Estimation bayésienne ciblée de modèles structurels marginaux

Herb Susmann[1,2] and Antoine Chambaz[2]

[1]Department of Biostatistics & Epidemiology, University of Massachusetts Amherst et [2]MAP5 (UMR CNRS 8145), Université de Paris

8 décembre, 2021

Introduction

- Slides: `herbsusmann.com/jmb2021`.
- Causal analysis is concerned with estimating *treatment effects*: the effect of an intervention on an outcome.
- We may also be interested in how individual characteristics affect the size of the treatment effect.
- Can we estimate how a set of *treatment effect modifiers* influence the size of the treatment effect?
- Our contribution: asymptotically efficient non-parametric Bayesian estimator

Cheatsheet

- Data structure
  - $W \in \mathbb{R}^d$: vector of covariates
  - $A \in \{0, 1\}$: binary indicator of treatment
  - $Y \in \mathbb{R}$: outcome
  - Observed data: $O = (W, A, Y) \sim P_0$
- Conditional mean:

$$\bar{Q}_P(a, w) = \mathbb{E}_P[Y \mid A = a, W = w]$$

- Conditional treatment effect:

$$\widetilde{Q}_P(w) = \bar{Q}_P(1, w) - \bar{Q}_P(0, w)$$

- Treatment effect modifiers
  - Let $V = (V_1, \ldots, V_k) \subset W$.
  - Let $X = (1, V_1, \ldots, V_k)$. ($X \in \mathbb{R}^p$, with $p = k + 1$.)
- Marginal structural model:

$$\boldsymbol{B}(P) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min}\, \mathbb{E}_P\left[\left(\widetilde{Q}_P(W) - X^\top \boldsymbol{\beta}\right)^2\right]$$

Data Structure

Data structure:

- $W \in \mathbb{R}^d$: vector of covariates
- $A \in \{0, 1\}$: binary indicator of treatment
- $Y \in \mathbb{R}$: outcome

Observed data:

- We assume that the observed data are n i.i.d. draws of $O = (W, A, Y)$ from a distribution P_0.

Average Treatment Effect

- Suppose we are interested in summarizing the effect of the treatment on the outcome over the entire population.
- Notation

$$\bar{Q}_P(a, w) = \mathbb{E}_P[Y \mid A = a, W = w]$$
$$\widetilde{Q}_P(w) = \bar{Q}_P(1, w) - \bar{Q}_P(0, w)$$

- Average Treatment Effect (ATE):

$$\begin{aligned}\Psi_P^{ATE} &= \mathbb{E}_P[\widetilde{Q}_P(W)] \\ &= \mathbb{E}_P[\bar{Q}_P(1, W) - \bar{Q}_P(0, W)] \\ &= \mathbb{E}_P[\mathbb{E}_P[Y \mid A = 1, W] - \mathbb{E}_P[Y \mid A = 0, W]]\end{aligned}$$

Treatment Effect Modification

- What if we are interested in whether the size of the effect is influenced by the covariates?
- Naive approach: estimate the conditional treatment effect for every value of W:

$$\begin{aligned}\widetilde{Q}_P(w) &= \bar{Q}_P(1, w) - \bar{Q}_P(0, w) \\ &= \mathbb{E}_P[Y \mid A = 1, W = w] - \mathbb{E}_P[Y \mid A = 0, W = w]\end{aligned}$$

- Two problems:
  - Could be difficult to estimate for every w, especially if there are strata with few observations.
  - How do we make sense of the results?

Marginal Structural Models

- Alternative approach: summarize the relationship between covariates and treatment effects via a low-dimensional *working model*.
- Let $V = (V_1, \ldots, V_k) \subseteq W$ be a set of possible *treatment effect modifiers*.
- Let $X = (1, V_1, \ldots, V_k)$.
  - $X \in \mathbb{R}^p$, where $p = k + 1$.
- Define $\boldsymbol{B}(P) \in \mathbb{R}^p$ as the solution to the optimization problem

$$\boldsymbol{B}(P) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min}\, \mathbb{E}_P \left[\left(\widetilde{Q}_P(W) - X^\top \boldsymbol{\beta} \right)^2 \right]$$

Marginal Structural Models

$$\boldsymbol{B}(P) = \operatorname*{arg\,min}_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathbb{E}_P \left[\left(\underbrace{\boxed{\widetilde{Q}_P(W)}}_{\text{treatment effect for W}} - X^\top \boldsymbol{\beta} \right)^2 \right]$$

Marginal Structural Models

$$\boldsymbol{B}(P) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min}\, \mathbb{E}_P \left[\left(\widetilde{Q}_P(W) - \underbrace{\boxed{X^\top \boldsymbol{\beta}}}_{\text{linear working model}} \right)^2 \right]$$

Marginal Structural Models

$$\boldsymbol{B}(P) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \underbrace{\boxed{\mathbb{E}_P \left[\left(\widetilde{Q}_P(W) - X^\top \boldsymbol{\beta} \right)^2 \right]}}_{\text{squared-error risk}}$$

Marginal Structural Models

$$\boxed{\boldsymbol{B}(P)} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\, \mathbb{E}_P\left[\left(\widetilde{Q}_P(W) - X^\top \boldsymbol{\beta}\right)^2\right]$$

defined in terms of P

Semi-parametric inference

- Define $\boldsymbol{\beta}_0 \equiv \boldsymbol{B}(P_0)$
- Estimation: we assume only that P_0 falls within the non-parametric model $\mathcal{M}$.
- We wish to construct regular, asymptotically normal and efficient estimator of $\boldsymbol{\beta}_0$.
- We start with a frequentist estimator, which we develop a Bayesian version of.

Semi-parametric inference

- We can write any regular estimator of $\boldsymbol{\beta}_0$ as:

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_0 + \frac{1}{n}\sum_{i=1}^{n} IC_{P_0}(O_i) + o_p(n^{-1/2}),$$

  where IC_{P_0} is called an *influence function* of the parameter $\boldsymbol{\beta}_0$.
- The influence function with the smallest variance is called the *efficient influence function* (EIF). The EIF of $\boldsymbol{\beta}_0$ is given by:

$$D_P(O) = M^{-1}\left[\frac{2A-1}{g_P(A,W)}(Y - \bar{Q}_P(A,W)) + \widetilde{Q}_P(W) - \boldsymbol{\beta}^\top X\right] X,$$

  with $g_P(a,w) \equiv \mathbb{P}_P[A = a | W = w]$ and $M = \mathbb{E}_P[X^\top X]$.
- The semi-parametric efficiency bound for estimating $\boldsymbol{\beta}_0$ is given by $\mathrm{var}_{P_0}(D_{P_0}(O))$

Targeted Maximum Likelihood Estimation

- It turns out we can construct an estimator that achieves this efficiency bound!
- *Targeted Maximum Likelihood Estimation (TMLE)* (van der Laan and Rose, 2011, 2018)
- Plug-in estimator of the form

$$\boldsymbol{\beta}^{TMLE} = \boldsymbol{B}(P_n^0(\boldsymbol{\epsilon}))$$

  where $\left\{P_n^0(\boldsymbol{\epsilon}) : \boldsymbol{\epsilon} \in \mathbb{R}^p\right\}$ is a *fluctuation* of an initial estimator P_n^0 of the pieces of P_0 relevant to $\boldsymbol{\beta}$.
- The TMLE estimator is designed to solve the EIF of the target parameter: $\mathbb{E}_{P_n}[D_{P_n^*}(O)] = 0$.

Bayesian Targeted Maximum Likelihood Estimation

- Can we make this procedure Bayesian?
- Core insight: the fluctuation submodel $\left\{ P_n^0(\epsilon) : \epsilon \in \mathbb{R}^p \right\}$ defines a proper likelihood. So we can use Bayesian inference to estimate ϵ! (Diaz et al., 2011; Díaz et al., 2020)
- Basic application of Bayes rule: posterior distribution of ϵ is given by

$$\Pi_\epsilon(\epsilon \mid O_1, \ldots, O_n) \propto \pi_\epsilon(\epsilon) \prod_{i=1}^{n} p_\epsilon(O_i \mid \epsilon)$$

Bayesian Targeted Maximum Likelihood Estimation

- In practice, we probably don't have any prior information for $\boldsymbol{\epsilon}$.
- Instead, we can put a prior on $\boldsymbol{\beta}$, which we map to a prior on $\boldsymbol{\epsilon}$:

$$\pi_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) = \pi_{\boldsymbol{\beta}}(\boldsymbol{B}(P_n^0(\boldsymbol{\epsilon})))|\det(J(\boldsymbol{\epsilon}))|,$$

where J is the Jacobian of the transformation $\boldsymbol{B}(P_n^0(\boldsymbol{\epsilon}))$.

Bayesian Targeted Maximum Likelihood Estimation

- Sampling techniques such as Markov-Chain Monte Carlo can be used to draw a set of samples from the posterior distribution of $\boldsymbol{\epsilon}$.
- Then, we can map each of these samples of $\boldsymbol{\epsilon}$ to a $\boldsymbol{\beta}$, yielding a set of samples from the posterior distribution of $\boldsymbol{\beta}$.
- Credible intervals can be constructed using the empirical quantiles of the posterior samples.

Bayesian Targeted Maximum Likelihood Estimation

Theorem

Under certain conditions, the posterior distribution converges in total variation:

$$\sup_{\mathbb{B}^p} \left| \Pi_{\boldsymbol{\beta}} \left\{ \sqrt{n}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}) \in \mathbb{B}^p \mid O_n \right\} - N_{\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}), P_n D_n^\top D_n}(\mathbb{B}^p) \right| \to 0$$

where $\mathbb{B}^p$ is a subset of $\mathbb{R}^p$, $N_{\mu,\Sigma}$ is the multivariate normal distribution with mean vector μ and covariance matrix Σ, D_n is the targeted estimate of the EIF, $\bar{\boldsymbol{\beta}} \sim \Pi_{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\beta}}$ is the targeted MLE.

Simulation

- Data generating process P_0:

$$\begin{aligned} W_1, W_2, W_3, V &\sim N_4(0, \boldsymbol{I}_4) \\ A | W_1, W_2, W_3, V &\sim \mathrm{Bernoulli}(\mathrm{logit}^{-1}(W_1 + W_2 + W_3)) \\ Y | A, W_1, W_2, W_3, V &\sim N(W_1 + W_2 + W_3 + 3A + 1.5AV, 0.1) \end{aligned}$$

- True average treatment effect: $\Psi_{P_0}^{ATE} = 3$.
- $X = (1, V)^\top$.
- Optimization problem:

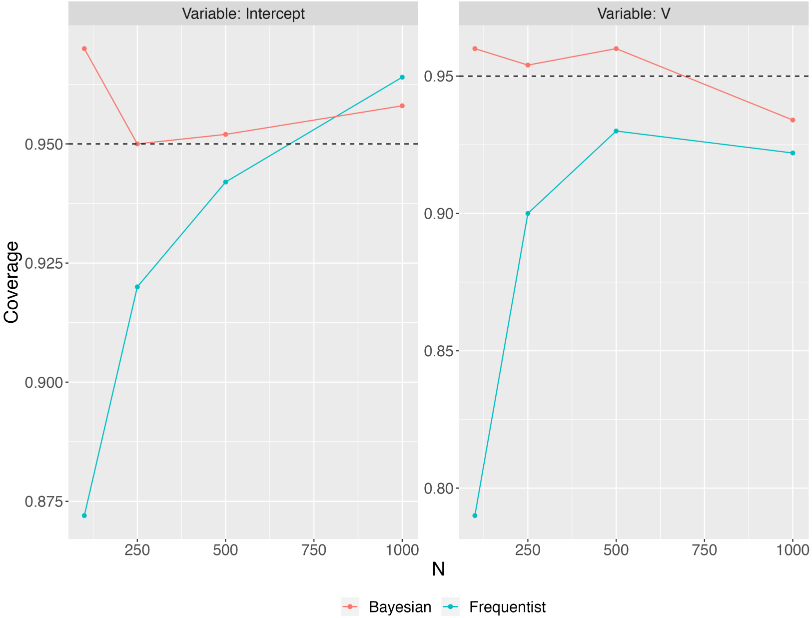
$$\boldsymbol{B}(P) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min}\, \mathbb{E}_P \left[\left(\widetilde{Q}_P(W) - X^\top \boldsymbol{\beta} \right)^2 \right]$$

- Target parameter: $\boldsymbol{\beta}_0 \equiv \boldsymbol{B}(P_0)$
- $\boldsymbol{\beta}_0 \in \mathbb{R}^2$: intercept term and linear effect of V on treatment effect
- For this data generating process, $\boldsymbol{\beta}_0 = (0, 1.5)^\top$.

Simulation

- Simulate 500 datasets for $n = (100, 250, 500, 1000)$.
- Estimate nuisance parameters g_P and $\bar{Q}_P$ with correctly specified regressions.
- Flat priors on β in Bayesian TMLE.
- Frequentist TMLE implemented in *R*, Bayesian TMLE implemented in *R* and *Stan*.
- Julia implementation with automatic differentiation

Simulation



Variable: Intercept
Variable: V
Coverage
0.875
0.900
0.925
0.950
0.80
0.85
0.90
0.95
250
500
750
1000
N
Bayesian
Frequentist

Conclusions

- Marginal Structural Models provide a method for summarizing impact of *treatment effect modifiers*.
- TMLE provides a framework for efficient estimation of Marginal Structural Model parameters within a semi-parametric model.
- Bayesian TMLE brings benefits of both frequentist and Bayesian approaches
- Ongoing work: does the Bayesian estimator inherit other useful properties of the frequentist TMLE? Does incorporation of prior information improve finite-sample performance?

References

I. Diaz, A. E. Hubbard, and M. J. van der Laan. Targeted Bayesian learning. In *Targeted Learning*, pages 475–493. Springer, 2011.

I. Díaz, O. Savenkov, and H. Kamel. Nonparametric targeted Bayesian estimation of class proportions in unlabeled data. *Biostatistics*, 2020.

M. J. van der Laan and S. Rose. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

M. J. van der Laan and S. Rose. *Targeted learning in data science: causal inference for complex longitudinal studies*. Springer, 2018.