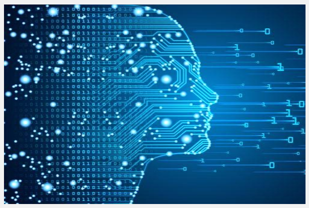


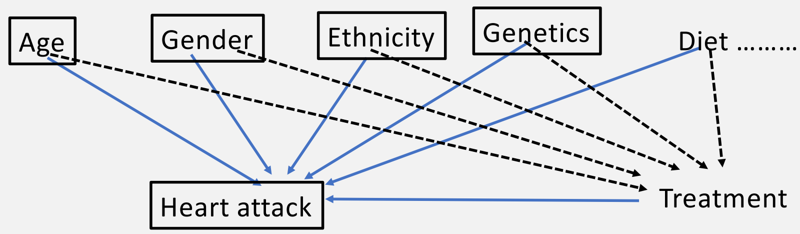
Principles of Causal Inference

Vasant G. Honavar

Dorothy Foehr Huck and J. Lloyd Huck Chair in Biomedical Data Sciences and Artificial Intelligence
Professor of Data Sciences, Informatics, Computer Science and Engineering, Bioinformatics & Genomics, Public Health Sciences and Neuroscience
Center for Artificial Intelligence Foundations and Scientific Applications
Institute for Computational and Data Sciences
Huck Institutes of the Life Sciences
Clinical and Translational Sciences Institute
Northeast Big Data Hub
Pennsylvania State University
vhonavar@psu.edu
http://faculty.ist.psu.edu/vhonavar
http://ailab.ist.psu.edu

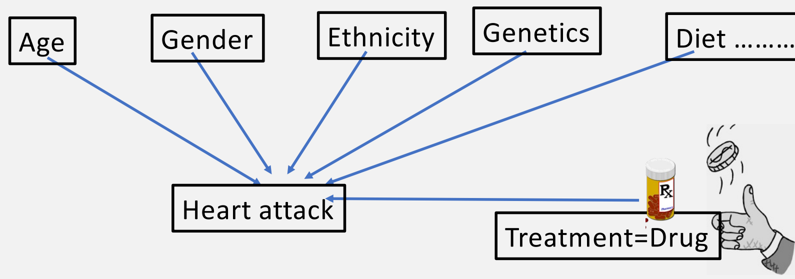
Randomized control trial (RCT)

Confounding



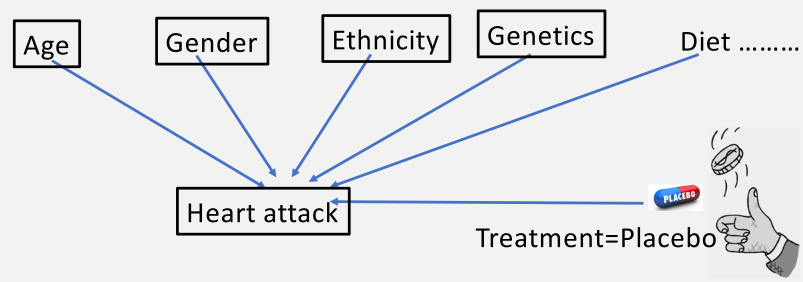
Randomization breaks the possible causal links between the potential confounders and the treatment

Why RCTs work



Randomization breaks the possible causal links between the potential confounders and the treatment

Why RCTs work



Randomization breaks the possible causal links between the potential confounders and the treatment

Why not always do RCT?

Randomized controlled trials are not always feasible

- RCT may be too costly
- RCT may not be ethical
- …

Causal effects defined

What is a causal effect?

- Individual causal effects
- Average causal effects
- Measures of causal effect
- Random variability
- Causation versus association

We introduce simple mathematical notation that formalizes the causal intuition that you already have

Causal Question

- What is the effect of the treatment on the outcome?
- If the opposite treatment had been received, how would the outcome differ?
- Example:
  - Treatment: choosing organic produce
  - Outcome: get cancer? (yes or no)
  - Question: Does choosing organic produce decrease risk of cancer?

Treatment and Potential outcomes

- For simplicity, we will consider only two possible treatments:
  - Treatment of interest ("treatment")
  - Control ("control") – often just not getting the treatment
- The two possible treatments (treatment, control) must be well defined
- We refer to the outcomes under treatment and control as "potential outcomes"
  - $Y^{a=1}$ and $Y^{a=0}$ where a is a binary treatment or simply $Y(1)$ and $Y(0)$
  - Remember that $Y^{a=1}$ is the RV that denotes the interventional outcome of treatment $a = 1$
  - The definitions and methods can be generalized to the setting of multiple treatments or even continuous valued treatments, e.g., dosage of a drug

Key Assumptions

- Non interference (units do not interfere with each other):
  - Treatment applied to one unit does not influence the outcome for another unit
  - Whether John underwent surgery does not influence the result of Jane's surgery
- Can you think of a situation where the "no interference" requirement is violated?
  - Suppose John is Jane's caregiver. John being incapacitated due to surgery will affect Jane's post-surgical care
- Suppose there are n individuals indexed by i, and a_i the treatment accorded to individual i
  - Non interference implies $Y_i(a_1, a_2, \ldots a_i \ldots a_n) = Y_i(a_i)$
  - That is, treatments given to all other individuals do not impact the outcome of treatment for individual i

Key Assumptions

Can you think of a situation where the "no interference" requirement is violated?

- Treatment – getting a raise
- Joe and James receive raises
- James is jealous of Joe
- James' happiness at receiving raise is adversely impacted by Joe receiving a raise as well
- Suppose there are n individuals indexed by i, and a_i the treatment accorded to individual i
  - Non interference implies $Y_i(a_1, a_2, \dots a_i \dots a_n) = Y_i(a_i)$
  - That is, treatments given to all other individuals do not impact the outcome of treatment for individual i

Key Assumptions

- Treatment must be well-defined
  - There is only a single version of each treatment
    - Taking Lipitor 20mg is a different treatment compared to taking Lipitor 40mg
  - Potential outcomes must be well defined

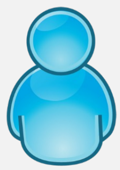
Individual causal effects

- Binary treatment variables: A_i(1: treated; 0: untreated)
- Binary outcome variable: Y_i (1: death; 0: survival) $Y_i^{A_i=v}$
- Outcomes under treatment $A_i = v, v \in \{0, 1\}$
- $Y_i^{A_i=1}$ and $Y_i^{A_i=0}$ are now random variables (why?)
- We sometimes drop the index i when the context does not require us to distinguish between individuals
- We sometimes write $a = v$ (dropping index i) and abusing notation
- We use $Y_i(1)$ and $Y_i(0)$ interchangeably with $Y_i^{A_i=1}$ and $Y_i^{A_i=0}$ respectively

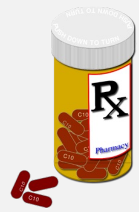
Individual causal effect

- Suppose James has $Y^{a=1} = 1$ and $Y^{a=0} = 0$ because he died when treated and somehow magically we know that he would have survived if he was untreated
- Suppose Jane has $Y^{a=1} = 0$ and $Y^{a=0} = 0$ because she survived when treated and somehow magically we know that she would have survived even if she was untreated
- Now for James, the treatment has a causal effect on outcome but for Jane, it does not
- The outcomes under different treatments $a = 0$ and $a = 1$ are called potential outcomes
- If an individual receives treatment $a = 0, Y^{a=0}$ is called the factual outcome and $Y^{a=1}$ the counterfactual outcome
- More precisely, individual causal effect $\tau_i = Y_i(1) - Y_i(0)$

Potential Outcomes framework

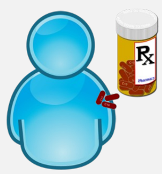


Zeus



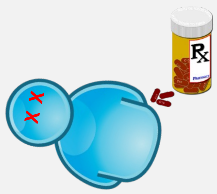
Treatment

Potential Outcomes framework



Zeus

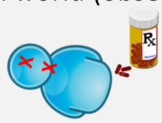
Potential Outcomes framework



Zeus

Potential Outcomes framework

Factual world (observed) Counterfactual world (imagined)



$Y_i^{a=1}$

$Y_i^{a=0}$

Causal effect of treatment $= Y_i^{a=1} - Y_i^{a=0}$

- In general, individual causal effects can, strictly speaking, never be identified, that is, expressed as a function of observed data
- Why? Because for any individual, only the factual outcome is observed, and the counterfactual outcome is not

Non-interference

- Non-interference: The definition of factual and counterfactual outcomes implicitly assumes that an individual's treatment outcome under treatment a does not depend on the treatment received by the other individuals

Individual causal effect

- To define an individual causal effect (ICE or ITE), we needed
  - An outcome of interest Y
  - Treatments (actions) $a = 0$ and $a = 1$
  - An individual for whom the factual and counterfactual outcomes $Y_i(1)$ and $Y_i(0)$ are to be compared

    $\tau_i = Y_i(1) - Y_i(0)$

Average causal effect

- To define an average causal effect (ACE or ATE), we need
  - An outcome of interest Y
  - Actions $a = 0$ and $a = 1$
  - A well-defined population of individuals for whom the factual and counterfactual outcomes are to be compared

$$\tau \quad = \mathbb{E}\,[Y_i(1) \; - \; Y_i(0)] \approx \frac{1}{n}\sum_{i=1}^{n} \tau_i$$

Average causal effect

- When the average causal effect of treatment on the outcome in the population is zero, we say that the causal null hypothesis of no average causal effect holds
- The non-null average causal effect exists when Average causal effect of treatment A on the outcome Y in the population $\tau \neq 0$
- Absence of average causal effect does not necessarily imply absence of individual causal effect
- When no individual has an individual causal effect, we say that the sharp causal null hypothesis holds
- Sharp causal null hypothesis implies causal null hypothesis, but not the other way around

Average causal effect

	$Y^{a=0}$	$Y^{a=1}$		$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1	Leto	0	1
Kronos	1	0	Ares	1	1
Demeter	0	0	Athena	1	1
Hades	0	0	Hephaestus	0	1
Hestia	0	0	Aphrodite	0	1
Poseidon	1	0	Cyclope	0	1
Hera	0	0	Persephone	1	1
Zeus	0	1	Hermes	1	0
Artemis	1	1	Hebe	1	0
Apollo	1	0	Dionysus	1	0

- Average causal effect of treatment A on the outcome Y in the population exists if $\Pr[Y^{a=1}=1]-\Pr[Y^{a=0}=1]\neq 0$
- For Zeus' extended family, $\Pr[Y^{a=1}=1]=\frac{10}{20}=\Pr[Y^{a=0}=1]$

Average causal effect

	$Y^{a=0}$	$Y^{a=1}$		$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1	Leto	0	1
Kronos	1	0	Ares	1	1
Demeter	0	0	Athena	1	1
Hades	0	0	Hephaestus	0	1
Hestia	0	0	Aphrodite	0	1
Poseidon	1	0	Cyclope	0	1
Hera	0	0	Persephone	1	1
Zeus	0	1	Hermes	1	0
Artemis	1	1	Hebe	1	0
Apollo	1	0	Dionysus	1	0

- 12 individuals have individual causal effects, of whom 6 were helped by the treatment and 6 were harmed by the treatment (causal sharp null hypothesis does not hold)
- Average causal effect is zero (causal null hypothesis holds)

Measures of (size of) average causal effect

- Causal risk difference or size of causal effect

$$\tau = \Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 0]$$

- Causal risk ratio or causal effect ratio $\frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]}$
- Causal odds ratio $\frac{\left(\frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=0]}\right)}{\left(\frac{\Pr[Y^{a=0}=1]}{\Pr[Y^{a=0}=1]}\right)}$

Random variability

- We typically have only a sample from the population of interest, and not the entire population
- Then the probabilities we used will need to be replaced by their sample estimates instead
- Different samples will yield slightly different estimates
- Error in the estimate due to sampling variability is random and obeys the law of large numbers
- We assume that the probability estimates are consistent – the error in the estimate approaches 0 as the number of samples approaches infinity
- Determining whether there is an average causal effect now becomes a hypothesis testing problem (rejecting the causal null hypothesis)

Stochastic potential outcomes

- We have assumed that the outcomes of treatment on each individual is deterministic
  - Zeus had a 100% chance of dying if treated and 0% chance of dying if untreated
- What if the outcomes are stochastic?
  - What if Zeus had a 90% chance of dying if treated and 10% chance of dying if untreated?
  - We expect the distribution of stochastic outcomes to vary across individuals
- Two mechanism behind stochastic outcomes
  - Effect of exogenous factors e.g., weather, food, etc.
  - Inherent nondeterminism as in quantum physics
- We will deal with random variability and stochastic outcomes using statistical techniques

Association

- In the real-world, we only observe the factual outcomes, and do not by definition, observe the counterfactual outcome
- All we have are the observed treatment A and observed outcome Y

	A	Y		A	Y		A	Y
Rheia	0	0	Zeus	1	1	Aphrodite	1	1
Kronos	0	1	Artemis	0	1	Cyclope	1	1
Demeter	0	0	Apollo	0	1	Persephone	1	1
Hades	0	0	Leto	0	0	Hermes	1	0
Hestia	1	0	Ares	1	1	Hebe	1	0
Poseidon	1	0	Athena	1	1	Dionysus	1	0
Hera	1	0	Hephaestus	1	1			

- We can obtain from data, the proportion of individuals who developed outcome Y among those who received treatment value a
- Note that observational data yield observational probabilities

Association

	A	Y		A	Y		A	Y
Rheia	0	0	Zeus	1	1	Aphrodite	1	1
Kronos	0	1	Artemis	0	1	Cyclope	1	1
Demeter	0	0	Apollo	0	1	Persephone	1	1
Hades	0	0	Leto	0	0	Hermes	1	0
Hestia	1	0	Ares	1	1	Hebe	1	0
Poseidon	1	0	Athena	1	1	Dionysus	1	0
Hera	1	0	Hephaestus	1	1			

- 7 individuals died (Y=1) among the 13 that were treated (A=1)
- $\Pr(Y=1|A=1)=7/13$; Similarly, $\Pr(Y=1|A=0)=3/7$
- When $\Pr(Y=1|A=1)=\Pr(Y=1|A=0)$, we say that A and Y are independent
- When $\Pr(Y=1|A=1)\neq\Pr(Y=1|A=0)$, we say that A and Y are associated or dependent

Measures of association

- Associational risk difference

$$\Pr[Y \quad = 1|A = 1] - \Pr[Y \quad = 1|A = 0]$$

- Associational risk ratio $\frac{\Pr[Y \quad =1|A=1]}{\Pr[Y \quad =1|A=0]}$
- Associational odds ratio $\frac{\left(\frac{\Pr[Y \quad =1|A=1]}{\Pr[Y \quad =0|A=1]}\right)}{\left(\frac{\Pr[Y \quad =1|A=0]}{\Pr[Y \quad =0|A=0]}\right)}$
- Contrast these with their causal counterparts e.g.,

$$\tau = \Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 0]$$

Exercise: Compute the average causal effect.

Potential Outcomes

$Y^{a=0}$	$Y^{a=1}$
13	14
6	0
4	1
5	2
6	3
6	1
8	10
8	9

Exercise: Compute the average causal effect.

Potential Outcomes

$Y^{a=0}$	$Y^{a=1}$
13	14
6	0
4	1
5	2
6	3
6	1
8	10
8	9

$$E(Y^{a=1} - Y^{a=0}) = \frac{(1 - 6 - 3 - 3 - 3 - 5 + 2 + 1)}{8} = -2$$

Exercise: Compute the association

- This is the same table as before, except
- Only factual outcomes are available
- Counterfactual outcomes are missing (denoted by ?)

A	$Y \mid A=0$	$Y \mid A=1$
1	?	14
0	6	?
0	4	?
0	5	?
0	6	?
0	6	?
1	?	10
1	?	9

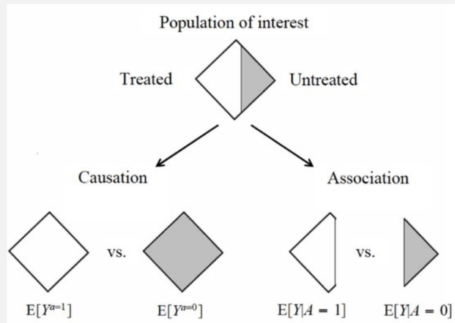
Exercise: Compute the association

A	$Y\|A=0$	$Y\|A=1$
1	?	14
0	6	?
0	4	?
0	5	?
0	6	?
0	6	?
1	?	10
1	?	9

$$E(Y|A=1) - E(Y|A=0)$$
$$= \left(\frac{14+10+9}{3}\right) - \left(\frac{6+4+5+6+6}{5}\right)$$
$$= 11 - 5.4 = 5.6$$

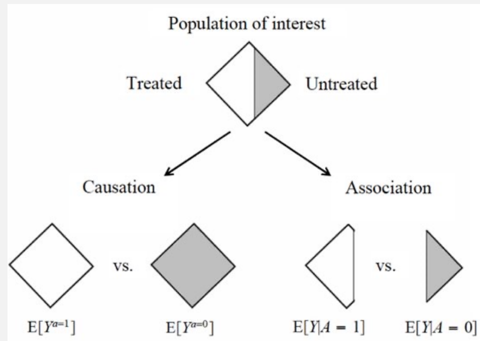
Causation versus association

- Causal inference is concerned with what if questions (imagined world)
  - What would the risk be if everyone in the population was treated (or untreated)?
- Statistical inference is concerned with observed world
  - What is the risk in the treated (or untreated) subpopulation?



Causation versus association

- Clearly, association is not causation
- The world in which we have defined causal effects is an imagined world
- Under what conditions can we identify causal effects in the real world, i.e., from observational data?
- Under what conditions is association is causation?

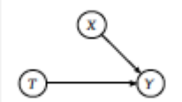


In general, causation is not association

- In order to distinguish causation from association, we need
  - Knowledge of the assignment mechanism

    OR
  - Additional assumptions

Ignorability



- Ignorability says that we can ignore the treatment mechanism
- Potential outcomes are independent of treatment
- This is tantamount to saying that the individuals were assigned to treatment and control groups at random
- If ignorability holds, association is causation

$(Y(1), Y(0)) \coprod T$ (where T is treatment)

$$\tau = \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$
$$= \mathbb{E}[Y(1)|T = 1] - \mathbb{E}[Y(0)|T = 0]$$
$$= \mathbb{E}[Y|T = 1] - \mathbb{E}[Y|T = 0]$$

Independence between counterfactual outcome and the observed treatment does not imply independence between the observed outcome and observed treatment!

Exchangeability

- Exchangability means that the treated and control groups are identical except for their treatment status
- If treatment and control groups were swapped, the new treatment group will experience the same treatment effect as the old treatment group, and the new control group would experience the same effect as the old control group

$$\mathbb{E}[Y(1)|T=1] = \mathbb{E}[Y(1)|T=0]$$
$$\mathbb{E}[Y(0)|T=1] = \mathbb{E}[Y(0)|T=0]$$

Which implies

$$\mathbb{E}[Y(1)|T=t] = \mathbb{E}[Y(1)]$$
$$\mathbb{E}[Y(0)|T=t] = \mathbb{E}[Y(0)]$$

Which is the “mean” version of ignorability

Causation is association if exchangeability holds

- If exchangeability holds, we can calculate ACE from the observational distribution $P(X, T, Y)$
- A causal quantity, e.g., $\mathbb{E}[Y\ \ (\mathrm{t})]$ is identifiable if we can calculate it from purely statistical quantity $\mathbb{E}[Y\ \ |t]$

Conditional Exchangeability

- In observational data, it is unrealistic to expect that treatment groups are exchangeable
  - Typically it is those who are sick that receive treatment and not those who are healthy
  - Treated and control groups are not exchangeable
- But what if we condition on (control for) all the relevant variables except treatment?

$$(Y(1), Y(0)) \coprod T \mid X$$

- While treatment and control groups may not have been comparable without conditioning on X, they are comparable upon conditioning on X (the covariates)
- The formal proof of this assertion has to wait until we introduce causal graphical models and do-calculus

Positivity

- For any value of covariates, the probability of receiving treatment is non-zero

$$\forall x \; 0 < P(T = 1|X = x) < 1$$

- Why do we need positivity?

$$\tau = \mathbb{E}\,[Y(1) - Y(0)]$$

$$= \mathbb{E}_X\left[\mathbb{E}[Y(1)|T = 1, X] - \mathbb{E}[Y(0)|T = 0, X]\right]$$

$$= \sum_x P(X = x)\left(\sum_y y\, P(Y = y|T = 1, X = x) - \sum_y y\, P(Y = y|T = 0, X = x)\right)$$

$$= \sum_x P(X = x)\left(\sum_y y \frac{P(Y = y, T = 1, X = x)}{P(T = 1|X = x)P(X = x)} - \sum_y y \frac{P(Y = y, T = 0, X = x)}{P(T = 0|X = x)P(X = x)}\right)$$

Without positivity, we will be conditioning on an event with 0 probability!

Positivity

- Violation of positivity means for some value of covariates, everyone is in the treatment group and no one in the control group!
  - We cannot compare treatment and control groups!
  - Causal effect is undefined!
- Conditioning on confounders is necessary to get rid of confounding
- Conditioning can in the extreme lead to violation of positivity
  - There is a tradeoff between unconfoundedness and positivity

Consistency

- If the treatment is T, the observed outcome is the same as the potential outcome under treatment T
- Consistency requires that $Y(t) = Y$ for every individual with $T = t$
  - The observed outcome for every treated individual equals her outcome had she been treated
  - The observed outcome for every untreated individual equals her outcome had she been left untreated
- Do not be fooled by the deceptive simplicity of consistency!
- Consistency relies on
  - Precise definition of treatments and potential outcomes $Y(t)$
  - Linkage of potential outcome to observed outcomes

Unpacking Consistency

- Precise definition of potential outcome $Y(t)$
  - There is only one version of each possible treatment
    - E.g., if treatment involves taking aspirin, you need to specify the exact composition, amount, and mode of delivery
    - This can be tricky in observational studies
    - Suppose $t =$ "obese at age 40". What does this really mean? Does it matter if the individual was obese all his life, as opposed to gradually becoming obese after turning 30?
    - Same goes for "not obese at age 40". Does it matter if absence of obesity is genetic? Due to diet? Due to microbiota?
    - If treatment values a are not well defined, the potential outcomes $Y(t)$ are not well defined
    - If $Y(t)$ are ill-defined, causal effect of treatment on outcome is ill defined

Unpacking Consistency

- Precise definition of potential outcome $Y(t)$
  - How can we be sure that a treatment is sufficiently well-defined?
  - We don't. Declaring that a treatment is sufficiently well-defined is a matter of agreement among experts based on available substantive knowledge
    - Today, we believe that the direction you face while lifting weights is irrelevant to the causal effect of weightlifting on your fitness
    - Ten years later, accumulating evidence might suggest that facing North while lifting weights is harmful, say, if your weights are made of magnetic material
  - The vagueness of causal questions can be reduced – but not entirely eliminated – by being precise about treatment!

Unpacking Consistency

- Linkage of potential outcome to observed outcomes
  - Suppose the treatments $T = 1$ and $T = 0$ and hence potential outcomes $Y(1)$ and $Y(0)$ are sufficiently well-defined
  - What does $Y(t) = Y$ for individuals with $T = t$ really imply?
  - Suppose we are interested in the effect of obesity on heart disease
  - Suppose Joe was not obese at 40 (due to genetics, exercise, diet)
    - This implies his observed treatment value $T \neq 1$, and hence his observed outcome Y may not equal the potential outcome $Y(1)$ under the imagined intervention $T = 1$
  - To link the potential outcomes to observed outcomes, we need data from some individuals where $Y(t) = Y$ for all treatments
    - This may be impossible when the data are simply not rich enough

Unpacking Consistency

- Do not be fooled by the deceptive simplicity of consistency!
- Consistency relies on
  - Precise definition of potential outcome $Y(t)$
  - Linkage of potential outcome to observed outcomes
- We should do our best to ensure consistency
- Because perfection is impossible, we should be as transparent as possible about the details so our causal claims can be challenged and if warranted, falsified
- There are alternate points of view – we don't need to precisely specify potential outcomes and link them to observed outcomes as long as T temporally precedes Y (Pearl, 2009)

Randomized experiments

- The gold standard for causal inference
- Suppose we want to determine whether the color (blue, black) of the ink used by a student to answer an exam has a causal effect on grade (pass versus fail)
- Can you think of a suitable experiment?
  - You stand by the door
  - As a student is about to enter the room, you flip a fair coin
    - If the outcome is a head, you give the student a black pen
    - If the outcome is a tail, you give the student a blue pen
  - After the instructor grades the exams, you compare the proportion of students receiving the passing grade among those receiving black pens with those receiving blue pens
  - If the two proportions are same, then the color of the pen has no causal effect on grade

Assumptions needed for causal inference from observational data

- Conditional exchangeability (absence of confounding)
- Positivity
- Non-interference
- Consistency

$\left.\begin{array}{l}\text{Non-interference}\\\text{Consistency}\end{array}\right\}$ SUTVA

Estimating causal effects

- To estimate causal effects
  - We need to ensure that exchangeability holds
  - We want treatment groups that are similar regarding covariates
  - We do so by creating covariate balance across treatment groups
  - Easiest way to accomplish this: randomized experiments
  - When randomized experiments are not possible, we need a bag of tricks

Randomized control trial (RCT)

Randomized Experiment

- The assignment mechanism is random, known, and controlled by the researcher
- Because the treatments are randomly assigned, the treatment groups should all look similar regarding covariates (observed and unobserved)
- In such "classical" randomized experiments, the assignment mechanism is unconfounded by design
- If possible, randomized experiments should be double-blind
    - the experimental subjects should be oblivious to the specific treatment given to them
    - the researchers should be ignorant of which treatment is being given to each subject

Analysis of RCT under the exchangeability assumption

Person	*T*	*Y (1)*	*Y (0)*
1	1 (Black)	1	?
2	0 (Blue)	?	1
3	1 (Black)	0	?
4	0 (Blue)	?	0
5	1 (Black)	1	?
6	0 (Blue)	?	0

- Exchangeability means that the treated and untreated groups are similar with respect to the covariates

- Assignment to Blue and Black groups is randomized
- The proportion of "Pass", i.e., outcome 1, among the Black group is expected to be identical to those in the Blue group had it been the case that the Blue group were treated (received Black pens) instead of the Black group
- The treated and untreated groups are exchangeable

Analysis of RCT under the exchangeability assumption

Person	*T*	*Y (1)*	*Y (0)*
1	1 (Black)	1	?
2	0 (Blue)	?	1
3	1 (Black)	0	?
4	0 (Blue)	?	0
5	1 (Black)	1	?
6	0 (Blue)	?	0

$\Pr[Y(1) = 1] = \Pr(Y = 1|T = 1)$

$\Pr[Y(1) = 0] = \Pr(Y = 0|T = 1)$

$\Pr[Y(0) = 1] = \Pr(Y = 1|T = 0)$

$\Pr[Y(0) = 0] = \Pr(Y = 0|T = 0)$

- When the treated and untreated groups are exchangeable, the unknown counterfactual probabilities are the same as observational probabilities
- In this case, causation is association!

Causal effect of treatment $= \Pr[Y(1) = 1] - \Pr[Y(0) = 1]$
$= \Pr(Y = 1|T = 1) - \Pr(Y = 1|T = 0) = (2/3) - (1/3) = 1/3$

Randomized control trial

- Randomization is expected to ensure exchangeability
- Hence, in the case of RCT, causation is association!
- Counterfactual outcome $Y(t)$ is statistically independent of the observed treatment T, i.e., $Y(t) \perp\!\!\!\perp T$
- Independence between counterfactual outcome and the observed treatment does not imply independence between the observed outcome and observed treatment
- When the treatment has a causal effect on the outcome, Y and T are associated and hence not independent!
- Experiments where exchangeability holds are called randomized (or more generally, marginally randomized) experiments

Propensity Score

- The propensity score $e(x)$ at $X = x$ is the average unit assignment probability for units with covariates $X = x$
- Assuming unconfounded assignment, the propensity score is just the probability of units with $X = x$ getting treated

Types of Randomized Experiments

- Bernoulli randomized experiment
- Completely randomized experiment
- Stratified randomized experiment
- Paired randomized experiment

• Increasingly restrictive treatment assignments

Bernoulli Experiment

- In a Bernoulli experiment, the treatment for each unit is determined by a coin flip
  - $T_i = 1$ if unit i received treatment
  - $T_i = 0$ if unit i did not receive treatment
- Observed outcome $Y_i = Y_i(t_i) = \begin{cases} Y_i(1) \text{ if } T_i = 1 \\ Y_i(0) \text{ if } T_i = 0 \end{cases}$

  That is, $Y_i = T_i \, Y_i(1) + (1 - T_i) \, Y_i(0)$
- Usually, the the probability of assigning a unit with $X = x$ to treatment group, i.e., its propensity score $e(x) = ½$
- The treatment and control groups are exchangeable
- The treatment assignments of subjects are independent
- But… there is a small probability that in any run of the experiment, all units are assigned to the treatment group or control group

Bernoulli Experiment

- In a Bernoulli experiment, the treatment for each unit is determined by a coin flip
  - $T_i = 1$ if unit i received treatment
  - $T_i = 0$ if unit i did not receive treatment
- Observed outcome $Y_i = Y_i(t_i) = \begin{cases} Y_i(1) \text{ if } T_i = 1 \\ Y_i(0) \text{ if } T_i = 0 \end{cases}$

  That is, $Y_i = T_i\, Y_i(1) + (1 - T_i)\, Y_i(0)$
- Usually, the the probability of assigning a unit with $X = x$ to treatment group, i.e., its propensity score $e(x) = ½$
- What if $e(x) = p$ where $0 < p < 1$?
- Are the treatment and control groups exchangeable?
- Are the treatment and control groups covariate balanced?

Question: How would you generalize the Bernoulli design to the setting with $K > 2$ treatments?

- Instead of a coin toss, use a K-sided die toss for assigning subjects to treatments
- Suppose one of the treatments is a reference treatment or control (e.g., placebo)
- Compute the causal effect of each of the other treatments relative to the control
- Order the treatments according to the magnitude of their causal effects (relative to the control)

Completely Randomized Experiment

- In a completely randomized experiment with N subjects sample sizes for each treatment group are fixed in advance
- N_T = size of treatment group
- $N_C = N - N_T$ = size of control group
- Often $N_T = N_C = N/2$, but not always
- $e(x) = \frac{N_T}{N}$
- We sample N_T subjects out of N without replacement
- Are the treatment assignments of subjects independent?
  - No! Assignment of a subject to the treatment group reduces the probability of another subject being assigned to the treatment group!
- Ensures exchangeability

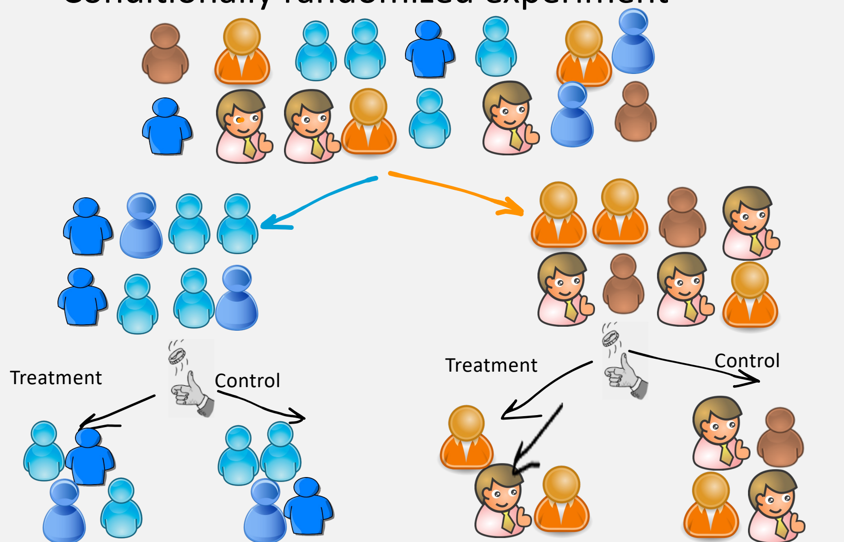
Stratified or conditionally randomized experiment

- In a stratified randomized experiment, population is partitioned into blocks or strata within which individuals are similar with respect to one or more covariates
  - Strata may correspond to individuals with similar education, demographics, etc.
- Individuals are completely randomized within each block/strata
- Ensures balance for important covariate(s)
- Also called blocking
- Heuristic: Block what you can, randomize what you cannot – may not always work (blocking can introduce confounding in some cases – more on this later)

Stratified or conditionally randomized experiments

- The probability of an individual being treated depends on, say, result of some test that is indicative of criticality (critical versus non-critical)
- Does exchangeability hold?
- No, because the treated and untreated groups may be unbalanced in terms of prognosis
- But we can ensure exchangeability within each of the “strata” using conditionally randomized experiments
  - Partition the subjects into strata based on the value of covariate(s) L e.g., critical versus non critical
  - Completely randomize the treatment assignment within each of the strata
- Estimate causal effect within each stratum
- Take weighted average across strata

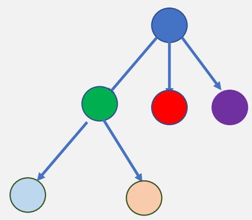
Conditionally randomized experiment



Stratified or conditionally randomized experiments

- The probability of an individual being treated depends on, say, result of some test that is indicative of criticality (critical versus non-critical)
- Exchangeability does not hold across the population, but holds within each stratum
  - We estimate causal effect within each stratum
  - Causal effects may be heterogeneous across strata
  - Average causal effect across population may be computed if desired by taking a weighted average of causal effects across strata (the weights are proportional to the size of the strata)

Stratification can be nested

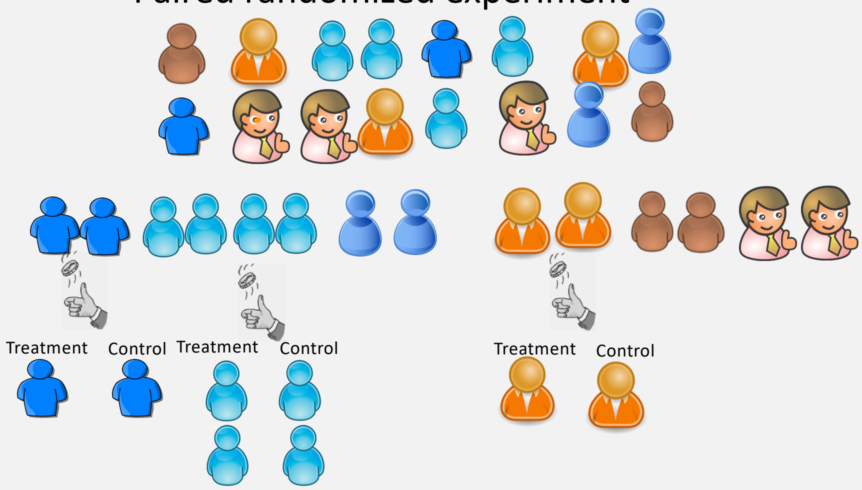


- Exchangeability holds within each stratum
  - Causal effects may be heterogeneous across strata
  - Average causal effect across population may be computed if desired by taking a recursive weighted average of causal effects up the tree defining the nested stratification

Paired randomized experiments

- In a paired randomized experiment, individuals are first matched into pairs that are similar with respect to covariates
- Within each pair, randomize which individual is treated
- Special case of stratification
- Also called matched pairs experiments

Paired randomized experiment



Clustered randomized experiment

- Suppose we want to study effect of an educational intervention
- Randomizing each individual student is not possible
- Causal effect of the intervention on a student may be affected by the treatment assignments of other students who she interacts with leading to violation of non-interference
- Clustered randomization is useful in such settings
- Cluster individuals so as to minimize inter-cluster interference
- Randomly assign clusters to treatments, keeping in mind, the need for exchangeability (at the cluster level)
  - Stratified randomization
  - Matched randomization

Summary of Randomized Experiments

- Randomly assigning individuals to treatments
  - Creates balanced treatment groups
  - Eliminates confounding of treatment and outcome by confounders

- Four types of classical randomized experiments:
  - Bernoulli randomized experiment
  - Completely randomized experiment
  - Stratified or conditionally randomized experiment
  - Paired randomized experiment

- Clustered randomization when there is interference

Summary of Randomized Experiments

Type of Experiment and Design	Number of Possible Assignments Cardinality of $\mathbb{W}^+$	Number of Units (N) in Sample			
		4	8	16	32
Bernoulli trial	2^N	16	256	65,536	4.2×10^9
Completely randomized experiment	$\binom{N}{N/2}$	6	70	12,870	0.6×10^9
Stratified randomized experiment	$\binom{N/2}{N/4}^2$	4	36	4,900	0.2×10^9
Paired randomized experiment	$2^{N/2}$	4	16	256	65,536

Summary of Randomized Experiments

- Completely randomized design is more informative than Bernoulli design
- Because it eliminates a priori uninformative treatment assignments, e.g., those with almost all units assigned a single treatment
- The stratified design is superior to completely randomized design when the information used to specify the strata is predictive of the potential outcomes
  - In the best case, the level of the pre-treatment covariate defining a stratum perfectly predicts both potential outcomes for the stratum
  - In the worst case, the strata correspond to random partitioning of units, and membership in a stratum is not predictive of potential outcomes for the stratum
- Similar arguments apply to the paired randomized design