Validating Causal Models

Dustin Tran†, Francisco R. Ruiz†, Susan Athey‡, David Blei†

†Columbia University, ‡Stanford University

Summary

- All causal inference requires assumptions—more so than for standard tasks in probabilistic modeling.
- Testing those assumptions is important to assess the validity of a causal model.
- We develop Bayesian model criticism for causal inference.
- We show how to separately criticize (1) the model of treatment assignments, (2) the model of outcomes, and (3) the central assumption of unconfoundedness.

Causal Models

- A *causal model* is a joint distribution of the potential outcomes $\mathbf{y}(0), \mathbf{y}(1)$, assignments $\mathbf{a}$, and governing parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, conditional on covariates $\mathbf{x}$.
- We focus on

$$p(\mathbf{y}(0), \mathbf{y}(1), \mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{x}) = \left(p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(y_i(0), y_i(1) \mid x_i, \boldsymbol{\theta})\right) \left(p(\boldsymbol{\phi}) \prod_{i=1}^{n} p(a_i \mid x_i, \boldsymbol{\phi})\right) \tag{1}$$

  The causal model has two components: the *outcome model* $p(\boldsymbol{\theta})p(\mathbf{y}(0), \mathbf{y}(1) \mid \mathbf{x}, \boldsymbol{\theta})$ and the *assignment model* $p(\boldsymbol{\phi})p(\mathbf{a} \mid \mathbf{x}, \boldsymbol{\phi})$.
- It assumes "unconfoundedness", $p(\mathbf{a} \mid \mathbf{y}(0), \mathbf{y}(1), \mathbf{x}) = p(\mathbf{a} \mid \mathbf{x})$.
- **Example.** The *average treatment effect* (ATE) is the expected difference in outcomes

$$\text{ATE} = \mathbb{E}[Y(0)] - \mathbb{E}[Y(1)],$$

  where the expectation is taken across the population of individuals.
- The "fundamental problem of causal inference" is that only one of the outcomes is observed for each data point.
- Ideally, we would collect data from an experiment where each assignment a_i is set according to known assignment parameters $\boldsymbol{\phi}^*$ (e.g., random assignment). This gives an alternative joint on the variables,

$$p(\mathbf{y}(0), \mathbf{y}(1), \mathbf{a}, \boldsymbol{\theta} \mid \mathbf{x}, \boldsymbol{\phi}^*) = \left(p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(y_i(0), y_i(1) \mid x_i, \boldsymbol{\theta})\right) \left(\prod_{i=1}^{n} p(a_i \mid x_i, \boldsymbol{\phi}^*)\right). \tag{2}$$

  We call this the *do model*. It is also called an "intervention" or "mutilation."
- These methods rest on the same assumption: unconfoundedness. Further, these methods also require assumptions on the outcome model $p(\boldsymbol{\theta})p(\mathbf{y}(0), \mathbf{y}(1) \mid \mathbf{x}, \boldsymbol{\theta})$ and the assignment model $p(\boldsymbol{\phi})p(\mathbf{a} \mid \mathbf{x}, \boldsymbol{\phi})$. We describe when and how we can check these assumptions.

Validating Causal Models

The central tool of model criticism is the posterior predictive check (PPC). The procedure is:

1. Design a discrepancy function, a statistic of the data and hidden variables.
2. Form the *realized discrepancy*, which is the statistic applied to observed data (along with posterior samples of hidden variables).
3. Form the *reference distribution*, the distribution of the discrepancy applied to many replicated data sets from the posterior predictive distribution.
4. Check if the realized discrepancy is unlikely to have come from the reference distribution.

Define a *causal discrepancy* to be a scalar function of the form,

$$T((\mathbf{y}(0), \mathbf{y}(1)), \mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\phi}). \tag{3}$$

There are two ingredients to a causal check: the *reference distribution* and the *realized discrepancy*.

Algorithm 1: Criticism of the assignment model

Input: Assignment model $p(\boldsymbol{\phi} \mid \mathcal{D}^{\text{obs}})\, p(\boldsymbol{a}^{\text{rep}} \mid \boldsymbol{x}, \boldsymbol{\phi})$, discrepancy $T(\boldsymbol{a}, \boldsymbol{\phi})$.
Output: Reference distribution $p(T)$ and realized test statistic T^{obs}.
for $s = 1, \ldots, S$ replications **do**
  Draw assignment parameters $\boldsymbol{\phi}^s \sim p(\boldsymbol{\phi} \mid \mathcal{D}^{\text{obs}})$.
  Draw assignments $\boldsymbol{a}^{\text{rep},s} \sim p(\boldsymbol{a}^{\text{rep}} \mid \boldsymbol{x}, \boldsymbol{\phi}^s)$.
  Calculate discrepancy $T^{\text{rep},s} = T(\boldsymbol{a}^{\text{rep},s}, \boldsymbol{\phi}^s)$.
  Calculate discrepancy $T^{\text{obs},s} = T(\boldsymbol{a}, \boldsymbol{\phi}^s)$.
end
Form reference distribution $p(T)$ from replications $\{T^{\text{rep},s}\}$.
Form realized discrepancy T^{obs} from replications $\{T^{\text{obs},s}\}$.

Algorithm 2: Criticism of the outcome model

Input: Causal model $p(\boldsymbol{\theta} \mid \mathcal{D}^{\text{do}})\, p(\boldsymbol{y}(0)^{\text{rep}}, \boldsymbol{y}(1)^{\text{rep}} \mid \boldsymbol{x}, \boldsymbol{\theta})\, p(\boldsymbol{\phi} \mid \mathcal{D}^{\text{obs}})\, p(\boldsymbol{a}^{\text{rep}} \mid \boldsymbol{x}, \boldsymbol{\phi})$, discrepancy $T((\boldsymbol{y}(0), \boldsymbol{y}(1)), \boldsymbol{\theta})$.
Output: Reference distribution $p(T)$ and realized test statistic T^{obs}.
for $s = 1, \ldots, S$ replications **do**
  Draw outcome parameters $\boldsymbol{\theta}^s \sim p(\boldsymbol{\theta} \mid \mathcal{D}^{\text{obs}})$.
  Draw outcomes $\boldsymbol{y}(0)^{\text{rep},s}, \boldsymbol{y}(1)^{\text{rep},s} \sim p(\boldsymbol{y}(0)^{\text{rep}}, \boldsymbol{y}(1)^{\text{rep}} \mid \boldsymbol{\theta}^s)$.
  Calculate discrepancy $T^{\text{rep},s} = T((\boldsymbol{y}(0)^{\text{rep}}, \boldsymbol{y}(1)^{\text{rep}}), \boldsymbol{\theta}^s)$.
  Calculate discrepancy $T^{\text{obs},s} = T\left(\left\{\frac{\delta_{A_i=0}(a_i)}{p(a_i \mid x_i)} y_i(0)\right\}, \left\{\frac{\delta_{A_i=1}(a_i)}{p(a_i \mid x_i)} y_i(1)\right\}, \boldsymbol{\theta}^s\right)$.
end
Form reference distribution $p(T)$ from replications $\{T^{\text{rep},s}\}$.
Form realized discrepancy T^{obs} from replications $\{T^{\text{obs},s}\}$.

Experiments

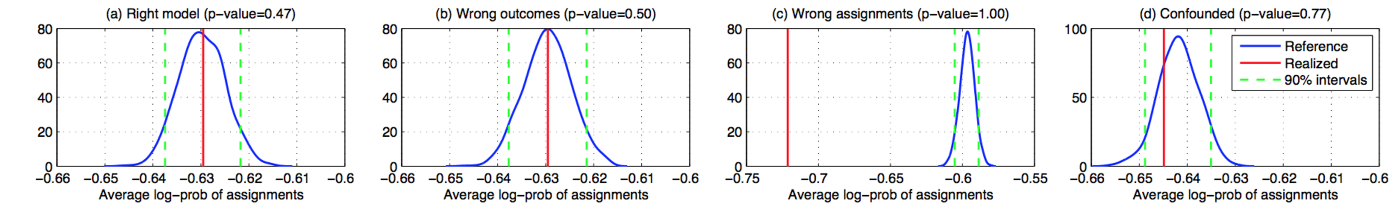
With synthetic data we compare our conclusions to the true mechanism that generated the data. (Real data experiments in paper!)

We generate 10,000 data points, each a 10-dimensional covariate x_i, a binary treatment a_i, and a set of potential outcomes $(y_i(0), y_i(1))$,

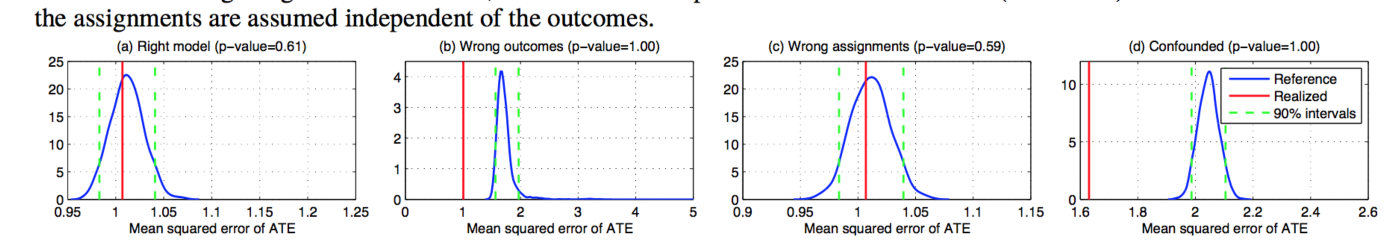
$$\begin{aligned}
x_i &\sim \text{Uniform}(x_i \mid [0,1]^{10}), \\
a_i \mid x_i &\sim \text{Bernoulli}\left(a_i \mid \text{logistic}(x_i^\top \phi)\right), \\
y_i(0) \mid x_i &\sim \mathcal{N}\left(y_i(0) \mid x_i^\top \theta^{(0)}, \sigma^2\right), \\
y_i(1) \mid x_i &\sim \mathcal{N}\left(y_i(1) \mid x_i^\top \theta^{(1)}, \sigma^2\right).
\end{aligned}$$

We place a standard normal prior over the model parameters ϕ, $\theta^{(0)}$, and $\theta^{(1)}$, and a Gamma prior with unit shape and rate on the variance σ^2.

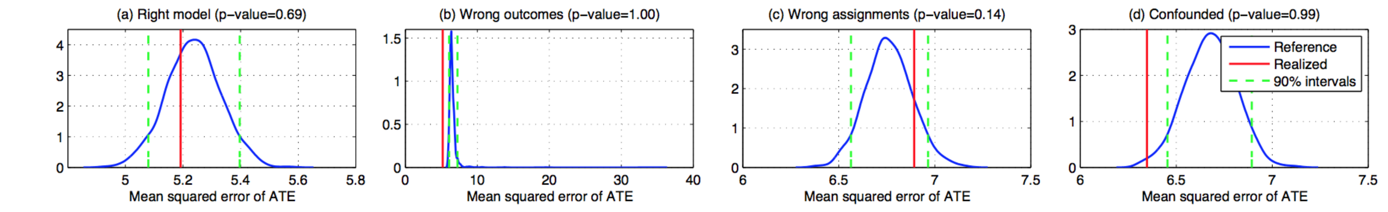
We study two scenarios: (i) In the "science-fiction" scenario, we have access simultaneously to $y_i(0)$ and $y_i(1)$. (This is, of course, not possible in the real world.) (ii) In the "fiction" scenario, we only have access to one counterfactual outcome, $y_i = a_i y_i(1) + (1 - a_i) y_i(0)$.



(a) Results of the assignment test in the science-fiction scenario, in which we have access to both counterfactual outcomes. Model (c), which has a wrong assignment mechanism, fails the test. The plots for the fiction scenario (not shown) are similar to these ones, as the assignments are assumed independent of the outcomes.



(b) Results of the outcome test in the science-fiction scenario. The test fails for the model in which the outcome model is mis-specified and for the confounded model.



(c) Results of the outcome test in the fiction scenario. The test fails for the model in which the outcome model is mis-specified and for the confounded model, and it also seems to suggest a flaw for the model in which the assignment mechanism is wrong.

References

[1] Athey, S. and Imbens, G. (2015). Machine Learning Methods for Estimating Heterogeneous Causal Effects. *arXiv preprint arXiv:1504.01132*.

[2] Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A. General*, 143(4):383–430.

[3] Pearl, J. (2000). *Causality*. Cambridge University Press.

[4] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.