Causal Inference in Machine Learning Lab

Daniel Jiwoong Im

February 10, 2024

Chapter 1

Lab3: Randomized controlled trials (RCT)

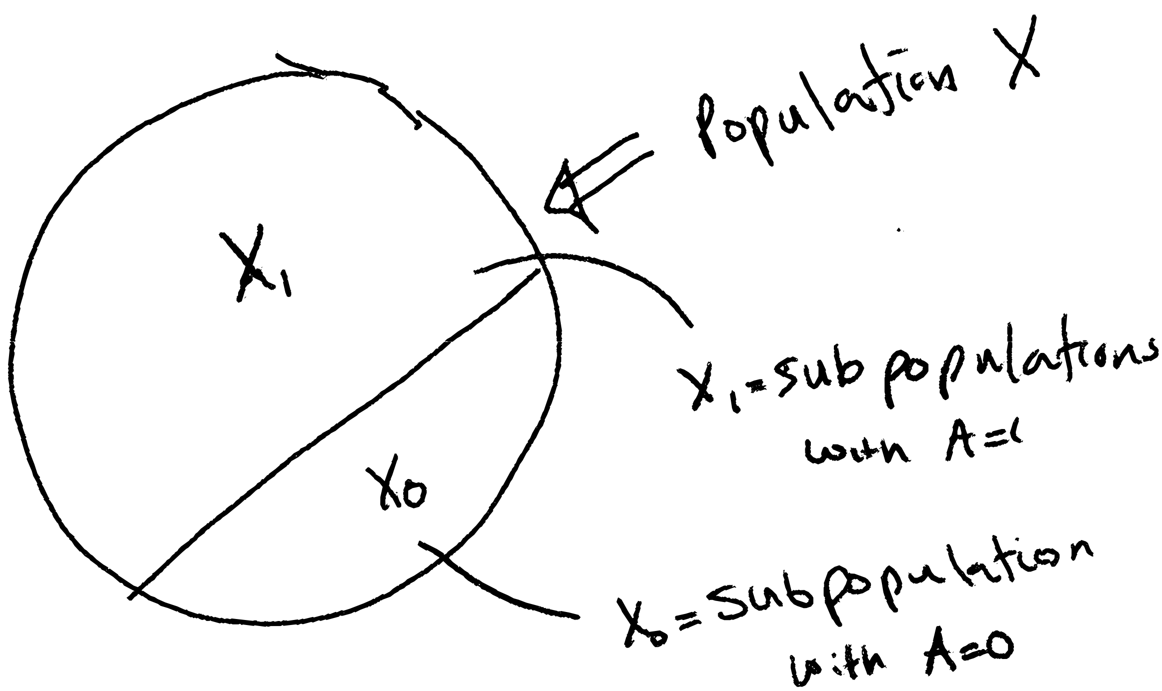
1.1 Recap & Notations

Assume Y is a Bernoulli random variable.

Conditional Probability

$$p_X(Y|A=1) = \mathbb{E}_{p(X)}[Y|A=1] = \frac{1}{p(A=1)} \sum_x p(Y|A=1,x)p(A=1|x)p(x)$$
$$p_X(Y|A=0) = \mathbb{E}_{p(X)}[Y|A=0] = \frac{1}{p(A=0)} \sum_x p(Y|A=1,x)p(A=0|x)p(x)$$

Let $X0$ and $X1$ be the subpopulations to which the action was applied.



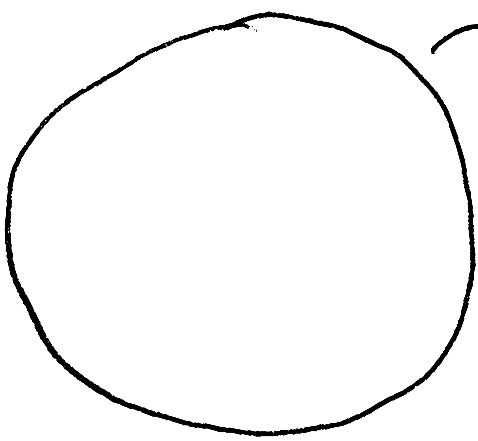
$P_X(Y|A=1) = P_{X_1}(Y|A=1)$

Interventional Probability (Potential Outcome)

$$p_X(Y|do(A=1)) = \mathbb{E}[Y[1]] = \mathbb{E}_{p(X)}[Y|do(A=1)] = \sum_x p(Y|A=1,x)p(x)$$

$$p_X(Y|do(A=0)) = \mathbb{E}[Y[0]] = \mathbb{E}_{p(X)}[Y|do(A=0)] = \sum_x p(Y|A=0,x)p(x)$$

Entire population X



$p_X(Y|do(A=a))$

↳ action a was applied to everyone, not subpopulation like conditional prob.

$\therefore\ p_X(Y|do(A=a)) \neq p_X(Y|A=a)$

Q when will they be equal?

Interventional Probability (Potential Outcome) applied to sub-population

$$\mathbb{E}_{p(X1)}[Y[1]] = p_{X1}(Y|do(A=1)) = \mathbb{E}_{p(X)}[Y|A=1]$$
$$\mathbb{E}_{p(X0)}[Y[1]] = p_{X0}(Y|do(A=1)) = \mathbb{E}_{p(X)}[Y|A=0]$$
$$\mathbb{E}_{p(X0)}[Y[1]] = p_{X0}(Y|do(A=1))$$
$$\mathbb{E}_{p(X1)}[Y[1]] = p_{X1}(Y|do(A=1))$$

where $X0$ and $X1$ are subpopulations to which the action was applied. Apply the intervention action A to the previous conditioned group $A^{'}$.

Average Treatment Effect

The ATE measures the difference in mean outcomes between units assigned to the treatment and units assigned to the control - wiki.

$$ATE := \mathbb{E}[Y[1]] - \mathbb{E}[Y[0]]$$

Q1: Why do we want action assignments to be randomized over the population $\mathcal{X}$? When would the dataset be truly randomized?

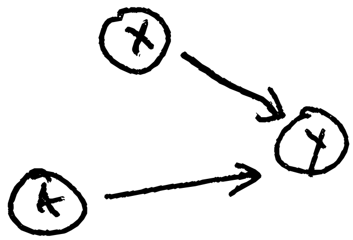
Randomization means that you randomly assign actions to the population.

$$p(y \mid do(A=a)) = p(y \mid A=a)$$

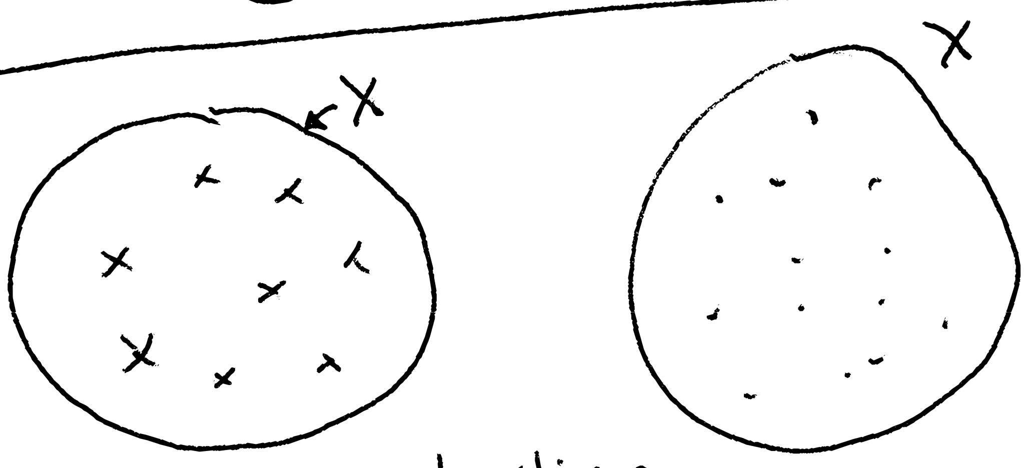
when action a are randomly assigned.

$\iff$ $p(a) = p(a \mid x)$

$\iff$ $a \perp x$ (independency)

$\iff$ 

(no edge between X, A)



The avg. outcome of action a applied to x samples vs. ∵ samples should result more or less the same.

i.e. randomly sampled subpopulations are exchangeable & gives the same result!

1.2 Treatment Effect $\equiv$ A/B Testing

- The objective is the same, that is we want to know the causal effect
- Methodology: random assignment
- mathematically the same

There is a difference in procedure though (experimentalist cares about these kinds of details). RCT takes a "passive approach" to collect data, that is define groups and then assign action to each group. A/B testing takes an "active approach" to collect data, that is randomly assign data as a stream of data comes in. Next lab, we will look into "Active CI learning"!