of the variables



How does it work? (2)





The graph is translated to the equivalent response function variable formulation. Since the observable variables are discrete, all possible response patterns can be enumerated.

How does it work? (3)

Example response functions in the graph $Z \to X; Z \leftarrow U \to X$:

Response Pattern	Function	Interpretation
$r = 0$	$f_X(z, 0) = 0$	never takes X, regardless of Z
$r = 1$	$f_X(z, 1) = z$	full compliance with assignment Z
$r = 2$	$f_X(z, 2) = 1 - z$	total defiance of assignment Z
$r = 3$	$f_X(z, 3) = 1$	always takes X, regardless of Z

We define a vector of unobservable parameters $\boldsymbol{q}$ each element of which represents the probability of a particular response pattern.

How does it work? (4)

for $\boldsymbol{p}$ the observable probabilities and $\boldsymbol{q}$ the unobservable response function probabilities, we have a linear system of equations

$$\boldsymbol{p} = R\boldsymbol{q} \text{ for some matrix } R \in \{0,1\}^{m \times n}.$$

Algorithm 1 of Sachs et al. [2023] gives a way to find the matrix R.
That paper also shows that these linear constraints are complete for all graphs that meet our criteria.

How does it work? (5)

Further, we can express potential outcomes of Y under intervention on X in terms of the parameters $\boldsymbol{q}$ by the adjustment formula;

$$P(Y(X = x) = y) = \sum_{r_X, r_Y} P(y|x, r_Y) q_{(r_X, r_Y)}$$

Hence, we have

$$\begin{aligned} \nu &= P(Y(X = 1) = 1) - P(Y(X = 0) = 1) \\ &= c^T \boldsymbol{q} \text{ for some vector } c \in \{0, 1, -1\}^n \end{aligned}$$

Algorithm 2 of Sachs et al. 2022 gives a way to find the c vector and describes conditions under which it is linear.

Linear Programming

Now we have a set of linear constraints as well as our effect of interest in terms of q and we are ready to optimize. The following LP gives a tight lower bound on ν:

$$\begin{aligned} \text{min} \quad & c^T q \\ st \quad & \Sigma q = 1 \\ \& \quad & Rq = p \\ \& \quad & q \geq 0 \end{aligned}$$

This "primal" problem is stated in terms of the q s, which are not estimable. We want a solution that gives us an expression in terms of the observable p s, thus we convert to the dual problem.

Linear Programming continued

- By *the Strong Duality Theorem* of convex optimization, the optimal value of this primal problem equals that of its dual.
- Furthermore, its constraint space is a convex polytope and by *the Fundamental Theorem of Linear Programming*, this optimum is attained at one of its vertices.
- Thus, we enumerate the vertices of the dual problem which are in terms of the p s. We use the *Double description method* as implemented in `cddlib` (very fast!)

The vertices correspond to the expressions given in the min/max output of the bounds.

About the package development

- Came out of a close collaboration with Arvid Sjölander and Erin Gabriel since 2017 when we were all at Karolinska in Sweden
- Needed help getting Balke's `C++` code from 1996 running again
- First released on Github in 2019 and on CRAN in 2020
- We've used it ourselves and "hacked" it to do some additional things
- Some of these hacks are now features in an unreleased version that will be close to a 1.0.0 release (stable API).

Some examples of what we've used it for: [Gabriel et al., 2021, 2022, 2023, 2024b,a]

It also includes Balke's original code from his thesis, which was previously unreleased

Some of the new hacks/features:

- Causal model object creation and tools

→ E.g., sampling from model, deriving observable constraints, testing linearity, testing observable constraints

- Flexibility in specifying causal model and observables
- Testing linearity of causal effect

Observable constraints

$\boldsymbol{p} = R\boldsymbol{q}$ relates the observables to the response pattern probabilities.

- Geometrically, it also describes a convex polytope by the locations of its vertices (the V-representation)
- These vertices are related to the bounds.

Another view of this same polytope gives us *observable constraints* that can be used to falsify the causal model. These are called the instrumental inequalities in the IV model.

- The other view is the H-representation, which describes the polytope in terms of its faces.
- These faces are described by inequalities in terms of the observable probabilities.

The was described by Bonet [2001] and we have implemented it in the causal model object creation.

New features demo

The future of `causaloptim`

- Plans to enhance flexibility and applicability
- Extend to non-linear cases – tricks for symbolic optimization, and numeric optimization in other cases
- Inference on the bounds (confidence intervals)
- Integrate into a complete causal pipeline

This is thanks to the creation of the new **Pioneer Center for SMART Biomed**, locally headed by Erin Gabriel. It specifically includes funding for the development of high-quality software for research on common complex diseases.

There are lots of opportunities to get involved. See more: `https://smartbiomed.dk/`

Discussion

What have you used `causaloptim` for?

Feature wishes?

Would you use it now if you haven't before?

What's the best wine from your region?

References I

Blai Bonet. Instrumentality tests revisited. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 48–55, 2001.

Erin E Gabriel, Arvid Sjölander, and Michael C Sachs. Nonparametric bounds for causal effects in imperfect randomized experiments. *Journal of the American Statistical Association*, pages 1–9, 2021.

Erin E Gabriel, Michael C Sachs, and Arvid Sjölander. Causal bounds for outcome-dependent sampling in observational studies. *Journal of the American Statistical Association*, 117(538): 939–950, 2022.

Erin E Gabriel, Arvid Sjölander, Dean Follmann, and Michael C Sachs. Cross-direct effects in settings with two mediators. *Biostatistics*, 24(4):1017–1030, 2023.

Erin E Gabriel, Michael C Sachs, and Andreas Kryger Jensen. Sharp symbolic nonparametric bounds for measures of benefit in observational and imperfect randomized studies with ordinal outcomes. *Biometrika*, page asae020, 04 2024a. doi: 10.1093/biomet/asae020. URL `https://doi.org/10.1093/biomet/asae020`.

References II

Erin E. Gabriel, Michael C. Sachs, and Arvid Sjölander. The impact of coarsening an exposure on partial identifiability in instrumental variable settings, 2024b. URL https://arxiv.org/abs/2401.17735.

Michael C Sachs, Gustav Jonzon, Arvid Sjölander, and Erin E Gabriel. A general method for deriving tight symbolic bounds on causal effects. *Journal of Computational and Graphical Statistics*, 32 (2):567–576, 2023.