

- 1 Causalidad y Machine Learning
- 2 Causalidad con BART
- 3 Árbol y Bosque causal
- 4 Bosque causal Bayesiano

## Causalidad y Machine Learning

- Marco: resultado potencial:  $\tau_i = Y_i(1) - Y_i(0)$
- Ignorabilidad:  $(Y(1), Y(0)) \perp D | X = x, 0 < Pr(D = 1|x) < 1$
- 2 grandes caminos: *ponderar* con  $e(x) = Pr(D = 1|x)$ , *modelar*  $E(Y|x, d) = f(x, d)$
- $ATE = E(Y_i(1) - Y_i(0))$

## Aporte de técnicas de ML

### Métodos de estimación más flexibles

- para estimar  $f(x) = E(Y|X = x)$
- para estimar  $e(x) = Pr(D = 1|X = x)$

### Relevancia:

- altas dimensiones
- efectos heterogéneos (CATE)
- generalización a población

### Distinguimos

- Métodos de ML usados para efectos causales
- Métodos diseñados para estimar causalidad

## Efecto de interés

Las técnicas que veremos se concentran en estimar CATE.

CATE: efecto promedio del tratamiento *condicional*

$$\tau_A = E(Y(1) - Y(0) | x \in A)$$

donde  $A$  es un valor o región de interés, es estimado en base a nuestro modelo como

$$\hat{\tau}_A = \sum_{x \in A} \hat{f}(x, 1) - \hat{f}(x, 0)$$

## Métodos ML para causalidad

Qingyuan Zhao and Trevor Hastie. [Causal interpretations of black-box models](#).  
*Journal of Business & Economic Statistics*, pages 1–10, 2019

3 ingredientes necesarios:

- Modelo *causal* descrito con DAG
- Modelo predictivo flexible
- Métodos de visualización (PPD)

en el contexto que usamos aquí:

bajo *ignorabilidad*, podemos estimar CATE con un modelo predictivo:

$$\hat{\tau}(x) = \hat{f}(x, D = 1) - \hat{f}(x, D = 0)$$

- los que vimos: CART, RF, BART
- otros: Boosting, Kernels, Redes neuronales, ...

## Métodos diseñados para causalidad

Si bien cualquier modelo predictivo se puede usar para estimar efectos causales, recientemente se proponen métodos diseñados para estimación causal:

- Causal Tree, Causal Forest
- Bayesian causal forest

1 Causalidad y Machine Learning

2 Causalidad con BART

3 Árbol y Bosque causal

4 Bosque causal Bayesiano

## BART para causalidad

- Jennifer L Hill. [Bayesian nonparametric modeling for causal inference.](#)  
*Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011
- Jennifer Hill and Yu-Sung Su. [Assessing lack of common support in causal inference using bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes.](#)  
*The Annals of Applied Statistics*, pages 1386–1420, 2013
- Holger L Kern, Elizabeth A Stuart, Jennifer Hill, and Donald P Green.  
[Assessing methods for generalizing experimental impact estimates to target populations.](#)  
*Journal of research on educational effectiveness*, 9(1):103–127, 2016



## BART para causalidad

$$\begin{aligned} Y &= f(X) + \varepsilon \\ &= \sum_j g_j(X | T_j, M_j) + \varepsilon \end{aligned}$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$\tau(x) = c(x, f) = \hat{f}(x, 1) - \hat{f}(x, 0)$$

Utilizando MCMC se obtiene  $s = 1, \dots, S$  simulaciones de la posterior de  $f(x_i)$  en cada punto de interés, y en base a esto construimos

$$\left\{ \tau^1(x_i), \dots, \tau^S(x_i) \right\}$$

## BART para causalidad

$$\tau_A^s = \sum_{x \in A} \tau^s(x_i)$$

es una simulación de la posterior para el efecto de interés.

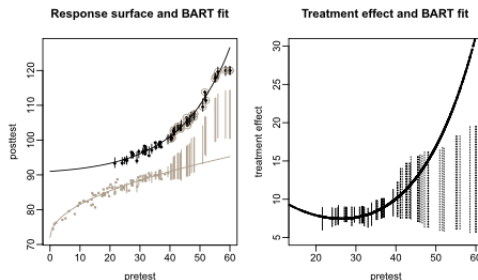
- 1 Estimación puntual:  $(1/S) \sum_s \tau_A^s$
- 2 Intervalo: percentiles 2.5% y 97.5%

Método flexible de modelado, con medidas de incertidumbre propias. Ha mostrado buena performance en competencias de estimación de efectos causales.

## Evaluar *overlap*

1390

J. HILL AND Y.-S. SU



descartar observaciones con mucha variabilidad:

$$\sigma_i^{1-d} > m_d + \text{sd}(\sigma_j^d)$$

$d$  es 0 o 1, y  $m_d$  el máximo desvío estimado.

1 Causalidad y Machine Learning

2 Causalidad con BART

3 **Árbol y Bosque causal**

4 Bosque causal Bayesiano

# Árboles causales

Susan Athey and Guido Imbens. **Recursive partitioning for heterogeneous causal effects.**

*Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016

Objetivo: Estimar efecto de tratamiento heterogeneo usando árboles de regresión.

## Árboles con doble muestra

Los árboles causales DM tiene dos principales diferencias con CART:

- Honesto. Dividir muestra en dos, usar una mitad para estimar particiones ( $S^{te}$ ) y la otra para estimar el efecto en los nodos terminales ( $S^{est}$ ).
- Si un dato  $x$  queda clasificado en la hoja  $l$ , el efecto del tratamiento se obtiene como:

$$\hat{\tau}_b(x) = \bar{Y}_{1l} - \bar{Y}_{0l}$$

Se asume que las hojas son suficientemente pequeñas tal que los pares  $(Y_i, D_i)$  para  $i \in L$  funcionan como que fueran de un experimento aleatorizado por eso es natural la estimación anterior.

## Árbol causal

Para particionar necesitan estimar MSE del efecto del tratamiento:

$$MSE_{\tau}(S^{te}, S^{est}, \pi) = \frac{1}{\#S^{te}} \sum_{i \in S^{te}} \{(\tau_i - \hat{\tau}(X_i; S^{est}, \pi))^2 - \tau^2\}$$

Se define a su vez  $EMSE(\pi) = E_{S^{te}, S^{est}}[MSE_{\tau}(S^{te}, S^{est}, \pi)]$

Tenemos que estimar  $EMSE$  porque no observamos  $\tau_i$ .

El criterio premia las particiones que encuentran gran heterogeneidad en el efecto de tratamiento.

## Árbol de propensión

Otra forma de obtener árboles honestos:

- Usar solamente  $D_i$  para las particiones y reservar  $Y_i$  para estimar  $\tau$
- Entreno el árbol usando como respuesta la asignación de tratamiento
- Cada hoja debe tener  $k$  o más observaciones de cada tratamiento.
- Estimo  $\tau$  como antes

Es honesto porque estoy usando solo  $Y_i$  para estimar el efecto y no para construir el árbol.



## Bosque causal

- Stefan Wager and Susan Athey. [Estimation and inference of heterogeneous treatment effects using random forests.](#)  
*Journal of the American Statistical Association*, 113(523):1228–1242, 2018
- Susan Athey, Julie Tibshirani, and Stefan Wager. [Generalized random forests.](#)  
*The Annals of Statistics*, 47(2):1148–1178, 2019

## Bosque causal

(Wagner, Athey 2017)

Proponen *bosques causal* (CF) construido de forma similar a RF. El bosque casual usa *árboles causales* como predictores individuales.

A partir de un conjunto de entrenamiento  $(Y, X)$

- Obtener  $B$  sub-muestras de tamaño  $s \ll n$
- Calcular  $\hat{\tau}_b(x)$  con un *árbol causal* (CT)
- Agregar para estimación del bosque  $\hat{\tau} = (1/B) \sum_b \hat{\tau}_b(x)$

2 diferencias claves con RF: sub-muestras en lugar de bootstrap, árbol causal en lugar de CART.

## Bosque causal: submuestras

### Usar sub-muestras

- Ventajas teóricas, permite desarrollar teoría asintótica para las predicciones del bosque
- Puede afectar performance en situaciones con muestras chicas o medias.

Permite tener resultados de normalidad asintótica para el estimador del efecto

$$(\tau(x) - \hat{\tau}(x)) / \sqrt{\text{Var}(\hat{\tau}(x))} \rightarrow N(0, 1)$$

con  $s \approx n^\beta$ .

- Normalidad asintótica basada en resultados de Meichhousen 2006
- Varianza puede estimarse como en Wagner 2014

- 1 Causalidad y Machine Learning
- 2 Causalidad con BART
- 3 Árbol y Bosque causal
- 4 Bosque causal Bayesiano**

## Bosque causal Bayesiano

(Hahn et al 2020) proponen, usar un modelo BART específico para el efecto causal de interés

$$Y = \mu(X) + \tau(W)D_i + \varepsilon_i$$

- $\mu(X)$ : BART con previas por defecto
- $\tau(W)$ : BART con previas que inducen arboles más chicos
- $W$  representan las variables predictoras que afectan  $\tau$ , puede ser  $W = X$
- En  $\mu(x)$  incluyen una estimación de  $Pr(D = 1|x)$  como covariable

- Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360, 2016.
- Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.
- Jennifer Hill and Yu-Sung Su. Assessing lack of common support in causal inference using bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *The Annals of Applied Statistics*, pages 1386–1420, 2013.
- Jennifer L Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.
- Holger L Kern, Elizabeth A Stuart, Jennifer Hill, and Donald P Green. Assessing methods for generalizing experimental impact estimates to target populations. *Journal of research on educational effectiveness*, 9(1):103–127, 2016.
- Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.
- Qingyuan Zhao and Trevor Hastie. Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, pages 1–10, 2019.