

Bayesian Causal Inference in Stan

Arman Oganisian

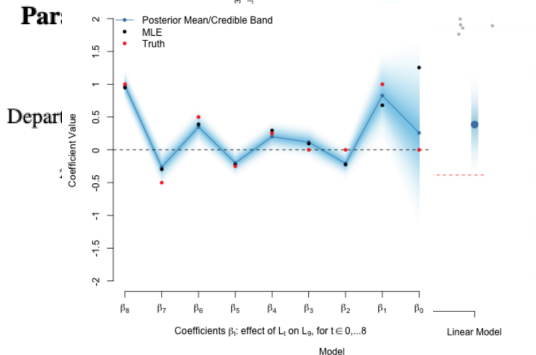
@StableMarkets 

Division of Biostatistics
Department of Biostatistics, Epidemiology, and Informatics
University of Pennsylvania

1/21

REVIEW / TUTORIAL PAPER

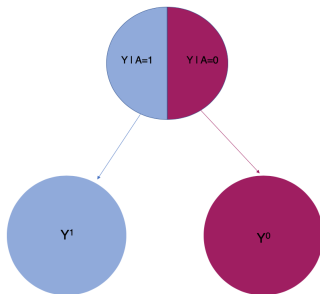
A Practical Introduction to Partial Identification of Causal Effects:



WHAT IS CAUSAL INFERENCE?

What **would have happened** had everyone in the target population if ...

- ▶ ... everyone took treatment 1 versus treatment 0?
- ▶ ... were vaccinated ?
- ▶ ... were enrolled in a job training program?



IDENTIFICATION VIA THE g-FORMULA

$$D = \{Y_i, A_i, L_i, V_i\}_{1:n}.$$

Define potential outcomes Y^a for $a \in \{0, 1\}$

$$\Psi(v) = E[Y^1 - Y^0 \mid V = v]$$

Under some **identification assumptions**

$$E[Y^a \mid V = v] = \int_{\mathcal{L}} \underbrace{E[Y \mid A = a, V = v, L]}_{\text{Regression, } \mu(a, v, l)} \underbrace{dP(L)}_{\text{Confounder}}$$

REGRESSION MODELING

- ▶ Parametric Approaches:

$$\mu(A, V, L) = g^{-1}(\beta_0 + \beta_1 A + \beta_2 V + \beta_3' L)$$

Need priors on β s.

- ▶ Nonparametric Approaches:

$$\mu(A, V, L) = g^{-1}(f(A, V, L))$$

Need prior for f .

WHY BAYES?

- ▶ Priors can help us compute causal effects under sparsity.
- ▶ Avoid *ad hoc* approaches.
- ▶ Powerful suite of nonparametric models (BART, DP, GP, etc).
- ▶ Probabilistic sensitivity analyses.

LOGISTIC MODEL FOR CATEs

Suppose Y is **binary** and $V \in \{1, 2, 3, 4, 5\}$ (e.g., race/ethnicity)

$$Y \mid A, V, L \sim \text{Ber}(\mu(A, V))$$

Specify **logistic regression**

$$\mu(A, V, L) = g^{-1}(\beta_v + \beta'_L L + \theta_v A)$$

with parameters $\omega = (\beta_1 \dots, \beta_5, \beta_L, \theta_1, \dots, \theta_5)$

PRIOR FOR RACE EFFECTS

$$\mu(A, V, L) = g^{-1}(\beta_v + \beta'_L L + \theta_v A)$$

- ▶ Consider “partial pooling” prior:

$$\theta_v \mid \theta^* \sim N(\theta^*, \phi)$$

- ▶ **Shrinkage**: shrinkage race effects towards common effect.
- ▶ **Belief**: the race effects shouldn't be *that* different.
- ▶ **Causal** intuition: small ϕ shrinks towards homogeneity.
- ▶ Note: implies

$$\theta_4 - \theta_5 \sim N(0, 2\phi)$$

As opposed to setting $\phi \approx 0$

$$\theta_4 - \theta_5 \sim \delta_0$$

WE HAVE A DATA MODEL...NOW WHAT?

Suppose we want to compute **Causal Odds Ratio**:

$$\Psi(v) = \frac{E(Y^1 | v) / [1 - E(Y^1 | v)]}{E(Y^0 | v) / [1 - E(Y^0 | v)]}$$

Using g -computation,

$$E(Y^a | v) = \int_{\mathcal{L}} \mu(a, v, L) dP(L)$$

But what about model for $P(L)$?

THE BAYESIAN BOOTSTRAP

- ▶ **Frequentist estimate:** $\hat{P}(L = l) = \sum_{i=1}^n \frac{1}{n} \cdot \delta_{L_i}(l)$
- ▶ **Bayesian model:** $P(L = l \mid p_{1:n}) = \sum_{i=1}^n p_i \cdot \delta_{L_i}(l)$

- ▶ prior:

$$p_{1:n} \sim \text{Dirichlet}(0_n)$$

- ▶ posterior:

$$p_{1:n} \mid L \sim \text{Dirichlet}(1_n)$$

$$E[p_i \mid L] = 1/n$$

FULL MCMC INFERENCE

1. Obtain m^{th} set of posterior draws $\omega^{(m)}$ and for each $A = a$

$$\mu^{(m)}(a, v, L_i) = \beta_v^{(m)} + \beta_L^{(m)} L_i + \theta_v^{(m)} a$$

2. Draw Bayesian Bootstrap weights from posterior:

$$p_1^{(m)}, p_2^{(m)}, \dots, p_n^{(m)} \sim \text{Dirichlet}(1_n)$$

3. Integrate of confounder distribution

$$E^{(m)}(Y^a | v) = \int_{\mathcal{L}} \mu(a, v, L) dP(L) \approx \sum_{i=1}^n \mu^{(m)}(a, v, L_i) p_i$$

4. Compute draw of causal odds ratio:

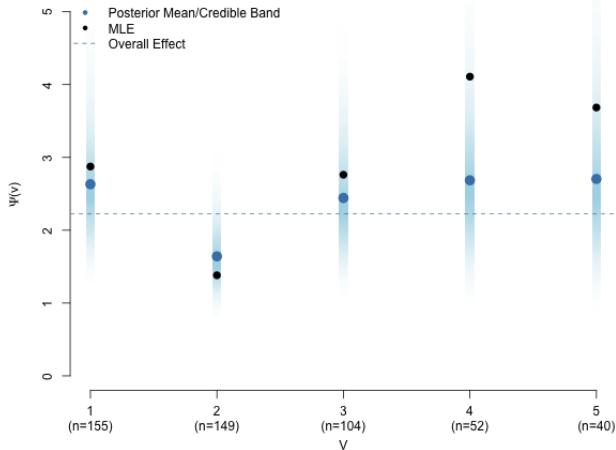
$$\Psi^{(m)}(v) = \frac{E^{(m)}(Y^1 | v) / [1 - E^{(m)}(Y^1 | v)]}{E^{(m)}(Y^0 | v) / [1 - E^{(m)}(Y^1 | v)]}$$

IMPLEMENTATION IN STAN

```
generated quantities {  
  vector[N] bb_weights = dirichlet_rng( rep_vector( 1, N) );  
  
  ...  
  
  for( v in 1:Pv ){  
    for(i in 1:N){  
      cond_mean_y1[i] = inv_logit( L[i]*beta_L + beta_v[v] + theta[v]);  
      cond_mean_y0[i] = inv_logit( L[i]*beta_L + beta_v[v] );  
    }  
    marg_mean_y1 = bb_weights' * cond_mean_y1 ;  
    marg_mean_y0 = bb_weights' * cond_mean_y0 ;  
  
    odds_1 = marg_mean_y1/(1 - marg_mean_y1);  
    odds_0 = marg_mean_y0/(1 - marg_mean_y0);  
    odds_ratio[v] = odds_1 / odds_0;  
  }  
  ...  
}
```

12/21

SYNTHETIC EXAMPLE



SENSITIVITY ANALYSIS

Identification requires **conditional ignorability**

$$Y^a \perp A \mid L, V = v$$

But, what if ignorability is violated?

$$E[Y^a \mid A = 1, L, v] \neq E[Y^a \mid A = 0, L, v]$$

CONSEQUENCE OF VIOLATION

Define,

$$\Delta^a(L) = E[Y^a \mid A = 1, L, v] - E[Y^a \mid A = 0, L, v]$$

$$\int \mu(1, v, L) - \mu(0, v, L) dP(L) = E[Y^1 - Y^0 \mid v] + \xi$$

Estimate of **risk difference** is biased by ξ .

FORM OF VIOLATION

Trade-offs involved in sensitivity analyses

$$\xi = \int \Delta^1(L)(1 - \pi(L)) + \Delta^0(L)\pi(L) dP(L)$$

Simplify $\Delta := \Delta^1 = \Delta^0$ and $\Delta \perp L$. Then,

$$\xi = \Delta$$

Now we can specify priors over Δ .

PRIORS OVER BIAS

Note that $-1 < E[Y^1 - Y^0 \mid v] < 1$ and recall:

$$\Delta = E[Y^a \mid A = 1, L, v] - E[Y^a \mid A = 0, L, v]$$

- ▶ Treated patients **systematically worse**:

$$\Delta \sim U(0, 1)$$

- ▶ Treated patients **systematically better**:

$$\Delta \sim U(-1, 0)$$

- ▶ Biased with **uncertain direction**:

$$\Delta \sim U(-1, 1)$$

MODIFIED MCMC INFERENCE

In Step 3 at m^{th} iteration:

Draw $\Delta^{(m)}$ from the prior and compute,

$$E^{(m)}(Y^a \mid v) = \left\{ \sum_{i=1}^n \mu^{(m)}(a, v, L_i) p_i \right\} + \Delta^{(m)}$$

IMPLEMENTATION IN STAN

- ▶ Could specify prior for Δ in “model” block. Manipulate in “generated quantities”.
- ▶ Could draw Δ from specified distribution in “generated quantities” block.

SOME RESOURCES

- ▶ A Practical Introduction to Bayesian Estimation of Causal Effects: Parametric and Nonparametric Approaches
<https://arxiv.org/pdf/2004.07375.pdf>
- ▶ Companion GitHub repo for paper: https://github.com/stablemarkets/intro_bayesian_causal
- ▶ GitHub Repo for this talk: https://github.com/stablemarkets/StanCon2020_BayesCausal

THANK YOU!

