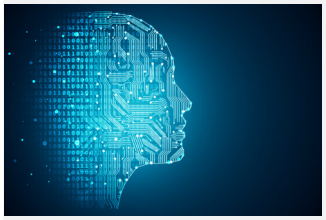


Data Science for Researchers and Scholars

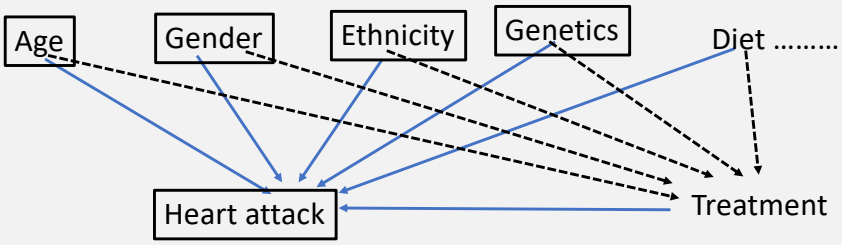
Vasant G. Honavar
Dorothy Foehr Huck and J. Lloyd Huck Chair in Biomedical Data Sciences and Artificial Intelligence
Professor of Data Sciences, Informatics, Computer Science and Engineering, Bioinformatics & Genomics, Public Health Sciences and Neuroscience
Director, Center for Artificial Intelligence Foundations and Scientific Applications
Associate Director, Institute for Computational and Data Sciences
Pennsylvania State University

vhonavar@psu.edu
http://faculty.ist.psu.edu/vhonavar
http://ailab.ist.psu.edu

Randomized experiments

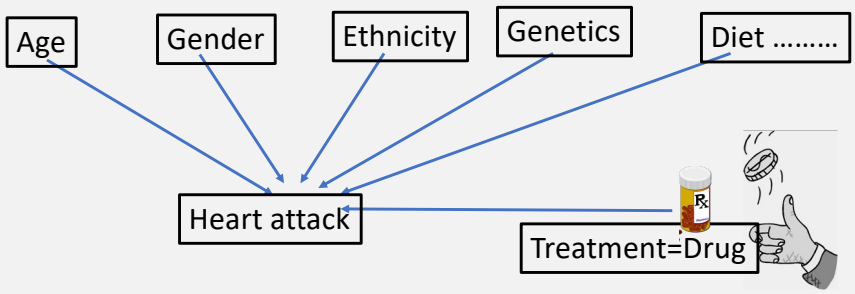
- The gold standard for causal inference
- Suppose we want to determine whether the color (blue, black) of the ink used by a student to answer an exam has a causal effect on grade (pass versus fail)
- Can you think of a suitable experiment?
  - You stand by the door
  - As a student is about to enter 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

Why RCT works despite 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
- Randomization ensures that the treated and untreated populations are exchangeable
- Exchangeability means that treated and untreated groups are statistically identical except for their treatment status

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

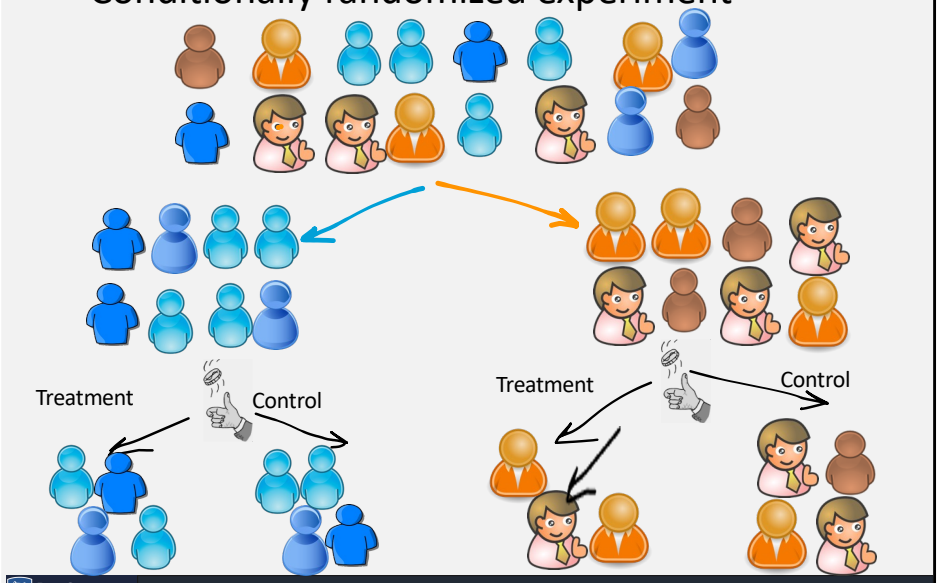
Question: How would you generalize the Bernoulli design to the setting with $K > 2$ treatments?

- 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 $N_T = NC = N/2$, but not always
- $e(x) = \frac{N_T}{N}$
- We sample N_T subjects out of N without replacement
- Ensures exchangeability

Conditionally randomized experiment



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)

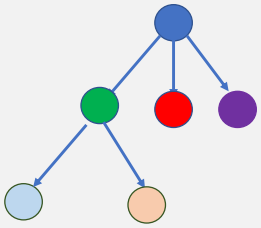
Stratified or conditionally randomized experiments

- 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

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

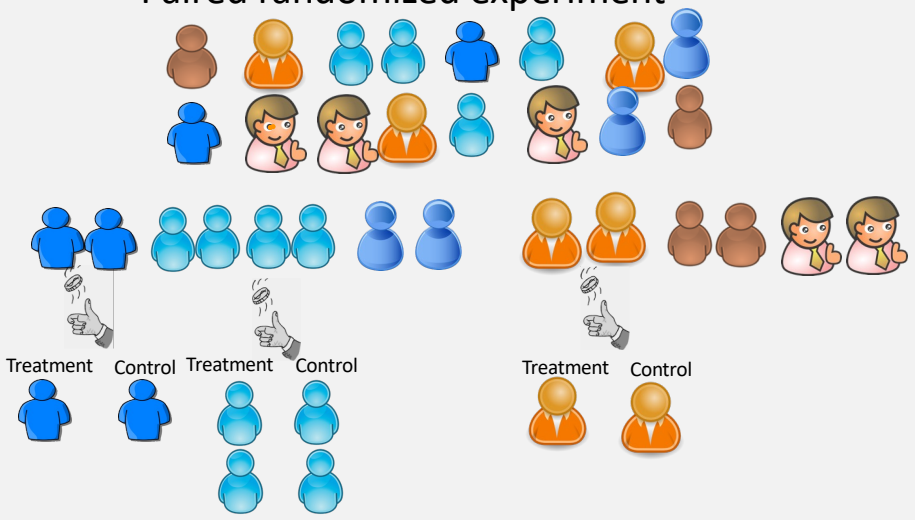


- Exchangeability holds within each stratum
  - Causal effects may be heterogeneous across strata
  - Average causal effect across population may be computed if desired by taking a recursive weighted average of causal effects up the tree defining the nested stratification

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



Summary of Randomized Experiments

- Randomly assigning individuals to treatments
  - Creates exchangeable 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

Estimating causal effects

Causal estimation

In the finite sample setting, the average causal effect of treatment is defined as:

$$\tau = \overline{Y^{T=1}} - \overline{Y^{T=0}} = \frac{\sum_{i=1}^{N} Y_i^{T=1}}{N} - \frac{\sum_{i=1}^{N} Y_i^{T=0}}{N}$$

Estimand: Average causal effect

- In the finite sample setting, the average causal effect of treatment is defined as:

$$\tau = \overline{Y^{T=1}} - \overline{Y^{T=0}} = \frac{\sum_{i=1}^{N_{T=1}} Y_i^{T=1}}{N_{T=1}} - \frac{\sum_{i=1}^{N_{T=0}} Y_i^{T=0}}{N_{T=0}}$$

where $T_i = 1$ if the ith individual is treated and 0 otherwise

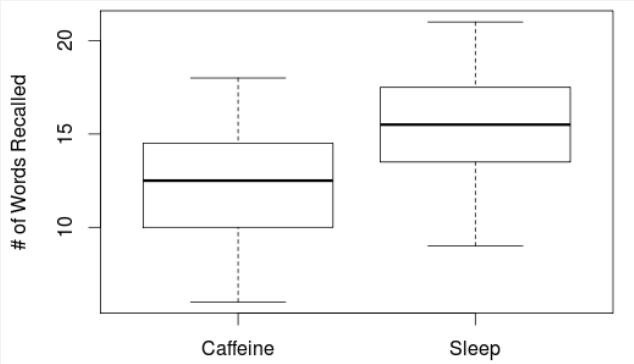
Note: As usual, we can calculate confidence intervals etc.

Example: Sleep or Caffeine?

- Is sleep or caffeine better for memory?
- 24 adults were given a list of words to memorize, then randomly divided into two groups and sent over to take a break
- During the break one group took a nap for an hour and a half, while the other group stayed awake and then took a caffeine pill after an hour
- Y: number of words recalled

Mednick et al., "Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory", Behavioral Brain Research, 2008; 193: 79-86.

Sleep or Caffeine



- Suppose the requisite assumptions (exchangeability etc.) hold
- Can we determine whether sleep or caffeine lead to better recall?

Sleep versus Caffeine

- Estimand:
  - The average word recall for all 24 people if they had slept – average word recall for all 24 people if they had caffeine
- Note that the estimator assumes exchangeability of the treated and untreated populations
- Estimate varies from one random assignment to another
- Estimator is unbiased if $E(\hat{\tau}) = \tau$

Sleep versus Caffeine

- Estimator is unbiased if $E(\hat{\tau}) = \tau$
- For completely randomized experiments,

$$\hat{\tau} = \frac{\sum_{i=1}^{N} T_i Y_i^{T=1}}{N_{T=1}} - \frac{\sum_{i=1}^{N} (1 - T_i) Y_i^{T=0}}{N_{T=0}}$$

is an unbiased estimator of

$$\tau = \overline{Y^{T=1}} - \overline{Y^{T=0}}$$

if the treated and untreated populations are exchangeable

Statistics

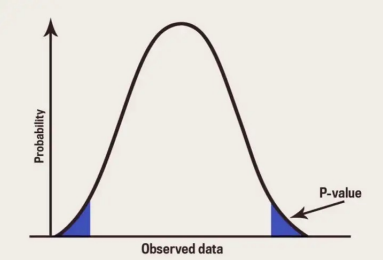
- We can compute the confidence interval around the estimate
- We can also formulate the null hypothesis (no causal effect) and perform hypothesis testing

How do we reject the null hypothesis?

- Compare at the behavior of the observed test statistic (in our case, causal effect) under random assignment of treatment to the test statistic under null hypothesis
  - Calculate a test statistic from the data (assuming random assignment of units to treatment groups)
  - Based on this statistic, with *some probability* we can reject the null hypothesis, that is, show that it does not hold
  - Calculate the 2-sided p value

p value and α value

- p *value* The probability of observing a test statistic at least as large as the one observed, by random chance, assuming that the null hypothesis is true.
- α *value* The p-value threshold at which you're okay with rejecting the null hypothesis (typically 0.01 or 0.05)



- 1-sided p-value offer evidence against the null hypothesis
- 2-sided p-value is used to reject the possibility that the observed effect is due to chance
- The smaller the p-value, the greater the confidence with which you can reject the null hypothesis

Why not always do a randomized experiment?

Randomized experiments are not always feasible

- They may be too costly
- They may not be ethical
  - Example: Making ½ the study population to smoke when we believe smoking is injurious to health

Causal inference from observational data

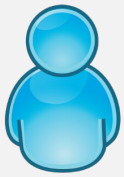
- An observational study can be viewed as a conditionally randomized experiment if the following conditions hold:
  - Treatments correspond to well-defined interventions that can be imagined in the data
  - The conditional probability of receiving every possible treatment, though not decided by the investigators, depends only on the measured covariates
  - The probability of receiving every treatment conditional on the observed covariates is greater than 0
- These conditions, taken together, are called identifiability assumptions
- We know how to draw valid causal inference from conditionally randomized experiments
- If we assume that the above identifiability conditions hold, we can draw valid causal inferences from observational data

Causal inference from Observational Data

Other possible approach to causal inference:

- A predictor of treatment, referred to as an instrumental variable, was randomly assigned conditional on the measured covariates"
- What we should do:
  - Carefully specify
    - The randomized experiment that we would like to, but cannot, conduct
    - How the observational study emulates that randomized experiment
- In ideal randomized experiments, the data contain sufficient information to identify causal effects
- In contrast, without identifiability assumptions, the information in observational data is insufficient to identify causal effects

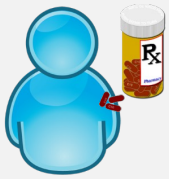
Potential Outcomes framework



Alice

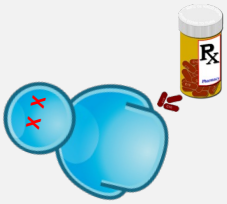
Treatment

Potential Outcomes framework



Alice

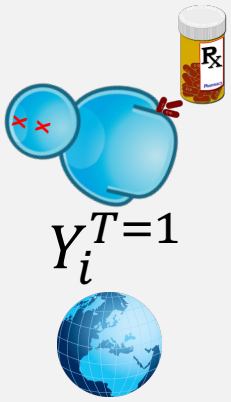
Potential Outcomes framework



Alice

Potential Outcomes framework

Factual world (observed) Counterfactual world (imagined)



$Y_i^{T=1}$

$Y_i^{T=0}$

Causal effect of treatment = $E[Y^{T=1} - Y^{T=0}]$

Causal inference requires the counterfactual outcome

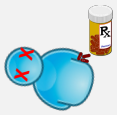
Person	T	$Y^{T=1}$	$Y^{T=0}$	Covariates
1	1	0.4	0.3	x_1
2	0	0.8	0.6	x_2
3	1	0.3	0.2	x_3
4	0	0.3	0.1	x_4
5	1	0.5	0.5	x_5
6	0	0.6	0.5	x_6
7	0	0.3	0.1	x_7

Causal effect of treatment = $E[Y^{T=1} - Y^{T=0}]$

Problem: counterfactual outcome is not observed!

- Missing data imputation problem
- Estimate missing data using various methods
  - Imputation from similar individuals

$\hat{Y}^{T=0}$



$Y^{T=1}$

- $Y_{T=0}$ now becomes an estimated quantity, based on outcomes of other individuals that are most similar to the treated individual that did not receive the treatment

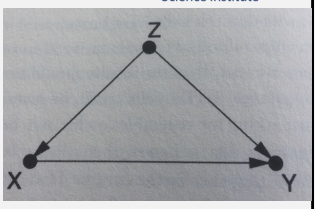
Estimating counterfactual outcomes

- Estimating counterfactual outcome based on the factual outcome of similar individuals (ideally identical twins in the opposite treatment group)
- Latent space methods
  - Learn a low dimensional latent representation from the covariates of treated and untreated individuals
  - Find similar individuals in the opposite treatment group based on distance in the latent space
- Machine learning methods
  - Predict the counterfactual outcome by training a model on a training set with factual outcomes
  - A zoo of recent methods

Assumptions underlying the potential outcomes framework

- Consistency – the potential outcome under $T=a$ is the same as the actual outcome under $T=a$
- Stable unit treatment value assumption (SUTVA) – Joe's response to drug does not depend on whether Mary was treated with the drug
  - Generally holds although one can imagine cases where it might not
- Unconfoundedness of treatment mechanism – Individuals, given their characteristics, are assigned treatment without regard to how they would respond to treatment (i.e., their potential outcomes)
  - Trivially holds under randomization
  - Otherwise unverifiable from observational data

Confounding revisited



- Confounding bias arises whenever a variable influences both who is selected for treatment and the outcome of the experiment
  - Sometimes the confounders are known
  - Sometimes the confounders are suspected
- The most basic version of confounding
  - The true causal effect X → Y is mixed with the spurious correlation induced by the fork X ←Z→ Y
  - Example: We are testing a drug but give it to patients who are younger, but not to those who are older – age becomes a confounder

How can we cope with confounders?

- Randomized controlled trials
  - – not always feasible, costly
- Potential outcomes framework (with matching)
  - – tantamount to identifying randomized experiments hidden in observational data
- Adjusting for confounders
  - – if *Z* is the only confounder and we have measured *Z*, we can compare the treatment and control groups for each possible value of *Z* and take a weighted average where the weights correspond to the fraction of the population represented by each value of *Z*
  - – need to know what the confounders are
  - – need to be able to measure the confounders

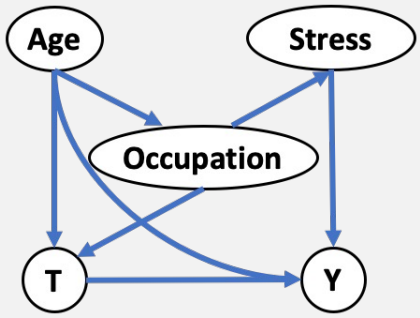
Confounding is a fact of life

- We can adjust for confounders if
  - We can identify them
  - They are observed (recorded in observational data)
- Standard statistical methodology provides little guidance for what variables to adjust for
  - You can end up controlling for the very thing you are trying to measure
  - You may fail to control for a confounder that you should control for
  - Even if you get lucky and control for the right confounders you have no way of knowing that and hence may have to avoid making causal claims even if they are justified

Confounding is a fact of life

- We can adjust for them if
  - We can identify them
  - They are observed (recorded in observational data)
- Most definitions of confounding, e.g., those used in the epidemiology and social sciences literature, are flawed
  - False positives as well as false negatives
  - Is it any wonder then that most scientific findings are false?
- Correct definition using the language of causal calculus
  - Confounder is any factor that leads makes $P(Y|X) \neq P(Y|do(X))$
- Checking this condition requires causal assumptions in the form of a causal graph!

Causal graph encodes causal assumptions

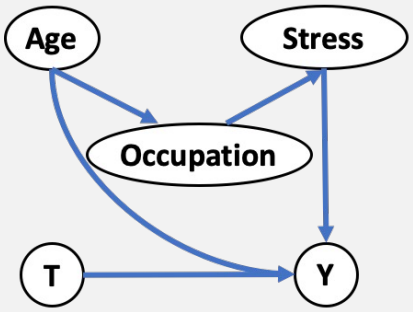


$$Occupation = h(Age, u_o)$$
$$Stress = k(Occupation, u_s)$$

$$T = g(Age, Occupation, u_t)$$
$$Y = f(T, Age, Stress, u_y)$$

48

Causal effect is implicit in the interventional distribution



Counterfactual (Intervention) world:

All edges to Treatment T removed, *keeping everything else the same*.

Observational distr. = P(Y|T)

Interventional distr.= P*(Y|T)

Causal Effect: $E^*[Y|T=1] - E^*[Y|T=0]$

Pearl's "do" notation

- We distinguish random X from X fixed by intervention by the notation "$do(X)$".
- Average causal effect of X on Y

$$P(Y = 1 \mid do(X = 1)) - P(Y = 1 \mid do(X = 0))$$

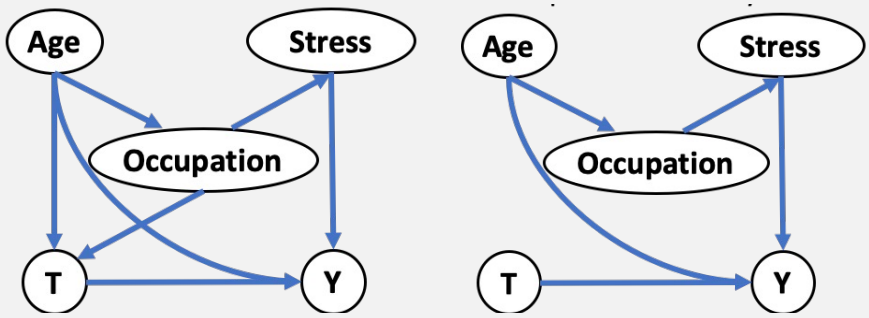
- In the vocabulary of statistics, this is the estimand. We may derive it from a model, and estimate it with an estimator.

Do not conflate the estimand, model, and estimator

- Conflation between estimand, model, and estimator is a major source of confusion in debates about causal inference
- Our primary focus is on specifying estimands and models
- We can leverage existing work on estimation (even deep learning)
- To do causal inference from observational data
  - We postulate a causal graph
  - Express the estimand as a function of the postulated causal graph
  - Check whether this function can be calculated from observational data (using the machinery of *do*-calculus)

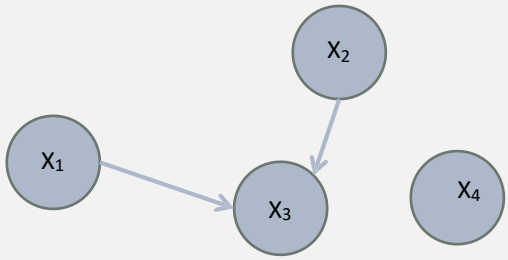
Causal graph permits reasoning about interventions

- $do(X)$ denotes intervention on X
- Intervention on X has the effect of removing all incoming links into X (or eliminating all direct causes of X)



What is a Direct Cause?

- The direct causes of X_i are the variables which will change the distribution of X_i as we vary them while holding everything else unchanged



$$P(X_3 = x_3 | do(X_1 = x_1), do(X_2 = x_2), do(X_4 = x_4)) \neq P(X_3 = x_3 | do(X_1 = x_1'), do(X_2 = x_2), do(X_4 = x_4))$$

$$P(X_3 = x_3 | do(X_1 = x_1), do(X_2 = x_2), do(X_4 = x_4)) = P(X_3 = x_3 | do(X_1 = x_1), do(X_2 = x_2), do(X_4 = x_4'))$$

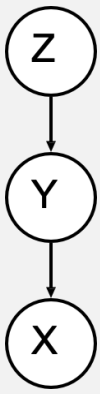
Information flow in causal graphs

- Information flows along
  - Causal paths
  - Non causal paths
- Non causal paths are the source of confounding
  - Why?
  - Because confounder of $X \rightarrow Y$ is any factor that leads makes

$$P\left(Y \mid X\right) \neq P\left(Y \mid do\left(X\right)\right)$$

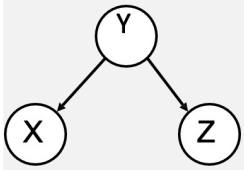
  - Do operator erases from the causal model, all arrows coming into X thus blocking information flow along non-causal paths

Information flow in causal graphs



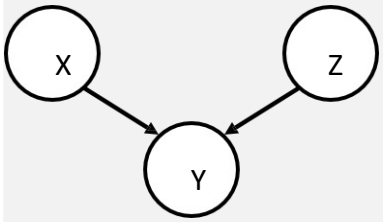
- Information flows along a chain Z → Y→ X
- Controlling for Y blocks information about Z getting to X (causal path) and vice versa (non causal path)

Information flow in causal graphs



- Information flows from X to Z
- Controlling for Y blocks information about Z getting to X and vice versa

Information flow in causal graphs



- No information flows from X to Z
- However, controlling for Y (or one of its descendents) allows information about Z from getting to X and vice versa

d-separation

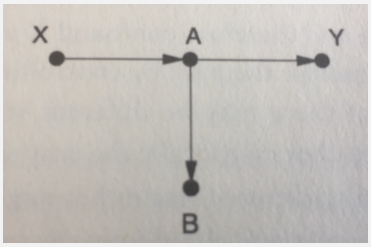
Definition: X and Z are d-separated by a set of de-confounders Y iff every undirected path from X to Z is "blocked" by controlling for Y

- *d*-separation can be computed using graph traversal in polynomial time

Confounding in the language of causal calculus
Backdoor criterion

- Confounder of $(X \text{ ? } Z)$ is any factor that makes $P(Y \mid X) \neq P(Y \mid do(X))$
- To de-confound two variables X and Y
  - We need to block all non-causal paths between X and Y without perturbing any causal paths
    - A backdoor path is any path from X to Y that starts with an arrow pointing into X
    - X and Y will be de-confounded if we block every such backdoor path
    - If we do this by controlling for some variables Z, we need to make sure that no member of Z is a descendent of X on a causal path
  - That is all there is to it!

Confounding through the lens of causal calculus



Source: The book of Why, Pearl and Mackenzie

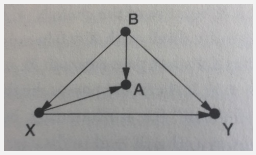
What do we need to control for in order to de-confound X and Y?

- Nothing!
  - There is no backdoor path into X
  - A, B are descendants of X (and hence should not be controlled for)

However,

- B passes a classical epidemiological definition of confounding
- But if we control for B, we introduce confounding rather than eliminating it!

Confounding through the lens of causal calculus

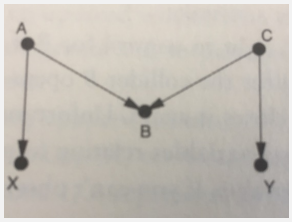


Source: The book of Why, Pearl and Mackenzie

- What do we need to control for in order to de-confound X and Y?

- There is a backdoor path X ←B →Y
- We can block it only by blocking B
- If B is observable, we are all set
- If B is unobservable
  - We cannot control for it, so there is no way we can de-confound X and Y, so there is no way to estimate the causal effect of X on Y without running a RCT
  - Current statistical practice would advocate controlling for A, a proxy of B – but this only partially eliminates the confounding bias and introduces a collider biasl

Confounding through the lens of causal calculus

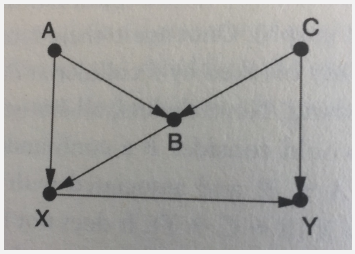


Source: The book of Why, Pearl and Mackenzie

- What do we need to control for in order to de-confound X and Y?

- There is a backdoor path X ←A →B ←C → Y which is already blocked by B
- Some of the correlation based statistical definitions of confounding would identify B as a confounder!
- B becomes a confounder when we control for it!
- Example
  - B – Seatbelt use, X – Smoking, A – Attitude towards societal norms, C – Attitude towards safety and health related measures, Y – lung cancer
  - A 2006 study found B to be correlated with both X and Y

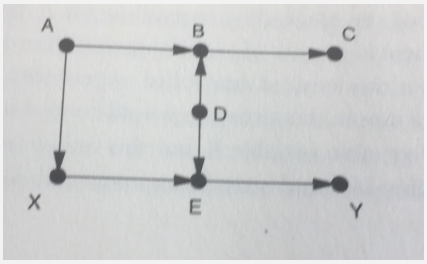
Confounding through the lens of causal calculus



- What do we need to control for in order to de-confound X and Y?
- A, B, C, D are pre-treatment variables, X is the treatment

- There is a backdoor path X ←A →B ←C → Y which is already blocked by B
- There is a second backdoor path X ←B ←C →Y
  - If we control B to block this path, we need to block A and C as well to ensure that the first backdoor path does not get unblocked
  - But blocking C alone suffices to block both the paths

Confounding through the lens of causal calculus

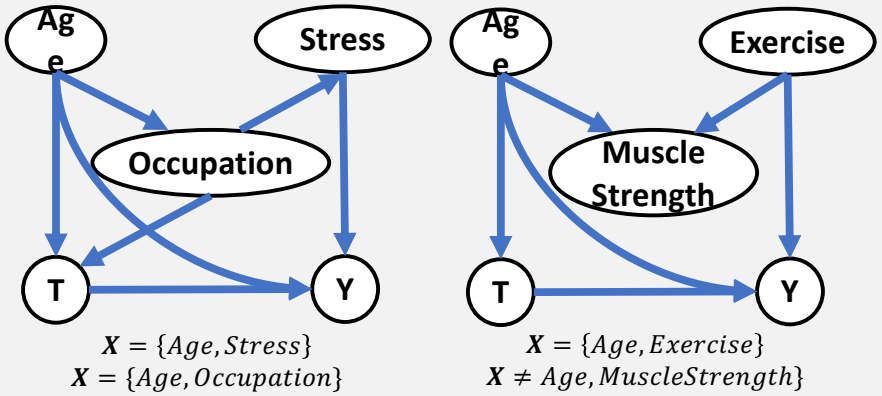


Source: The book of Why, Pearl and Mackenzie

- What do we need to control for in order to de-confound X and Y?

- A, B, C, D are pre-treatment variables, X is the treatment
- The only backdoor path X ←A → B ←D→ E→Y is already blocked by the collider B, so no need to control for anything!
- Standard statistical practice would be to control for B and C
  - "To avoid conditioning on some observed covariates … is scientific ad hockery" (sic)
  - Controlling fro B and C introduces confounding (unless we control for A or D as well)

Back-door criterion provides a precise way to find variables to control for



$\boldsymbol{X} = \{Age, Stress\}$
$\boldsymbol{X} = \{Age, Occupation\}$

$\boldsymbol{X} = \{Age, Exercise\}$
$\boldsymbol{X} \neq Age, MuscleStrength\}$

Confounding and causal models

- If we can identify and measure the confounders, we can control for them
- But as Pearl's work has shown, standard criteria for identifying confounders are flawed
- Both false positive and false negative confounders can yield misleading conclusions
- Causal calculus and tools based on graph theoretic criteria like d-separation provide effective methods for identifying the confounders (and only the confounders)

Identifying causal effects

- Postulate a causal graph (causal assumptions)
- See how the estimand can be written as a function of the postulated causal graph
- Check whether this function can be calculated from observations (using do calculus)

Identifying causal effects

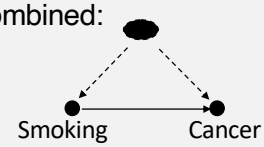
Surgeon General (1964):

$P(c \mid do(s)) \approx P(c \mid s)$

Tobacco Industry:

$P(c \mid do(s)) = P(c)$

Combined:



$P(c \mid do(s))$ - not identifiable

Structural Causal Models: The Story So Far

- Causal conclusions require causal assumptions
- Structural causal models encode causal assumptions
- Causal assumptions have testable implications – conditional independence relations (via d-separation)
- Causal effects are defined in terms of interventions
  - Average causal effect of (binary) D on Y is given by
    - $E[Y|do(D = 1)] - E[Y|do(D = 0)]$
  - We observe (samples from) $P(Y, X, D)$ and hence we can obtain $P(X)$ and $E[Y|X]$ etc.
  - Unless we have the resources and ability to experiment, we seldom observe $P(Y|do(D))$ and hence can't use it to obtain $E[Y|do(D)]$

Structural Causal Models: The Story So Far

- Identification of causal effects from observational data entails reducing $E[Y|do(D)]$ to an expression that is free of $do(\,)$, e.g., $\sum_x E[Y|D,X]P(X=x)$ using the causal assumptions encoded in the causal graph
- Once such reduction is done, $E[Y|do(D)]$ can be estimated from observational data
- In some cases, such identification is trivial. In other cases, it is not
- Primary challenge: observed or unobserved confounders
  - If we know the confounders X, and they are observed, we can adjust for them $E[Y|do(D)] = \sum_x E[Y|D,X]P(X=x)$
  - How do we know which confounders to adjust for?
  - Confounders are precisely those variables which make $P(Y|do(D)) \neq P(Y|D)$

- Unless we actually are in a situation where we have resources to intervene, we don't observe $E[Y|do(D)]$

Structural Causal Models: The Story So Far

- If we know the confounders X, and they are observed, we can adjust for them $E[Y|do(D)] = \sum_x E[Y|D,X]P(X=x)$
- How do we know which confounders to adjust for?
- Confounders are precisely those variables which make $P(Y|do(D)) \neq P(Y|D)$
- Backdoor criterion allows us to identify the confounders
- A path that starts with an arrow into D is called a **back-door path**
  - Blocking back-door paths ensures we block bad, i.e., non-causal, paths
  - Not conditioning on descendants of D ensures that we leave all good, i.e., causal, paths open and that we do not open up new bad paths

Identifying causal effects

- Given a causal Bayesian network G, we say that a causal effect $P(Y|do(X))$ is identifiable when $P(Y|do(X))$ can be computed using only the joint distribution over the observable variables

General recipe for identifying causal effects

- Postulate a causal graph (causal assumptions)
- See how the estimand (causal effect of interest) can be written as a function of the postulated causal graph
- Check whether this function can be calculated from observations (using do calculus)

Some notation

- Let G be a causal model on a graph, and W, X, Y, Z be disjoint disjoint subsets of the variables in the causal model.
- Let $G_{\overline{X}}$ denote the perturbed graph in which all edges pointing to X from the parents of X in G have been deleted. This is the graph that models the results of an intervention on X.
- Let $G_{\underline{X}}$ denote the graph in which all edges out of X to the children of X in G have been deleted.
- We will also freely use notations like $G_{\overline{X}\underline{W}\overline{Z}}$ to denote combinations of the above operations.

Do-calculus

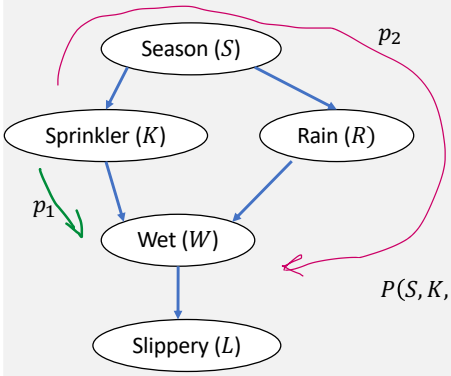
Theorem (Rules of do-calculus): Given a causal graph G, an associated distribution P, and disjoint sets of variables Y, T, Z, and W, the following rules hold:

- $P(Y \mid do(T = t), Z = z, W = w)$
  $= P(Y \mid do(T = t, W = w)$ if $Y \coprod_{G_{\overline{T}}} Z \mid T, W.$
- $P(Y \mid do(T = t), do(Z = z), W = w)$
  $= P(Y \mid do(T = t), Z = z, W = w)$ if $Y \coprod_{G_{\overline{T}\underline{Z}}} Z \mid T, W$
- $P(Y \mid do(T = t), do(Z = z), W = w)$
  $= P(y \mid do(T = t), W = w)$ if $Y \coprod_{G_{\overline{TZ(W)}}} Z \mid T, W$
  where $Z(W)$ denotes the set of nodes of Z that aren't ancestors of any node of W in $G_{\overline{T}}$.

Identifying causal effects from observations

- Postulate a causal graph
- Write the causal effect of interest according to the postulated causal graph
- Apply do-calculus rules repeatedly to reduce the causal effect expression containing *do* into one free of *do*
  - If the expression for the causal effect containing a *do* can be reduced to one that is free of *do,* the causal effect is identifiable from observations
  - If the expression cannot be reduced to one that is free of do, the causal effect is not identifiable from observations

Causal and non-causal associations



$$P(S,K,R,W,L) = P(L|W)P(W|K,R)P(K|S)P(R|S)$$

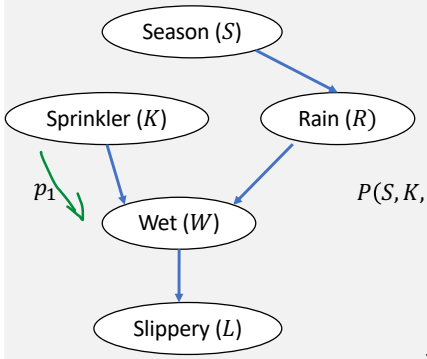
$$P(W|K=1) = P(p_1) + P(p_2)$$

$$= \frac{P(K=1,W)}{P(K=1)}$$

$$= \frac{\sum_{s,r} P(W|K=1,r)P(K=1|s)P(r|s)P(s)}{\sum_{s,} P(K=1|s)P(s)}$$

Type equation here.

Causal and non-causal associations



$$P(S,K,R,W,L) = P(L|W)P(W|K,R)P(K|S)P(R|S)$$

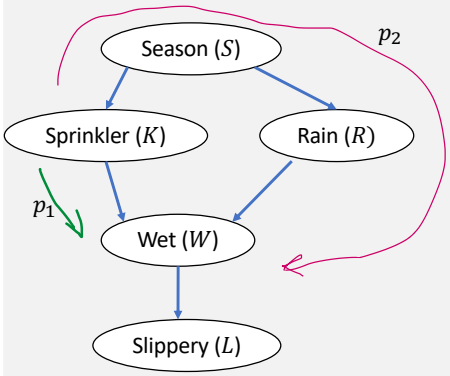
$$P(W|do(K=1)) = P(p_1)$$

$$= \frac{\sum_{s,r} P(W|K=1,r)P(K=1)P(r|s)P(s)}{P(K=1)}$$

$$= \sum_{s,r} P(W|K=1,r)P(r|s)P(s)$$

Type equation here.

Causal and non-causal associations



$P(W|K = 1) = P(p_1) + P(p_2)$ (Association)

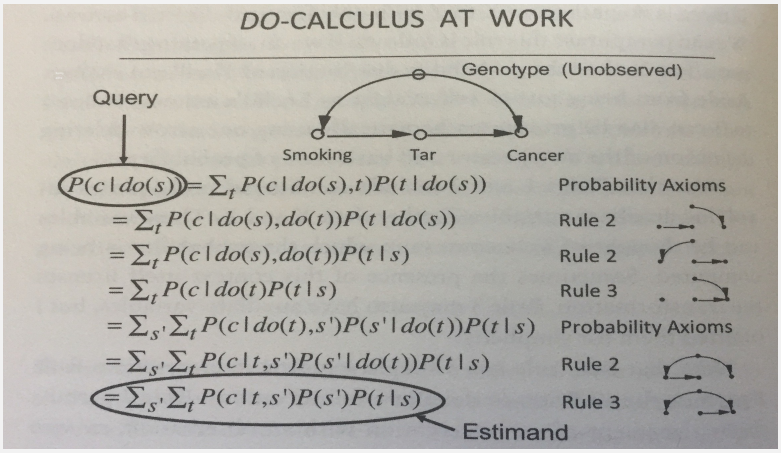
$P(W|do(K = 1)) = P(p_1)$ (Causation)

Causal models show us

- Why not all associations are causal
- When it is possible to distinguish one from the other
- How to identify causal effects (when they are identifiable)

Type equation here.

DO-CALCULUS AT WORK



Source: The book of Why, Pearl and Mackenzie