Causality

Macartan Humphreys

Section 1

Causality. What's a cause?

Subsection 1

Potential outcomes and the counterfactual approach

Potential outcomes and the counterfactual approach

Causation as difference making

Motivation

The *intervention* based motivation for understanding causal effects:

- We want to know if a particular intervention (like aid) caused a particular outcome (like reduced corruption).
- We need to know:
  1. What happened?
  2. What would the outcome have been if there were no intervention?
- The problem:
  1. ... this is hard
  2. ... this is impossible

The problem in 2 is that you need to know what would have happened if things were different. You need information on a **counterfactual**.

Potential Outcomes

- For each unit, we assume that there are two **post-treatment** outcomes: $Y_i(1)$ and $Y_i(0)$.
- For example, $Y(1)$ is the outcome that *would* obtain *if* the unit received the treatment.
- The **causal effect** of Treatment (relative to Control) is: $\tau_i = Y_i(1) - Y_i(0)$
- Note:
  - The causal effect is defined at the *individual level*.
  - There is no "data generating process" or functional form.
  - The causal effect is defined relative to something else, so a counterfactual must be conceivable (did Germany cause the second world war?).
  - Are there any substantive assumptions made here so far?

Potential Outcomes

Idea: A causal claim is (in part) a claim about something that did not happen. This makes it metaphysical.

Potential Outcomes

Now that we have a concept of causal effects available, let's answer two **questions**:

- TRANSITIVITY: If for a given unit A causes B and B causes C, does that mean that A causes C?

Potential Outcomes

Now that we have a concept of causal effects available, let's answer two **questions**:

- TRANSITIVITY: If for a given unit A causes B and B causes C, does that mean that A causes C?
- A boulder is flying down a mountain. You duck. This saves your life.
- So the boulder caused the ducking and the ducking caused you to survive.
- So: *did the boulder cause you to survive?*

Potential Outcomes

CONNECTEDNESS Say A causes B — does that mean that there is a spatiotemporally continuous sequence of causal intermediates?

Potential Outcomes

CONNECTEDNESS Say A causes B — does that mean that there is a spatiotemporally continuous sequence of causal intermediates?

- Person A is planning some action Y; Person B sets out to stop them; person X intervenes and prevents person B from stopping person A. In this case Person A may complete their action, producing Y, without any knowledge that B and X even exist; in particular B and X need not be anywhere close to the action. So: *did X cause Y*?

Causal claims: Contribution or attribution?

The counterfactual model is all about contribution, not attribution, except in a very conditional sense.

- Focus is on non-rival contributions
- Not: what caused Y but what is the effect of X?
- At most it provides a conditional account

Causal claims: Contribution or attribution?

Consider an outcome Y that might depend on two causes X_1 and X_2:

$$Y(0,0) = 0$$
$$Y(1,0) = 0$$
$$Y(0,1) = 0$$
$$Y(1,1) = 1$$

What caused Y? Which cause was most important?

Causal claims: Contribution or attribution?

The counterfactual model is about attribution in a very conditional sense.

- Focus is on non-rival contributions
- Not: what caused Y but what is the effect of X?
- At most it provides a conditional account
- This is problem for research programs that define "explanation" in terms of figuring out the things that cause Y
- Real difficulties conceptualizing what it means to say one cause is more important than another cause. What does that mean?

Causal claims: Contribution or attribution?

Erdogan's increasing authoritarianism was the most important reason for the attempted coup

- More important than Turkey's history of coups?
- What does that mean?

Causal claims: No causation without manipulation

- Some seemingly causal claims not admissible.
- To get the definition off the ground, manipulation must be imaginable (whether practical or not)
- This renders thinking about effects of race and gender difficult
- What does it mean to say that Aunt Pat voted for Brexit because she is old?

Causal claims: No causation without manipulation

- Some seemingly causal claims not admissible.
- To get the definition off the ground, manipulation must be imaginable (whether practical or not)
- This renders thinking about effects of race and gender difficult
- **Compare**: What does it mean to say that Southern counties voted for Brexit because they have many old people?

Causal claims: Causal claims are everywhere

- Jack exploited Jill
- It's Jill's fault that bucket fell
- Jack is the most obstructionist member of Congress
- Melania Trump stole from Michelle Obama's speech
- Activists need causal claims

Causal claims: What is actually seen?

- We have talked about what's potential, now what do we *observe*?
- Say Z_i indicates whether the unit i is assigned to treatment $(Z_i = 1)$ or not $(Z_i = 0)$. It describes the treatment process. Then what we observe is:

$$Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$$

This is sometimes called a "switching equation"

In `DeclareDesign` Y is realised from potential outcomes and assignment in this way using `reveal_outcomes`

Causal claims: What is actually seen?

- Say Z is a random variable, then this is a sort of data generating process. BUT the key thing to note is
  - Y_i is random but the randomness comes from Z_i — the potential outcomes, $Y_i(1)$, $Y_i(0)$ are fixed
  - Compare this to a regression approach in which Y is random but the X's are fixed. eg:

$$Y \sim N(\beta X, \sigma^2) \text{ or } Y = \alpha + \beta X + \epsilon, \epsilon \sim N(0, \sigma^2)$$

Causal claims: The estimand and the rub

- The causal effect of Treatment (relative to Control) is:

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

- This is what we want to estimate.
- BUT: We never can observe both $Y_i(1)$ and $Y_i(0)$!
- This is the **fundamental problem** (@holland1986statistics)

Causal claims: The rub and the solution

- Now for some magic. We really want to estimate:

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

- BUT: We never can observe both $Y_i(1)$ and $Y_i(0)$
- Say we lower our sights and try to estimate an *average* treatment effect:

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

- Now make use of the fact that

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

- In words: *The average of differences is equal to the difference of averages*; here, the average treatment effect is equal to the difference in average outcomes in treatment and control units.
- The magic is that *while we can't hope to measure the differences; we*

Causal claims: The rub and the solution

- So we want to estimate $\mathbb{E}[Y(1)]$ and $\mathbb{E}[Y(0)]$.
- We know that we can estimate averages of a quantity by taking the average value from a random sample of units
- To do this here we need to select a random sample of the $Y(1)$ values and a random sample of the $Y(0)$ values, in other words, we **randomly assign** subjects to treatment and control conditions.
- When we do that we can in fact estimate:

$$\mathbb{E}_N[Y_i(1)|Z_i = 1) - \mathbb{E}_N(Y_i(0)|Z_i = 0]$$

which in expectation equals:

$$\mathbb{E}[Y_i(1)|Z_i = 1 \text{ or } Z_i = 0] - \mathbb{E}[Y_i(0)|Z_i = 1 \text{ or } Z_i = 0]$$

- This highlights a deep connection between **random assignment** and **random sampling**: when we do random assignment *we are in fact randomly sampling from different possible worlds*.

Causal claims: The rub and the solution

This provides a **positive argument** for causal inference from randomization, rather than simply saying with randomization "everything else is controlled for"

Let's discuss:

- *Does the fact that an estimate is unbiased mean that it is right?*
- *Can a randomization "fail"?*
- *Where are the covariates?*

Idea: random assignment is random sampling from potential worlds: to understand anything you find, you need to know the sampling weights

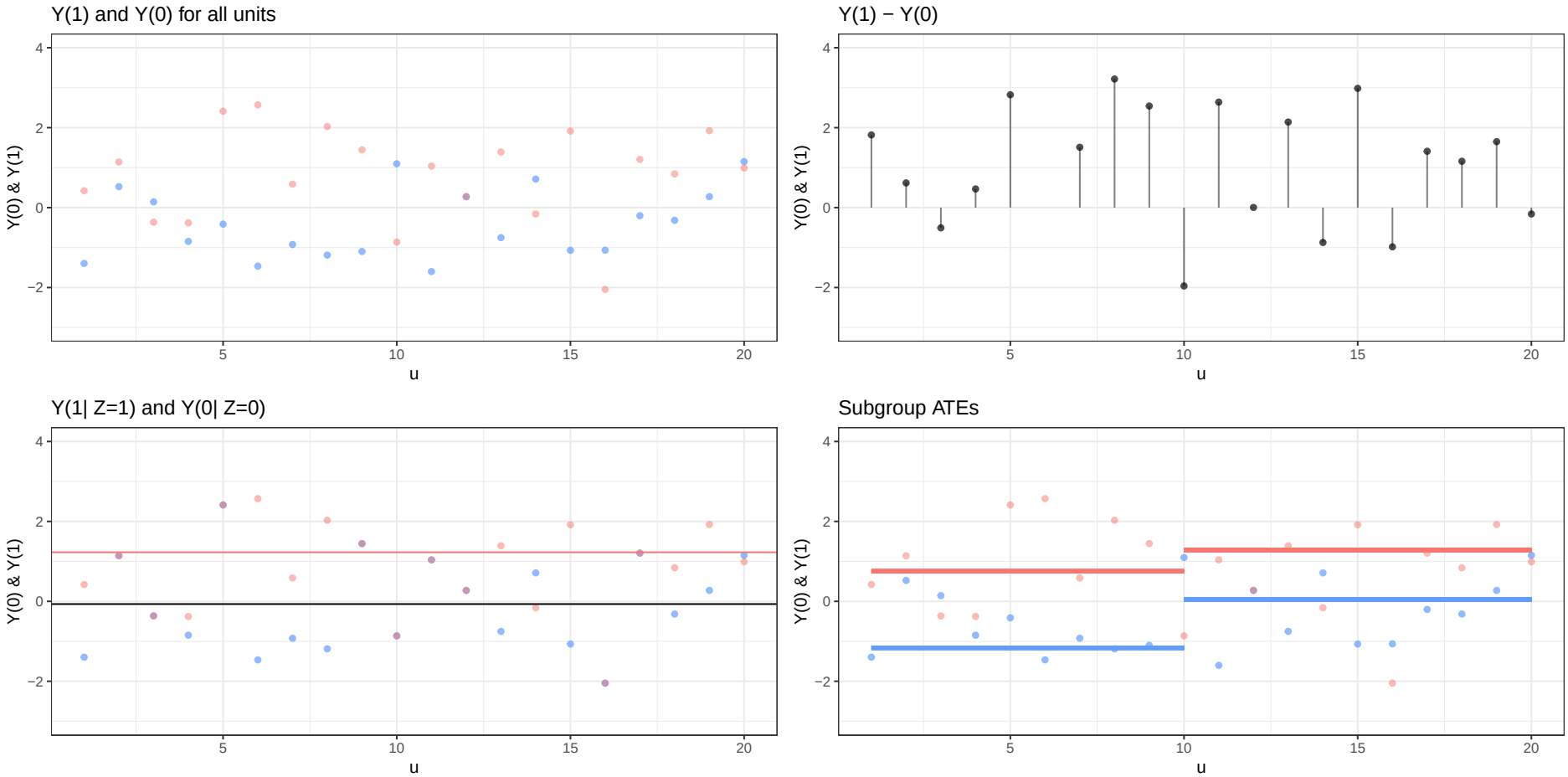
Reflection

Idea: We now have a *positive* argument for claiming unbiased estimation of the average treatment effect following random assignment

But is the average treatment effect a quantity of *social scientific* interest?

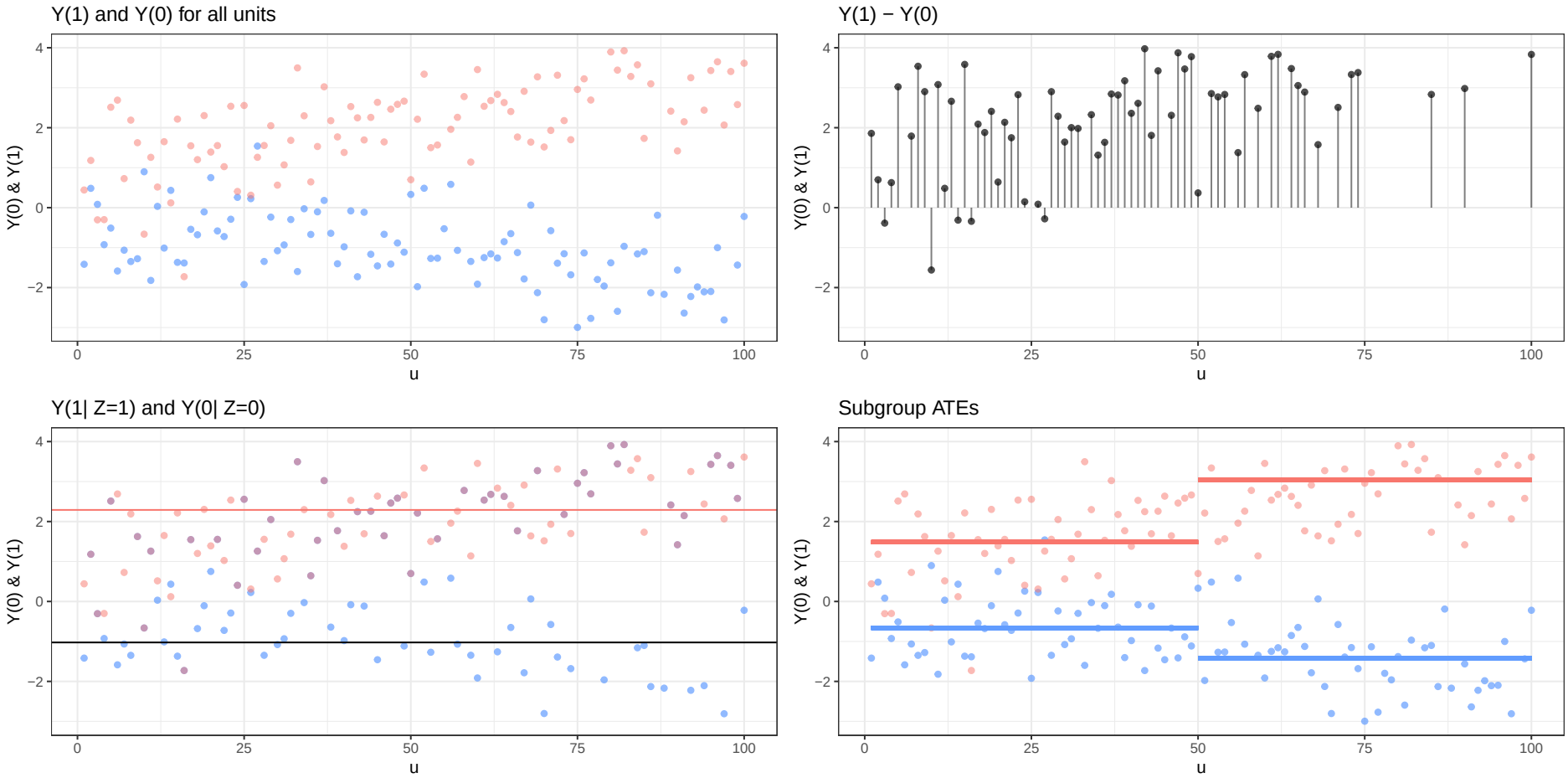
Potential outcomes: why randomization works

The average of the differences $\approx$ difference of averages



Potential outcomes: heterogeneous effects

The average of the differences $\approx$ difference of averages



Potential outcomes: heterogeneous effects

Question: $\approx$ or $=$?

Exercise your potential outcomes 1

Consider the following potential outcomes table:

Unit	Y(0)	Y(1)	τ_i
1	4	3	
2	2	3	
3	1	3	
4	1	3	
5	2	3	

Questions for us: What are the unit level treatment effects? What is the average treatment effect?

Exercise your potential outcomes 2

Consider the following potential outcomes table:

In treatment?	Y(0)	Y(1)
Yes		2
No	3	
No	1	
Yes		3
Yes		3
No	2	

Questions for us: Fill in the blanks.

- Assuming a constant treatment effect of $+1$
- Assuming a constant treatment effect of -1
- Assuming an *average* treatment effect of 0

What is the actual treatment effect?

Subsection 2

Endogeneous subgroups

Endogeneous Subgroups

Experiments often give rise to endogenous subgroups. The potential outcomes framework can make it clear why this can cause problems.

Heterogeneous Effects with Endogeneous Categories

- Problems arise in analyses of subgroups when the categories themselves are affected by treatment
- Example from our work:
  - You want to know if an intervention affects reporting on violence against women
  - You measure the share of all subjects that experienced violence that file reports
  - The problem is that which subjects experienced violence is itself a function of treatment

Heterogeneous Effects with Endogeneous Categories

It is possible that in truth no one's reporting behavior has changed, what has changed is the propensity of people with different propensities to report to experience violence:

```
\begin{table} \scriptsize
    \centering
    \begin{tabular}{rcc|cc|cc}

        & \multicolumn{ 2}{c}{Violence(Treatment)} & \multicol

        &       V(0) &       V(1) &     R(0,1) &     R(1,1) &

        Type 1 (reporter) &          1 &          1 &

        Type 2 (non reporter) &          1 &          0 &
    \end{tabular}
\end{table}
```

Heterogeneous Effects with Endogeneous Categories

- **Violence(Treatment)**
- **Reporting(Treatment, Violence)**

	V(0)	V(1)	R(0,1)	R(1,1)	R(0,0)	R(1,0)
Type 1 (reporter)	1	1	1	1	0	0
Type 2 (non reporter)	1	0	0	0	0	0

Expected reporting given violence in control = Pr(Type 1)

Expected reporting given violence in treatment = 100%

Question: What is the actual effect of treatment on the propensity to report violence?

Heterogeneous Effects with Endogeneous Categories

It is possible that in truth no one's reporting behavior has changed, what has changed is the propensity of people with different propensities to report to experience violence:

	Reporters		Non reporters		
	Experience Violence		Experience Violence		
	No	Yes	No	Yes	% Report
Control	25	25	25	25	$\frac{25}{25+25} = 50\%$
Treatment	25	25	50	0	$\frac{25}{25+0} = 100\%$

Heterogeneous Effects with Endogeneous Categories

This problem can arise as easily in seemingly simple field experiments. Example:

- In one study we provided constituents with information about performance of politicians
- we told politicians in advance so that they could take action
- we wanted to see whether voters punished poorly performing politicians
- what's the problem?

Heterogeneous Effects with Endogeneous Categories

Question for us:

Setting:

```
* Quotas for women are randomly placed in a set of constit
* In year 2 these quotas are then lifted.
```

Questions Which problems face an endogenous subgroup issue?:

1. You want to estimate the likelihood that a woman will stand for reelection in treatment versus control areas in year 2.
2. You want to estimate how much incumbents are more likely to be reelected in treatment versus control areas in year 2.
3. You want to estimate how much treatment areas have more relected incumbents in elections in year 2 compared to control.

Heterogeneous Effects with Endogeneous Categories

In such cases you can:

- Examine the joint distribution of multiple outcomes
- Condition on pretreatment features only
- Engage in mediation analysis

Missing data can create an endogeneous subgroup problem

- It is well known that missing data can undo the magic of random assignment.
- One seemingly promising approach is to match into pairs *ex ante* and drop pairs together *ex post*.
- Say potential outcomes looked like this (four units divided into two pairs):

Pair	I	I	II	II	
Unit	1	2	3	4	Average
Y(0)	0	0	0	0	
Y(1)	-3	1	1	1	
τ	-3	1	1	1	

Missing data

- Say though that cases are likely to drop out of the sample if things go badly (eg they get a negative score or die)
- Then you might see no attrition in cases in which people that are likely to drop out if treated do not get treated.
- You might assume you have no problem (after all, no attrition).
- No missing data when the normal cases happens to be selected

Pair	I	I	II	II	
Unit	1	2	3	4	Average
Y(0)	0		0		0
Y(1)		1		1	1
$\widehat{\tau}$					1

Missing data

- But in cases in which you have attrition, dropping the pair doesn't necessarily help.
- The problem is potential missingness still depends on potential outcomes
- The kicker is that the method can produce bias even if (*in fact*) there is no attrition!

Missing data when the vulnerable cases happens to be selected

Pair	I	I	II	II	
Unit	1	2	3	4	Average
Y(0)		[0]	0		0
Y(1)	[-3]			1	1
$\widehat{\tau}$					1

Missing data

Note: The right way to think about this is that bias is a property of the strategy over possible realizations of data and not normally a property of the estimator conditional on the data.

Multistage games

Multistage games can also present an endogenous group problem since collections of late stage players facing a given choice have been created by early stage players.

Multistage games

Question: Does **visibility** alter the extent to which subjects follow norms to punish antisocial behavior (and reward prosocial behavior)? Consider a trust game in which we are interested in how information on receivers affects their actions

Table 7: Return rates given investments under different conditions

		% invested	Average % returned	
		(average)	...when 10% invested	...when 50% invested
Treatment	Masked information on respondents	30% (avg)	20%	40%
	Full information on respondents	30% (avg)	0%	60%

What do we think? Does visibility make people react more to investments?

Multistage games

Imagine you could see all the potential outcomes, and they looked like this:

Table 8: Potential outcomes with (and without) identity protection

		Responder's return decision (given type)						Avg.
		Nice 1	Nice 2	Nice 3	Mean 4	Mean 4	Mean 6	
Offerer behavior	Invest 10%:	60%	60%	60%	0%	0%	0%	30%
	Invest 50%:	60%	60%	60%	0%	0%	0%	30%

Conclusion: Both the offer and the information condition are **completely irrelevant** for all subjects.

Multistage games

Unfortunately you only see a sample of the potential outcomes, and that looks like this:

Table 9: Outcomes when respondent **is visible**

		Responder's return decision (given type)						Avg.
		Nice 1	Nice 2	Nice 3	Mean 4	Mean 4	Mean 6	
Offerer behavior	Invest 10%:				0%	0%	0%	0%
	Invest 50%:	60%	60%	60%				60%

False Conclusion: When not protected, responders condition behavior *strongly* on offers (because offerers can select on type accurately)

Multistage games

Unfortunately you only see a sample of the potential outcomes, and that looks like this:

Table 10: Outcomes when respondent **is not visible**

		Responder's return decision (given type)						Avg.
		Nice 1	Nice 2	Nice 3	Mean 4	Mean 4	Mean 6	
Offerer behavior	Invest 10%:			60%		0%	0%	20%
	Invest 50%:	60%	60%		0%			40%

False Conclusion: When protected, responders condition behavior less strongly on offers (because offerers can select on type less accurately)

Multistage games

What to do?

Solutions?

1. Analysis *could* focus on the effect of treatment on respondent behavior, directly.
   - This would get the correct answer but to a different question [Does information affect the share of contributions returned by subjects on average? No]
2. **Strategy method** can sometimes help address the problem, **but** that is also (a) changing the question and (b) putting demands on respondent imagination and honesty
3. First mover action could be **directly manipulated**, but unless deception is used that is also changing the question
4. First movers could be **selected** because they act in predictable ways (bordering on deception?)

Idea: Proceed with extreme caution when estimating effects beyond the

Subsection 3

DAGs

Subsection 4

Key insight

Key insight

The most powerful results from the study of DAGs are procedures for figuring out when conditioning aids or hinders causal identification.

- You can read off a **confounding** variable from a DAG.
  - You have to condition on such a variable for causal identification.
- You can read off **"colliders"** from a DAG
  - Sometimes you have *avoid* conditioning on these
- Sometimes a variable might be both, so
  - you have to condition on it
  - you have to avoid conditioning on it
  - Ouch.

Subsection 5

Key resource

Key resource

- Pearl's book *Causality* is the key reference. @pearl2009causality (Though see also older work such as @pearl1985graphoids)
- There is a lot of excellent material on Pearl's page http://bayes.cs.ucla.edu/WHY/
- See also excellent material on Felix Elwert's page http://www.ssc.wisc.edu/~felwert/causality/?page_id=66

Subsection 6

Challenge for us

Challenge for us

- Say you don't like graphs. Fine.
- Consider this causal structure:
  - $Z = f_1(U_1, U_2)$
  - $X = f_2(U_2)$
  - $Y = f_3(X, U_1)$

Say Z is temporally prior to X; it is correlated with Y (because of U_1) and with X (because of U_2).

Question: Would it be useful to "control" for Z when trying to estimate the effect of X on Y?

Subsection 7

Challenge for us

Challenge for us

- Say you don't like graphs. Fine.
- Consider this causal structure:
  - $Z = f_1(U_1, U_2)$
  - $X = f_2(U_2)$
  - $Y = f_3(X, U_1)$

Question: Would it be useful to "control" for Z when trying to estimate the effect of X on Y?

Answer: Hopefully by the end of today you should see that that the answer is obviously (or at least, plausibly) "no."

Subsection 8

Conditional independence

Conditional independence

Variable sets A and B are conditionally independent, given C if for all a, b, c:

$$\Pr(A = a|C = c) = \Pr(A = a|B = b, C = c)$$

Informally; given C, knowing B tells you nothing more about A.

Subsection 9

Causal graphs basics 1

Causal graphs basics 1

- Consider a situation with variables $X_1, X_2, ... X_n$
- The probability of outcome x can always be written in the form $P(X_1 = x_1)P(X_2 = x_2|X_1 = x_1)(X_3 = x_3|X_1 = x_1, X_2 = x_2)....$
- This can be done with any ordering of variables.
- However the representation can be greatly simplified if you can make use of a set of "parentage" relationships
- Given an ordering of variables, the **Markovian parents** of variable X_j are the minimal set of variables such that when you condition on these, X_j is independent of all other prior variables in the ordering
- In this case we can write: $P(x) = \prod_j (x_j|pa_j)$
- No graphs yet

Subsection 10

Causal graphs basics 2

Causal graphs basics 2

- We want to use causal graphs to represent these relations of conditional independence.
- Informally, an arrow, $A \rightarrow B$ means that A is a cause of B: that is, under some conditions, a change in A produces a change in B.
  - Arrows carry no information about the type of effect; e.g. sign, size, or whether different causes are complements or substitutes
- We say that arrows point from *parents* to *children*, and by extension from *ancestors* to *descendants*.
- These are parents *on the graph*; but we will connect them to Markovian parents in a probability distribution P.

Subsection 11

Causal graphs basics 2

Causal graphs basics 2

- A DAG is just a graph in which some or all nodes are connected by *directed* edges (arrows) and there are no cyclical paths along these directed edges.
- Consider a DAG, G, and consider the ancestry relations implied by G: the distribution P is *Markov relative to the graph* G if every variable is independent of its nondescendants (in G) conditional on its parents (in G).
  - This is the **Markov condition**: conditional on its parents, a variable is independent of its non-descendants.
- OK now we have a link from probability distributions to graphs. But we have not talked about causality.

Subsection 12

Causal graphs basics 3

Causal graphs basics 3

We want the graphs to be able to represent the effects of interventions.

Pearl uses do notation to capture this idea.

$$\Pr(X_1, X_2, ... | do(X_j = x_j))$$

or

$$\Pr(X_1, X_2, ... | \hat{x}_j)$$

denotes the distribution of X when a particular node (or set of nodes) is intervened upon and forced to a particular level, x_j.

Subsection 13

Causal graphs basics 3

Causal graphs basics 3

Note, in general:

$$\Pr(X_1, X_2, ... | do(X_j = x'_j)) \neq \Pr(X_1, X_2, ... | X_j = x'_j)$$

as an example we might imagine a situation where for men binary X always causes $Y = 1$ and for women $Y = 1$ regardless of X. We imagine that $X = 1$ for men only.

In that case $\Pr(Y = 1|X = 1) = 1$ but $\Pr(Y = 1|do(X = 1)) = .5$

Subsection 14

Causal graphs basics 3

Causal graphs basics 3

- Let P_z denote the resulting distribution on all variables that arises when vector Z is "set" (forced, controlled...) to the value z. That is when we have `do(Z=z)`.
- Let P_* denote the set of all such distributions that can result from any set of interventions on variables.
- A DAG, G, is a **causal Bayesian network compatible with** P_* iff, for all interventions z:
  1. P_z is Markov relative to G
  2. $P_z(x_i) = 1$ for all x_i consistent with z
  3. $P_z(x_j|pa_j) = P(x_j|pa_j)$ for all $x_j \notin Z$ when pa_j is consistent with z

Subsection 15

Causal graphs basics 3

Causal graphs basics 3

- That all means that the probability distribution resulting from setting some set X_i to $\hat{x'}_i$ (i.e. `do(X=x')`) is:

$$P_{\hat{x}_i} = P(x_1, x_2, ... x_n | \hat{x}_i) = \prod_{-i} P(x_j | pa_j) \mathbb{1}(x_i = x'_i)$$

This means that there is only probability mass on vectors in which $x_i = x'_i$ (reflecting the success of control) and all other variables are determined by their parents, given the values that have been set for x_i.

Subsection 16

Conditional Independence and d-separation

Conditional Independence and d-separation

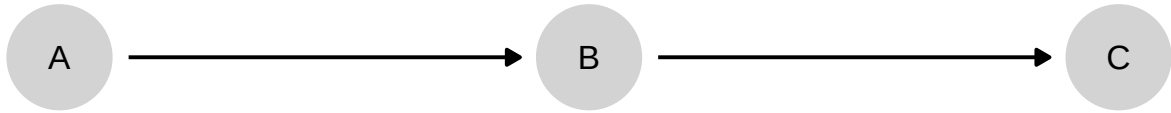
- We now have a well defined sense in which the arrows on a graph represent a causal structure and capture the conditional independence relations implied by the causal structure.
- Of course any graph might represent many different probability distributions P
- We can now start reading off from a graph when there is or is not conditional independence between sets of variables

Subsection 17

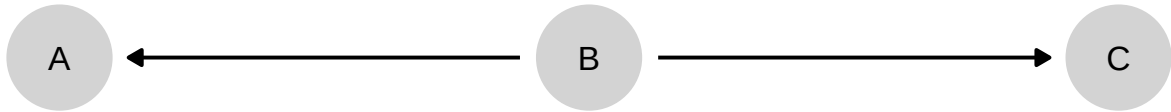
Conditional independence on paths

Conditional independence on paths

(1) A path of arrows pointing in the same direction



(2) A forked path



(3) An inverted fork (collision)

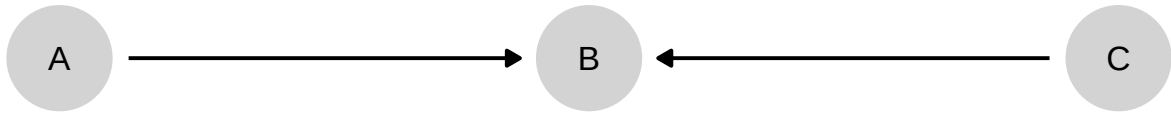


Figure 1: Three elemental relations of conditional independence.

Subsection 18

Conditional independence

Conditional independence

A and B are *conditionally independent*, given C if on *every* path between A and B:

- there is some chain ($\bullet \rightarrow \bullet \rightarrow \bullet$ or $\bullet \leftarrow \bullet \leftarrow \bullet$) or fork ($\bullet \leftarrow \bullet \rightarrow \bullet$) with the central element in C,

or

- there is an inverted fork ($\bullet \rightarrow \bullet \leftarrow \bullet$) with the central element (and its descendants) *not* in C

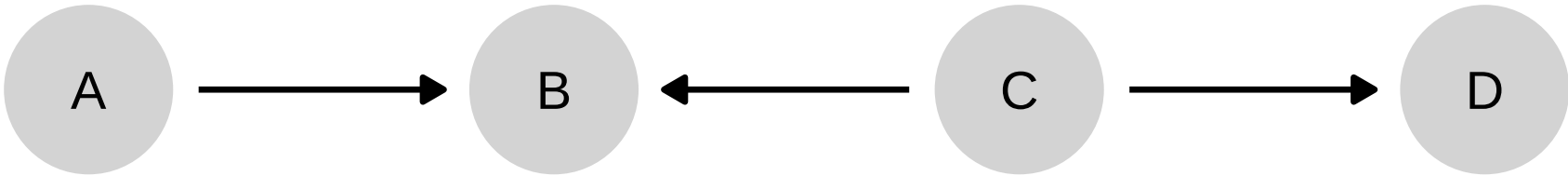
Notes:

- In this case we say that A and B are d-separated by C.
- A, B, and C can all be sets
- Note that a path can involve arrows pointing any direction
  $\bullet \rightarrow \bullet \rightarrow \bullet \leftarrow \bullet \rightarrow \bullet$

Subsection 19

Test yourself

Test yourself



Are A and D unconditionally independent:

- if you do not condition on anything?
- if you condition on B?
- if you condition on C?

Subsection 20

Back to this example

Back to this example

* $Z = f_1(U_1, U_2)$
* $X = f_2(U_2)$
* $Y = f_3(X, U_1)$

1. Let's graph this
2. Now: say we removed the arrow from X to Y
   - Would you expect to see a correlation between X and Y if you did not control for Z
   - Would you expect to see a correlation between X and Y if you did control for Z

Subsection 21

From graphs to Causal Models

From graphs to Causal Models

A "**causal model**" is:

1.1: An ordered list of n endogenous nodes, $\mathcal{V} = (V^1, V^2, \ldots, V^n)$, with a specification of a range for each of them

1.2: A list of n exogenous nodes, $\Theta = (\theta^1, \theta^2, \ldots, \theta^n)$

2: A list of n functions $\mathcal{F} = (f^1, f^2, \ldots, f^n)$, one for each element of $\mathcal{V}$ such that each f^i takes as arguments θ^i as well as elements of $\mathcal{V}$ that are *prior* to V^i in the ordering

and

3: A probability distribution over Θ

Subsection 22

From graphs to Causal Models

From graphs to Causal Models

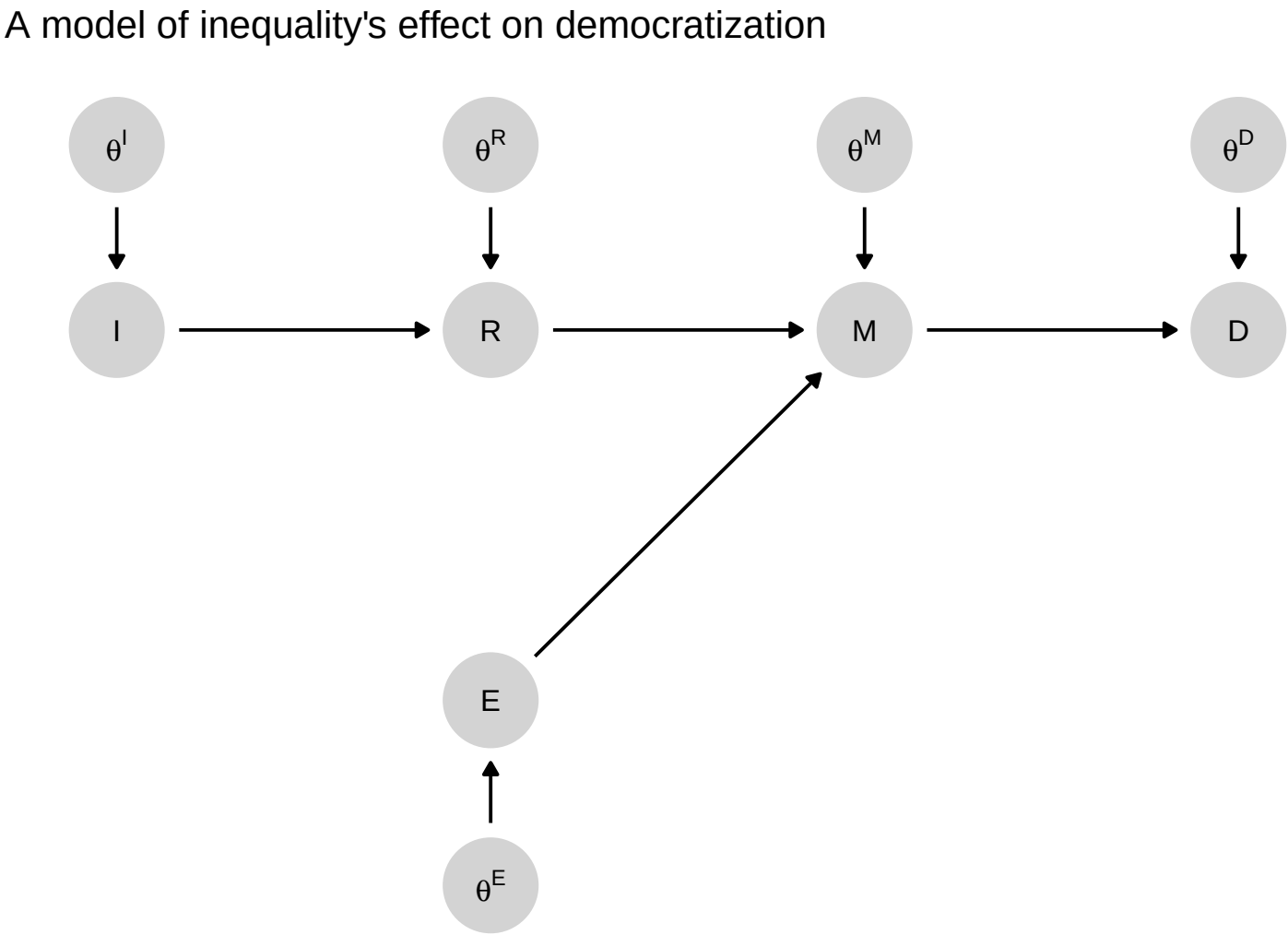


Figure 2: A simple causal model in which high inequality (I) affects democratization (D) via redistributive demands (R) and mass mobilization (M)

Effects on a DAG

Learning about effects *given* a model means learning about F and *also* the distribution of shocks (Θ).

For discrete data this can be reduced to a question about learning about the distribution of Θ only.

Subsection 23

Recap: Key features of graphs

Recap: Key features of graphs

- Directed
- Acyclic
- The missing arcs are the really important ones
- Implicitly there are shocks going into every node
- These graphs represent Nonparametric structural equation models NPSEMs
- But you cannot read off the size or direction of effects from a DAG

Recap: Ten things you need to know about causal inference

1. A causal claim is a statement about what didn't happen.
2. There is a fundamental problem of causal inference.
3. You can estimate average causal effects even if you cannot observe any individual causal effects.
4. If you know that A causes B and that B causes C, this does not mean that you know that A causes C.
5. The counterfactual model is primarily about contribution, and about attribution in a limited sense.
6. X can cause Y even if there is no "causal path" connecting X and Y.
7. Correlation is not causation.
8. X can cause Y even if X is not a necessary condition or a sufficient condition for Y.
9. Estimating average causal effects does not require that treatment and control groups are identical.
10. There is no causation without manipulation.

http://egap.org/resources/guides/causality/