# Monte Carlo Simulation on Causal Forest

Jiacheng He

December 13, 2017

# Introduction

Sometimes we are interested to estimate the heterogeneous treatment effect of a binary treatment on different subgroups.

e.g.
Suppose we run a randomized trial to test a new drug. We spilt the subjects into a control group and a treatment group. What's the effect of this drug for adult white males and teenager Asian females respectively?

# Introduction

Wager and Athey (2017) developed causal forest method to predict heterogeneous treatment effect conditional on observed covariates.

In this project, I test the prediction accuracy and confidence interval coverage rate of causal forest via simulation.

# Causal Forest

Model setup

$Y_i$: The outcome variable

$W_i$: $W_i = 1$ if individual $i$ receives treatment, $W_i = 0$ if not treated

$X_i$: A vector of covariates

$\tau_i$: Individual treatment effect. Never observed.

$$Y_i = m(X_i) + \frac{W_i}{2}\tau(X_i) + \frac{1-W_i}{2}\tau(X_i) + \epsilon_i$$

$m(X_i) = E[Y_i|X_i]$: The conditional mean of outcome

$\tau(X_i) = E[Y_i|X_i, W_i = 1] - E[Y_i|X_i, W_i = 0]$: The heterogeneous treatment effect (conditional on covarites $X_i$)

$e(X_i) = E[W_i|X_i]$: The treatment propensity

# Causal Forest

Goal is to predict $\tau(X_i)$ (while random forest aims to predict $m(X_i)$)

Difficulty:
1. Disentangle $\tau(X_i)$ from $m(X_i)$ and $e(X_i)$
2. Cannot perform cross-validation, because we never observe the true $\tau_i$ (while in random forest we observe the true $Y_i$)

# Algorithm

Similar to random forest

Place a split at point $\tilde{x}_i$ which maximize the difference of $\hat{E}[Y_i|X_i = x_i, W_i = 1] - \hat{E}[Y_i|X_i = x_i, W_i = 0]$ across the two sides of $\tilde{x}_i$

(while random forest maximize the difference of $\hat{E}[Y_i|X_i = x_i]$)

# Simulation Setup

DGP1 (constant $\tau$)

$$\tau(X_i) = 0$$
$$e(X_i) = (1 + f_{beta}^{2,4}(X_{1i}))/4$$
$$m(X_i) = 2X_{1i} - 1$$

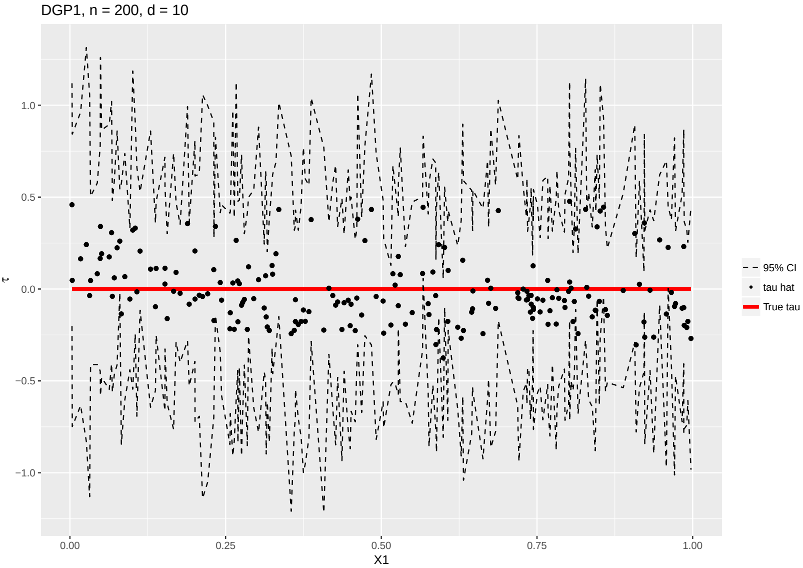
DGP2 (heterogeneous $\tau$)

$$\tau(X_i) = 1 + \frac{1}{(1+e^{-20(X_{1i}-1/3)})(1+e^{-20(X_{2i}-1/3)})}$$
$$e(X_i) = 0.5$$
$$m(X_i) = 0$$

# Simulation Setup

1. Draw $X_i \sim U(0,1)^d$, $W_i \sim binom(1, e(X_i))$, $\epsilon_i \sim N(0,1)$
2. Run the causal forest on a training set, then evaluate the trained model on a test set. ($n_{train} = n_{test}$)
3. For each senario, replicate it for 100 times. Then compute the average MSE and the 0.95 confidence interval coverage rate

# An Example



DGP1, n = 200, d = 10
1.0
0.5
0.0
−0.5
−1.0
τ
0.00
0.25
0.50
0.75
1.00
X1
95% CI
tau hat
True tau

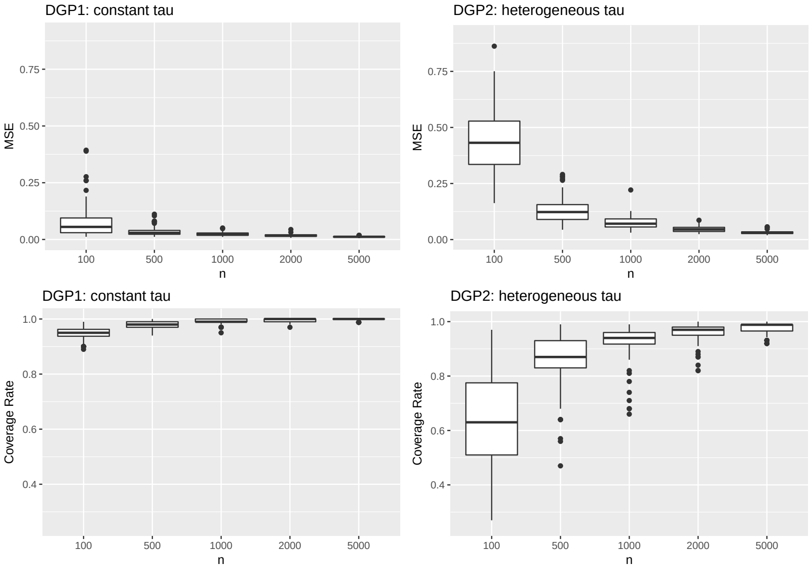
# Sample Size and Covariate Size

Vary the sample size $n$ and covariate size $d$ to see what happens:

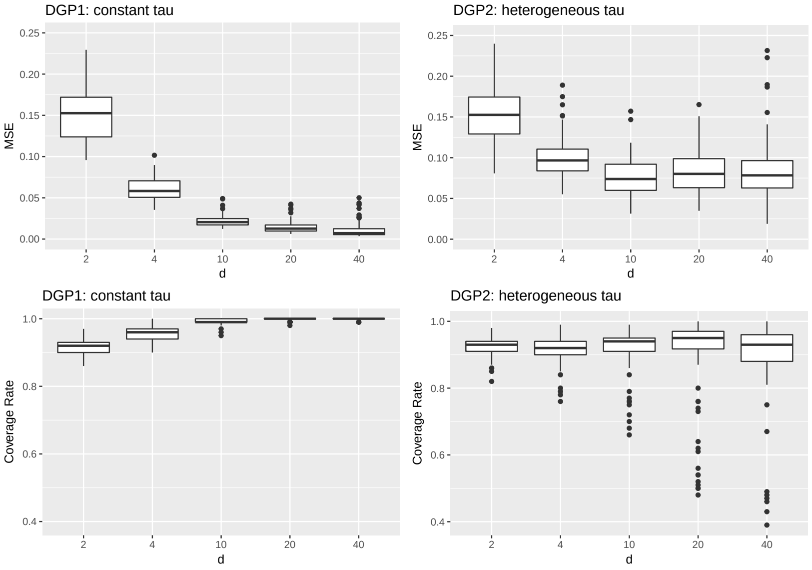
Fix $d = 10$, try $n = 100, 500, 1000, 2000, 5000$;

Fix $n = 1000$, try $d = 2, 4, 10, 20, 40$

# Sample Size and Covariate Size



# Sample Size and Covariate Size



# Tuning Parameters

I try varying five tuning parameters, one at a time. I use DGP2 and fix $n = 1000, d = 10$

1. Sample fraction used in each tree training; (default 0.5)
2. Covariates used in each tree training; (default $\frac{2}{3}d$)
3. Number of trees; (default 2000)
4. Minimun # observations in each terminal node; (default NULL)
5. Regularization parameter $\lambda$; (default 0)

# Tuning Parameters

1. Try sample fraction $s = 0.1, 0.2, 0.3, 0.4, 0.5$

Table 1: Different Sample Fraction s

| s | MSE | coverage |
|---|---|---|
| 0.1 | 0.242 | 0.535 |
| 0.2 | 0.123 | 0.87 |
| 0.3 | 0.082 | 0.92 |
| 0.4 | 0.073 | 0.94 |
| 0.5 | 0.075 | 0.94 |

# Tuning Parameters

2. Try # covariates in each tree training $t = 4, 5, 6, 7, 8$

Table 2: Different Number of Training Covariates t

| t | MSE | coverage |
|---|---|---|
| 4 | 0.102 | 0.865 |
| 5 | 0.082 | 0.92 |
| 6 | 0.079 | 0.92 |
| 7 | 0.07 | 0.94 |
| 8 | 0.074 | 0.94 |

# Tuning Parameters

3. Try # trees $b = 500, 1000, 2000, 4000, 6000$

Table 3: Different Number of Trees b

| b | MSE | coverage |
|---|---|---|
| 500 | 0.088 | 0.97 |
| 1000 | 0.079 | 0.96 |
| 2000 | 0.076 | 0.95 |
| 4000 | 0.077 | 0.94 |
| 6000 | 0.075 | 0.915 |

# Tuning Parameters

4. Try minimun node size = 0, 10, 20, 40, 80

Table 4: Different Minimum Node Size

| size | MSE | coverage |
| --- | --- | --- |
| 0 | 0.068 | 0.94 |
| 10 | 0.066 | 0.89 |
| 20 | 0.066 | 0.9 |
| 40 | 0.063 | 0.915 |
| 80 | 0.076 | 0.905 |

# Tuning Parameters

5. Try $\lambda = 0.1, 1, 5, 10, 100$

Table 5: Different Regularization Parameter lambda

| lambda | MSE | coverage |
|---|---|---|
| 0 | 0.067 | 0.95 |
| 0.1 | 0.069 | 0.94 |
| 1 | 0.066 | 0.94 |
| 5 | 0.076 | 0.925 |
| 10 | 0.079 | 0.92 |