PURDUE UNIVERSITY

BAYESIAN CAUSAL INFERENCE

Team: Ohnishi Yuki, Satoshi Ido, Noah Hung

Ref: Fan Li, Peng Ding, Fabrizia Mealli: Bayesian Causal Inference: A Critical Review

Background Knowledge

- For each unit i, $\{Y_i(Z_i = 0), Y_i(Z_i = 1), Z_i, X_i\}$

  where outcome Y_i; treatment levels $Z_i \in \{0, 1\}$; a vector of p covariates X_i

- Individual Treatment Effect (ITE) for unit i is $\tau_i = Y_i(Z_i = 1) - Y_i(Z_i = 0)$

- Estimands:
  - SATE: $\tau^s \equiv N^{-1} \sum_{i=1}^{N} \tau_i$
  - CATE: $\tau(x) \equiv \mathbb{E}\{Y_i(1) - Y_i(0) | X_i = x\} = \mu_1(x) - \mu_0(x)$ where $\mu_z(z) \equiv \mathbb{E}\{Y_i(z) | X_i = x\}$ for $z = 0, 1.$
  - PATE: $\tau^p \equiv \mathbb{E}\{Y_i(1) - Y_i(0)\} = \mathbb{E}(\tau(X_i))$

- counterfactual problem
  - Only one of $Y_i(0)$ $Y_i(1)$ can be observation and the other be potential and missing

Background Knowledge

For each unit, one of the potential outcomes is observed and the rest are missing,
→ Causal inference is inherently a missing data problem

→ Bayesian paradigm offers a unified framework for inference with missing data
See all quantities as r.v. and assume them governed by a parameter $\theta = (\theta_X, \theta_Z, \theta_Y)$

Factorize the joint density $P(Y_i(0), Y_i(1), Z_i, X_i|\theta)$ for each unit i as,

$$\underbrace{P\{Z_i|Y_i(0), Y_i(1), X_i; \theta_Z\}}_{\text{assignment mechanism}} \times \underbrace{P\{Y_i(0), Y_i(1)|X_i; \theta_Y\}}_{\text{potential outcomes}} \times \underbrace{P(X_i; \theta_X)}_{\text{covariates}} \quad \text{r.v. are iid}$$

Assume Ignorability

$$\Pr\{Z_i|Y_i(0), Y_i(1), X_i\} = \Pr(Z_i|X_i) \text{ and } 0 < e(X_i) < 1 \text{ for all } i, \text{ where } e(x) \equiv \Pr(Z_i = 1|X_i = x)$$

Under ignorability

$$= P(Z_i|X_i; \theta_Z)$$

Background Knowledge

Assumption

- The parameters are a priori distinct and independent

Posterior distribution of (governed parameters)

$$P(\theta_X, \theta_Y, \theta_Z) \propto P(\theta_X) \prod_{i=1}^{n} P(X_i|\theta_X) \times P(\theta_Z) \prod_{i=1}^{n} P(Z_i|X_i;\theta_Z)$$
$$\times P(\theta_Y) \prod_{i=1}^{n} P(Y_i(1), Y_i(0)|X_i;\theta_Y)$$

- Calculating SATE is more complex because it depends on $Y_i(0), Y_i(1)$

which requires posterior sampling of $\theta_y|Y^{mis}$

Procedure (Iteratively)

1. Simulate $p(\theta_y|Y^{mis}, Y^{obs}, Z, X)$
2. Simulate $P(Y^{mis}|\theta_y, Y^{obs}, Z, X)$

where $p(\theta_y|Y^{mis}, Y^{obs}, Z, X) \propto p(\theta_y) \prod p(Y_i(1), Y_i(0)|X_i;\theta_y)$

$$p(Y^{mis}|\theta_y, Y^{obs}, Z, X) \propto \prod_{i=1, z=1}^{i=n} p(Y_i(0)|Y_i(1), X_i;\theta_y) \prod_{i=1, z=0}^{i=n} p(Y_i(1)|Y_i(0), X_i;\theta_y)$$

Simulation Study

- The outcome function we want to obtain $\mu_z(x) = E[Y_i|X_i = x, Z_i = z; \theta_y]$
- Single model (single tree): $\mu(x, z) = x + z + xz$
- Two model (two tree): $\mu_1(x) = \mu(x, 1); \mu_2(x) = \mu(x, 2)$
  - Linear model:

$$\mu(z, x) = f_z(x) + \epsilon_i \text{ with } \epsilon_i \sim N(0, 1), \text{ where } f_z(x) = \alpha_z + \beta_z(x)$$
$$(\alpha_z, \beta_z) \sim \text{ Gaussian Prior}$$

  - Gaussian Process:

covariance function with signal-to-noise ratio ρ and inverse-bandwidth parameter λ

$$\text{where } \lambda : (f_z(x_1), ..., f_z(x_n))' \sim N(0, \Sigma) \text{ and } \Sigma_{ij} = \delta^2\rho^2 \exp\{-\lambda^2||x_i - x_j||^2\}$$

  - BART Prior

Bayesian version of random forest

Data/ Setup

Synthesis data

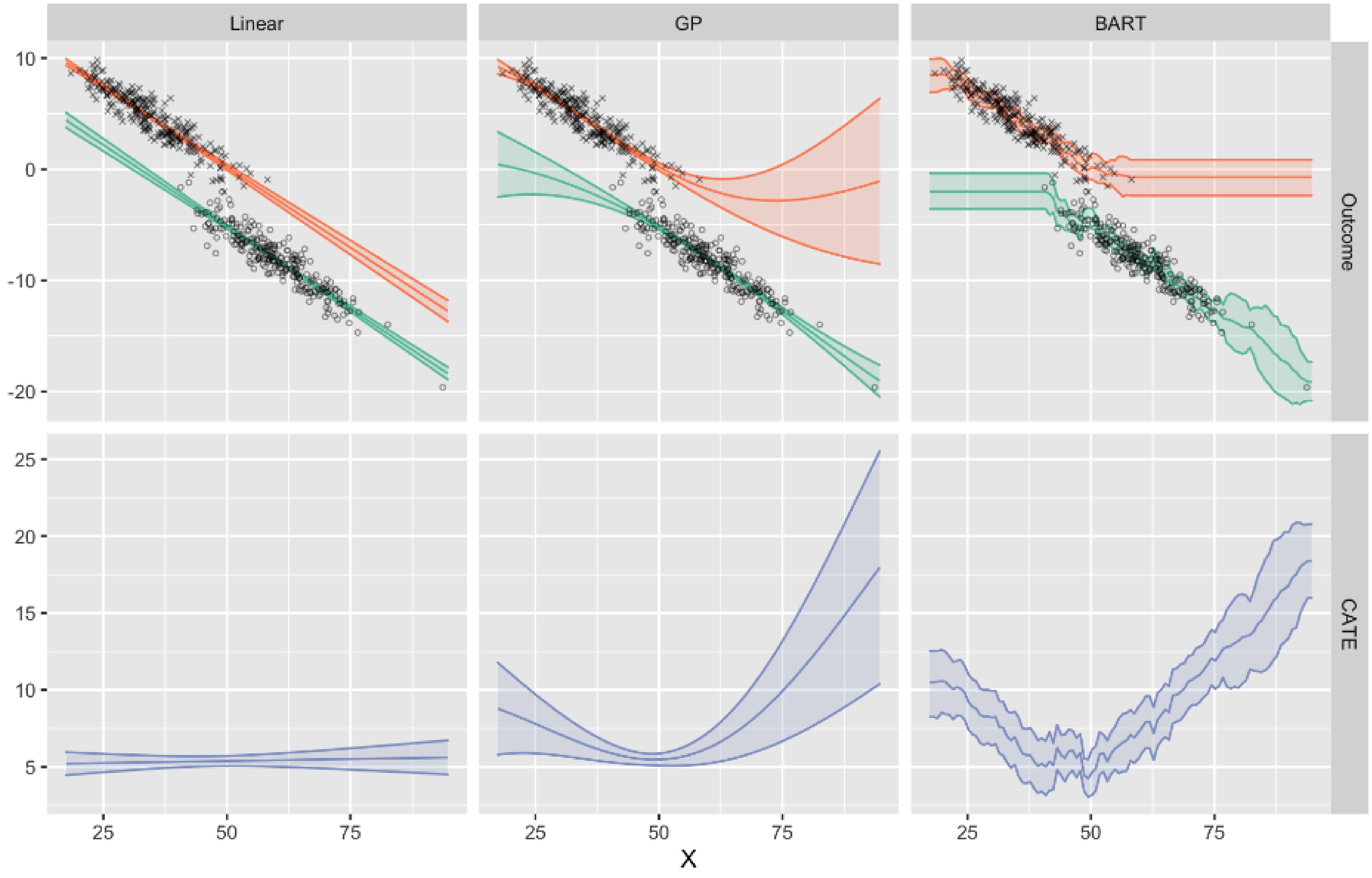
- A study with **250 treated (Z=1)** and **250 control (Z=0) units**.
- Each unit has a single covariate *X* **follow Gamma dist**.

$$X_{z_i=0} \sim \gamma(\frac{a}{b} = 60, \frac{a}{b^2} = 64)$$
$$X_{z_i=1} \sim \gamma(\frac{a}{b} = 35, \frac{a}{b^2} = 64)$$

- Set the true outcome model with constant treatment effects:

$$Y_i(z) = 10 + 5z - 0.3X_i + \epsilon_i, \text{ where } \epsilon_i \sim N(0,1)$$

Result



- (Top panel): Outcome Mean Value; (Lower panel): CATE
- (Left -> Right): Linear model, Gaussian Process, BART

Reference

1. Ref: Fan Li, Peng Ding, Fabrizia Mealli: Bayesian Causal Inference: A Critical Review

Thank you.