Counterfactuals

Kaizhao Liu

May 2023

Contents

This note aims to present the basic idea of counterfactuals in the language of probability theory. In this way, I hope to introduce new concepts and structures to enrich probability theory, and borrow tools and insights from probability theory to study causal inference. In the following I fix $(\Omega, \mathcal{F}, \mathbb{P})$ to be a probability space.

1 Binary Treatment

Let X, Y be random variables on Ω.

Example 1.1 (binary treatment). *You can imagine Ω as the collection of all people being investigated, each element $\omega \in \Omega$ stands for a single person being investigated. Suppose X is a binary treatment variable, where $X = 1$ means 'treated' and $X = 0$ means 'not treated'. Let Y be some outcome varible such as the absence of disease. The goal is to study the relationship between Y and X.*

Definition 1.1 (potential decomposition). If a random variable Y can be written as $Y = C_0 1_{A^c} + C_1 1_A$ where C_0 and C_1 are two random variables, then we say that Y admits a potential decomposition C_0, C_1 w.r.t. A.

Remark. *If $X = 1_A$ is the binary treatment varible associated with A, then we also say that Y admits a potential decomposition w.r.t X. In this case, we can call C_0, C_1 the potential outcomes with the following interpretation: C_0 is the outcome if not treated and C_1 is the outcome if treated.*

Theorem 1.1 (existence). *For any random variable Y on Ω and any event $A \in \mathcal{F}$, there exists random variable C_0 and C_1 s.t.*

$$Y = C_0 1_{A^c} + C_1 1_A.$$

This theorem is self-evident. We can look at the following cases, where we assume $X = 1_A$ is a binary treatment variable.

Example 1.2. *Let $C_0 = C_1 = Y$, then it is a potential decomposition of Y w.r.t. X. In this example, the outcome is the same whether treated or not. We can interpret this as X has no causal effect on Y.*

Example 1.3. *Let $C_0 = Y 1_D$ and $C_1 = Y 1_E$, where $A^c \subset D$, $A \subset E$ and $D, E \in \mathcal{F}$, then it is a potential decomposition of Y w.r.t. X. In this example, D, E can be chosen rather arbitarily. This shows that potential decomposition is not unique.*

The problem of the above example is that we can decompose a random variable *a posteriori*. To model the causal effect of the real world, we want the decomposition to be *a priori*. Namely, we are given a treatment X and potential outcomes C_0, C_1 first, then we construct Y naturally. We express this special type of potential decomposition more succinctly by

$$Y = C_X, \tag{1}$$

which is called the **consistency relationship**.

Now we can define statistics.

Definition 1.2 (average causal effect). Define the average causal effect or average treatment effect to be

$$\theta = \mathbb{E}C_1 - \mathbb{E}C_0. \tag{2}$$

2

Definition 2.1 (counterfactual function)**.** A random variable which is parameterized by X.

Definition 2.2 (causal regression function)**.** Define the causal regression function to be

$$\theta(x) = \mathbb{E}_\omega C(x, \omega). \tag{3}$$

Note that x is fixed.