

Simulation Study

Setting + Assumptions

Suppose we have random sample (i.i.d.) of the reference population, each containing a feature vector X_i , a treatment indicator D_i and an outcome Y_i . In short: $\{(Y_i, X_i, D_i)\}_{i=1}^n$.

Generally, we'll either assume unconfoundedness (ignorability) or purely random treatment assignment in our DGPs. More formally, random treatment assignment amounts to

$$(Y(1), Y(0)) \perp\!\!\!\perp D. \quad (1)$$

In situations usually analyzed within the social sciences, this assumption turns out to be too restrictive (ref.). Therefore, as we've seen throughout the course, many methods were developed under the assumption of unconfoundedness:

$$(Y(1), Y(0)) \perp\!\!\!\perp D \mid X. \quad (2)$$

This is also used by (Athey and Wager 2018) to develop the theory of causal trees/forests.

Within the simulation study we'll also try to stress the consequences for estimation when switching from (1) to (2).

Heterogeneous Treatment Effects

To develop an intuition for the importance of heterogeneity in treatment effects, we define the Average Treatment Effect (ATE) and the Conditional Average Treatment Effect (CATE) as follows.

$$\delta \equiv \mathbf{E}(Y(1) - Y(0)) \quad (3)$$

$$\delta(x) \equiv \mathbf{E}(Y(1) - Y(0) \mid X = x) \quad (4)$$

These definitions are conceptually different: While the CATE is a real-valued function mapping a realization of the random variable X to a real number, the ATE is a real number. Thus, by definition, the CATE allows to have *distinct* treatment effects for different realizations of X . By the law of iterated expectation we have the following relationship:

$$\mathbf{E}(\delta(X)) = \delta$$

Thus, the ATE is a summary statistic of the CATE. Example for illustration + ref.

First Model

For the whole first part of the simulations, I'll stick to models that are linear in parameters when it comes to generating the potential outcomes. More specifically,

$$Y(0) = \gamma_0 + \gamma_1 \cdot g(X) + \gamma_2 \cdot h(X) + \varepsilon \quad (5)$$

$$Y(1) = \phi_0 + \phi_1 \cdot f(X) + \phi_2 \cdot c(X) + \varepsilon, \quad (6)$$

where

- $\varepsilon \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(0, 1)$, $X \stackrel{\text{i.i.d}}{\sim} U[0, 1]$
- $g(\cdot), h(\cdot), f(\cdot)$, and $c(\cdot)$ are real-valued functions
- $\phi_j, \gamma_j \in \mathbb{R}, j = 1, 2, 3$

In this model, ATE (3) and CATE (4) become

$$\delta = \phi_0 - \gamma_0 + \phi_1 \mathbf{E}(f(X)) - \gamma_1 \mathbf{E}(g(X)) + \phi_2 \mathbf{E}(c(X)) - \gamma_2 \mathbf{E}(h(X))$$

$$\delta(x) = \phi_0 - \gamma_0 + \phi_1 f(x) - \gamma_1 g(x) + \phi_2 c(x) - \gamma_2 h(x).$$

The workhorse function to generate the simulations from the first part looks as follows.

```
sim <- function(n, het_linear = FALSE, random_assignment = TRUE,
               non_linearY = FALSE, non_linearD = FALSE,
               gamma0=1, gamma1=3, gamma2=0, gamma3=1,
               phi0=5, phi1=3, phi1p=5, phi2=0, phi3=5){
  x <- runif(n)
  eps <- rnorm(n)
  #
  if(random_assignment){D <- rbinom(n,size=1,prob=0.5)}
  else{
    D <- rep(0,n)
    if(non_linearD){
      c_1<- 0.1; c_2 <- 0.89
      prt <- function(x){c_1+0.01*x+c_2*sin(pi*x)}
    }
    else{prt <- function(x){x}}
    for(j in seq_along(D)){D[j] <- rbinom(1,size=1,prob=prt(x[j]))}
  }
  if(non_linearY){
    nl <- function(z,const,t=1){const+3*z+20*sin(pi*z*t)}
```

```

y1 <- nl(z=x,const=phi0,t=phi3) + eps
y0 <- nl(z=x,const=gamma0,t=gamma3) + eps
}
else{
  if(het_linear){y1 <- phi0 + phi1p*x + eps}
  else{y1 <- phi0 + phi1*x + eps}
  y0 <- gamma0 + gamma1*x + eps
}
res <- tibble(Y0=y0[order(x)], Y1=y1[order(x)], X=sort(x),
              D=D[order(x)], Y_obs=rep(0,n), IntXD=x[order(x)]*D[order(x)])
res$Y_obs[res$D==1] <- res$Y1[res$D==1]
res$Y_obs[res$D==0] <- res$Y0[res$D==0]
return(res)
}

```

Constant treatment effects (no heterogeneity) + linearity

For the first simulation we'll study a straightforward setup to illustrate where CATE and ATE coincide. We'll use the following specification.

- $f : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \mapsto x$ and $f = g = c = h$
- $\phi_1 = \gamma_1$ and $\phi_2 = \gamma_2 = 0$
- $\phi_0 = 5, \gamma_0 = 1$

Thus we obtain

$$\begin{aligned} \delta &= \phi_0 - \gamma_0 + \mathbf{E}[(\phi_1 - \gamma_1) \cdot X] \\ &= \phi_0 - \gamma_0 \end{aligned}$$

$$\begin{aligned} \delta(x) &= \mathbf{E}(\phi_0 - \gamma_0 | X = x) \\ &= \phi_0 - \gamma_0 \\ &= 4 \end{aligned}$$

Packages

```

library(tidyverse)
library(np)

```

References

Athey, Susan, and Stefan Wager. 2018. “Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests.” *Journal of the American Statistical Association* 113 (523). Taylor & Francis: 1228–42.