

# Causal Inference - Homework 1

Introduction to Causal Inference -  
a Machine Learning Perspective 0365-4094-01  
Winter 2019/2020

**Submission data: 25 November, 2019**

## Question 1

Using potential outcomes notation, give an example of a data generating process (a joint distribution) which includes a hidden confounder  $H$ , a binary treatment  $T$ , and two potential outcomes  $Y_0$  and  $Y_1$ , such that:

1. Ignorability does *not* hold, and
2.  $\mathbb{E}[Y_1 - Y_0] \neq \mathbb{E}[Y|T = 1] - \mathbb{E}[Y|T = 0]$ ,

where  $Y = T \cdot Y_1 + (1 - T) \cdot Y_0$ .

## Question 2

Let  $(x_1, t_1, y_1), \dots, (x_n, t_n, y_n)$  be a sample from a randomized controlled trial, where for each  $i = 1, \dots, n$ ,  $x_i \in \mathbb{R}$  is a covariate measured before treatment assignment,  $t_i \in \{0, 1\}$  is a binary treatment, and  $y_i \in \mathbb{R}$  is an outcome measured after the treatment.

Let  $\pi : \mathbb{R} \rightarrow \{0, 1\}$  be a *policy*: a function which for each value of the covariate  $x$  assigns a treatment 0 or 1. Let  $V(\pi) = \mathbb{E}[y|t = \pi(x)]$  be the value of the policy  $\pi$ : what we expect the outcome to be if treatment were assigned according to  $\pi$ , as opposed to randomly. Give an unbiased estimator of  $V(\pi)$  for the sample above.

## Question 3

Give an example of a dataset with features  $X$  and one or more observed outcome variables  $Y_1, \dots, Y_k$ . For this dataset give:

1. Two examples of interesting causal questions relating one of the features and one of the outcomes. Explain what would be the treatment and what would be the *potential* outcomes in this case.
2. Two examples of interesting prediction questions which do not require causal reasoning.

Examples can come from the fields of politics, biology, sports, economics, entertainment, medicine, transportation and so on - use your imagination.