CompSci 295, Causal Inference

Rina Dechter, UCI

Lecture 6b: Counterfactuals

Slides: Primer, chapter 4

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.
  - The 3-steps
  - Do operators are limited
  - Expressing do by counterfactuals
  - The graphical representation of counterfactuals

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.

The three ladder of cause and effect

- What if I see? (a customer buy toothpaste… will he buy dental floss)
  - Answer: from data P(buy DF| buy toothpaste). First ladder is observing
- What if I act: (What would happen to our toothpaste sale if we double the price?) P(Y| do(x))?
- What if I had acted differently: Google example (Bozhena): "it is all about counterfactuals" how to determine the price of an addvertizement. A customer bought an item Y and ad x was observed. What is the likelihood he would have bought the product has ad x not been used.
- "No learning machine in operation today can answer the questions about questions not taken before. Moreover, most learning machine today do not utilize a representation from which such questions can be answered" (Pearl, position paper)

The three ladders of cause and effect

- Counterfactuals subsumes the higher levels.

Counterfactuals

Common sense: “While driving home last night, I came to a fork in the road where I had to make a choice: to take the freeway (X = 1) or go on a surface street named Sepulveda Boulevard (X = 0). I took Sepulveda, only to find out that the traffic was touch and go. As I arrived home, an hour later, I said to myself: “Gee, I should have taken the freeway.”

Economy: Would a customer by the shoes online had the advertisement not been there?

Politics: Had Hillery won the election had Comey not announced 10 days before election that FBI reopen The investigation into her email servers?

This kind of statement: an “if” statement in which the “if” portion is untrue or unrealized—is known as a **counterfactual**. The “if” portion of a counterfactual is called the hypothetical condition, or more often, the antecedent.

Require a new language beyond “do” or intervention

Counterfactual Representation

If we try to express this estimate using do-expressions, we come to an impasse. Writing

$$E(\text{driving time} \mid do(\text{freeway}), \text{driving time} = 1 \text{ hour})$$

leads to a clash between the driving time we wish to estimate and the actual driving

1. Actual driving time
2. Hypothetical driving time under freeway conditions when actual surface driving time is known to be 1 hour.

The do-operator allows us to distinguish between two probabilities, P(driving time$|do$(freeway))) and P(driving time$|do$(Sepulveda)), it does not offer us the means of distinguishing between the two variables themselves, one standing for the time on Sepulveda, the other for the hypothetical time on the freeway. We need this distinction in order to let the actual driving time (on Sepulveda) inform our assessment of the hypothetical driving time.

$$E(\text{hypothetical driving time} \mid do(\text{freeway}), \text{actual driving time} = 1 \text{ hour})$$

Counterfactual Notation

Fortunately, making this distinction is easy; we simply use different subscripts to label the two outcomes. We denote the freeway driving time by $Y_{X=1}$ (or Y_1, where context permits) and Sepulveda driving time by $Y_{X=0}$ (or Y_0). In our case, since Y_0 is the Y actually observed, the quantity we wish to estimate is

$$E(Y_{X=1}|X = 0, Y = Y_0 = 1) \tag{4.1}$$

The difficulty is that Y_{X=1} and X=0 are event occurring or different worlds.
This problem does not occur when we “intervene”. E(Y|do(X=x)) cannot capture this.

We cannot reduce the expression to a do-expression, which means that it cannot be estimated from interventional experiments. Indeed, a randomized controlled experiment on the two decision options will never get us the estimate we want.

Definition of Counterfactuals

- M is a structural causal modes (V,U,F) U, exogenous variables U (latent) for which we know the potential domain values.
- U=u implies a single entity in the population (e.g., a person, a situation in Nature)
- X(u) is a characteristic at world (e.g., salary(joe))
- The counterfactual sentence: Y would be y had X been x in situation U=u denoted $Y_x(u) = y$, where Y and X are any two variables in V.
- "had X been x" can be thought of as an instruction to make a minimal modification in the current model so as to establish the antecedent condition $X = x$,

The SCM Encodes Many Counterfactuals

Table 4.1 The values attained by $X(u), Y(u), Y_x(u)$, and $X_y(u)$ in the linear model of Eqs. (4.3) and (4.4)

u	$X(u)$	$Y(u)$	$Y_1(u)$	$Y_2(u)$	$Y_3(u)$	$X_1(u)$	$X_2(u)$	$X_3(u)$
1	1	2	2	3	4	1	1	1
2	2	4	3	4	5	2	2	2
3	3	6	4	5	6	3	3	3

Assume U= 1,2,3
Assume a-b=1

$$X = aU$$

$$Y = bX + U$$

We see that X is not affected counterfactually by Y

The Difference Between "Do" Operator and Counterfactuals

In this example we computed not merely the probability or expected value of Y under one intervention or another, but the actual value of Y under the hypothesized new condition $X = x$. For each situation $U = u$, we obtained a definite number, $Y_x(u)$, which stands for that hypothetical value of Y in that situation.

The do-operator, is only defined on probability distributions and, after deleting the factor $P(x_i|pa_i)$ always delivers probabilistic results such as $E[Y\ |do(x)]$.

the $do(x)$-operator captures the behavior of a population under intervention, whereas $Y_x(u)$ describes the behavior of a specific individual, $U = u$, under such interventions.

The Fundamental Law of Counterfactuals

Let M_x stand for the modified version of M, with the equation of X replaced by $X = x$. The formal definition of the counterfactual $Yx(u)$ reads

$$Y_x(u) = Y_{M_x}(u)$$

:

The counterfactual $Y_x(u)$ in model M is defined as the solution for Y in the

"surgically modified" submodel M_x.

In words: The counterfactual $Y_x(u)$ in model M is defined as the solution for Y in the "surgically modified" submodel M_x. Equation (4.5) is one of the most fundamental principles of causal inference. It allows us to take our scientific conception of reality, M, and use it to generate answers to an enormous number of hypothetical questions of the type "What would Y be had X been x?" The same definition is applicable when X and Y are sets of variables, if by M_x we mean a model where the equations of all members of X are replaced

The Fundamental Law of Counterfactuals

In general, counterfactuals obey the following *consistency rule*:

$$if\ X = x\ then\ Y_x = Y \tag{4.6}$$

If X is binary, then the consistency rule takes the convenient form:

$$Y = XY_1 + (1 - X)Y_0$$

which can be interpreted as follows: Y_1 is equal to the observed value of Y whenever X takes the value one. Symmetrically, Y_0 is equal to the observed value of Y whenever X is zero. All these constraints are automatically satisfied if we compute counterfactuals through Eq. (4.5).

From Population to Individual –Illustration in a Structural Equation Model (SEM)

$$X = U_X$$
$$H = a \cdot X + U_H$$
$$Y = b \cdot X + c \cdot H + U_Y$$
$$\sigma_{U_i U_j} = 0 \quad \text{for all } i, j \in \{X, H, Y\}$$

X = time in remedial program
H= the amount of homework
Y= student's score in exam

The value of each variable is the number of standard deviations above the mean the student falls
Students are assigned to the remedial sessions randomly.
Assume all U factors are independent and a = 0.5, b = 0.7, c = 0.4

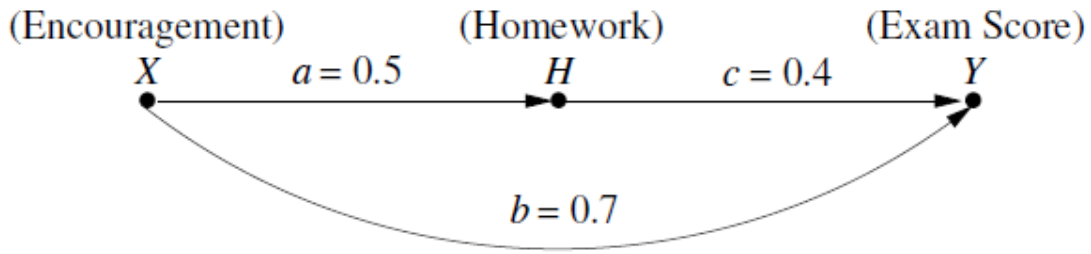


Figure 4.1: A model depicting the effect of Encouragement (X) on student's score

Assume Joe has X = 0.5,H = 1, and Y = 1.5.

What would Joe's score have been had he doubled his study time?

Next, we simulate the action of doubling Joe's study time by replacing the structural equation for H with the constant $H = 2$. The modified model is depicted in Figure 4.2. Finally, we compute the value of Y in our modified model using the updated U values, giving

$$\begin{aligned} Y_{H=2}(U_X = 0.5, U_H = 0.75, U_Y = 0.75) \\ &= 0.5 \cdot 0.7 + 2.0 \cdot 0.4 + 0.75 \\ &= 1.90 \end{aligned}$$

We thus conclude that Joe's score, had he doubled his homework, would have been 1.9 instead of 1.5. This, according to our convention, would mean an increase to 1.9 standard deviations above the mean, instead of the current 1.5.

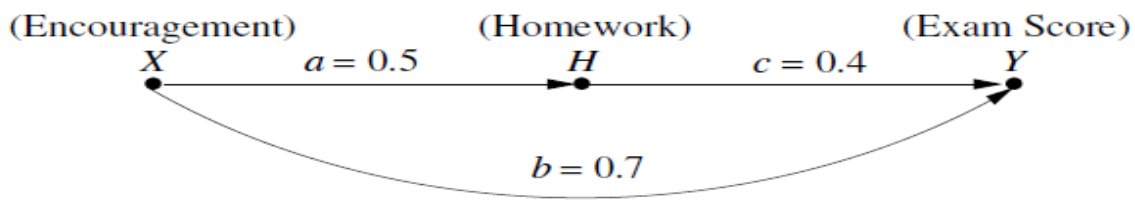


Figure 4.1: A model depicting the effect of Encouragement (X) on student's score

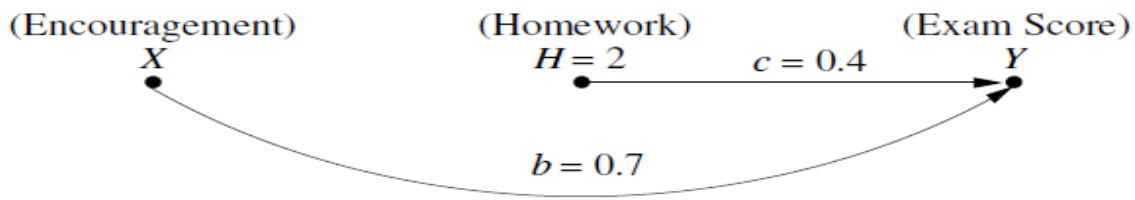


Figure 4.2: Answering a counterfactual question about a specific student's score, predicated on the assumption that homework would have increased to $H = 2$

Three Steps for Computing Deterministic Counterfactuals

There is a three-step process for computing any deterministic counterfactual:

- **Abduction**: Use evidence $E = e$ to determine the value of U.

- **Action**: Modify the model, M, by removing the structural equations for the variables in X and replacing them with the appropriate functions $X = x$, to get M_x.

- **Prediction**: Use the modified model, M_x, and the value of U, to compute the value of Y , the consequence of the counterfactual.

In temporal metaphors, Step (i) explains the past (U) in light of the current evidence e;
Step (ii) bends the course of history (minimally) to comply with the hypothetical antecedent $X = x$; finally, Step (iii) predicts the future (Y) based on our new understanding of the past and our newly established condition, $X = x$.
This process will solve any deterministic counterfactual, enabled in structural models

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.
  - The 3-steps
  - Do operators are limited
  - Expressing do by counterfactuals
  - The graphical representation of counterfactuals

Non-Deterministic Counterfactuals

- Counterfactuals can also be probabilistic, pertaining to a class of units within the population; for instance, in the after-school program example, we might want to know what would have happened if all students for whom $Y < 2$ had doubled their homework time.

- Nondeterminism enters causal models by assigning probabilities $P(U = u)$ over the exogenous variables U.

- The exogenous probability $P(U = u)$ induces a unique probability distribution on the endogenous variables V, $P(v)$, and we can compute not only the probability of any single counterfactual, $Y_x = y$, but also the joint distributions of all combinations of observed and counterfactual variables.

X = time in remedial program
H= the amount of homework
Y= student's score in exam

Non-deterministic 3-steps computing of counterfactuals

Given that we observe feature $E = e$ for a given individual, what would we expect the value of Y for that individual to be if X had been x?

The expectation is denoted $E[Y_{X=x} | E = e]$, where $E = e$ can conflict with the antecedent $X = x$.

Given a counterfactual $E[Y_{X=x} | E = e]$, the three-step process is:

(i) Abduction: Update $P(U)$ by the evidence to obtain $P(U | E = e)$.

(ii) **Action**: Modify the model, M, by removing the structural equations for the variables in X and replacing them with appropriate functions $X = x$, yielding M_x.

(iii) **Prediction:** Use M_x, and the updated probabilities $P(U | E = e)$, to compute the expectation of Y .

Revisiting earlier example; Adding P(U)

$$X = aU$$

$$Y = bX + U$$

$$P(U = 1) = \frac{1}{2},\ P(U = 2) = \frac{1}{3} \text{ and } P(U = 3) = \frac{1}{6}.$$

Table 4.1 The values attained by $X(u), Y(u), Y_x(u)$, and $X_y(u)$ in the linear model of Eqs. (4.3) and (4.4)

u	$X(u)$	$Y(u)$	$Y_1(u)$	$Y_2(u)$	$Y_3(u)$	$X_1(u)$	$X_2(u)$	$X_3(u)$
1	1	2	2	3	4	1	1	1
2	2	4	3	4	5	2	2	2
3	3	6	4	5	6	3	3	3

For instance, we can compute the proportion of units for which Y would be 3 had X been 2, or $Y_2(u) = 3$. This occurs only in the first row when $U = 1$, and therefore $P(Y_2 = 3) = 1/2$. Similarly:
$P(Y\_1 = 4) = 1/6$, $P(Y\_1 = 3) = 1/3$, $P(Y\_2 > 3) = 1/2$

We can compute joint probability of any combination

$$P(Y_2 > 3,\ Y_1 < 4) = \frac{1}{3}$$

$$P(Y_1 < 4,\ Y - X > 1) = \frac{1}{3}$$

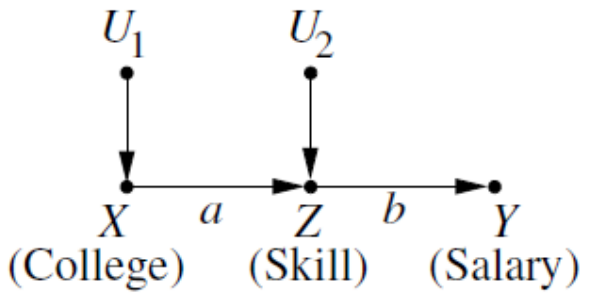
$$P(Y_1 < Y_2) = 1$$

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.
  - The 3-steps
  - Do operators are limited and Expressing do by counterfactuals
  - The graphical representation of counterfactuals

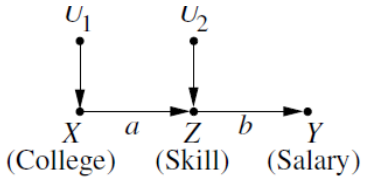
The Do operator is limited.

- Example model: $X = U_1 \quad Z = aX + U_2, \quad Y = bZ$
- X=1 has college education
- U_2 = professional experience
- Z = skill level
- Y = salary



The Do operator is limited.

$$X = U_1 \quad Z = aX + U_2, \quad Y = bZ$$



Let's compute $E[Y_{X=1}|Z = 1]$ = the expected salary of individuals with skill level $Z = 1$, had they received a college education.

- $E[Y\ |do(X = 1), Z = 1]$ will not work: The do-expression stands for the expected salary of individuals who all finished college and have since acquired skill level $Z = 1$. The salaries of these individuals, as the graph shows, depend only or their skill, and are not affected by whether they obtained the skill through college or work experience.
- Conditioning on $Z = 1$, in this case, cuts off the effect of the intervention that we're interested in.

In contrast, some of those who currently have $Z = 1$ might not have gone to college and would have attained higher skil (and salary) had they gotten college education. Their salaries are of great interest to us, but they are not included in the do-expression.

Thus, in general, the do-e $E[Y|do(X = 1), Z = 1] \neq E[Y_{X=1}|Z = 1]$ uestion:

$E[Y\ |do(X = 1), Z = 1] = E[Y\ |do(X = 0), Z = 1]$, but $E[Y_{X=1}|Z = 1]$ is not equal to $E[Y_{X=0}|Z = 1]$;

Counterfactual notations can capture the Do

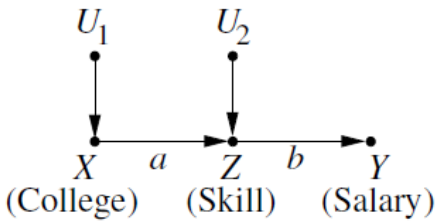
- Can counterfactual notation capture the postintervention, single-world expression $E[Y \mid do(X = 1), Z = 1]$?
- Yes! being more flexible, counterfactuals can capture both single-world and cross-world probabilities.
- The translation of $E[Y \mid do(X = 1), Z = 1]$ is $E[Y_{X=1} | Z_{X=1} = 1]$ Where the event $Z = 1$ is a postintervention.
- In general

$$P[Y = y | do(X = 1), Z = z] = \frac{P(Y = y, Z = z | do(X = 1))}{P(Z = z | do(X = 1))}$$

This shows explicitly how the dependence of Z on X should be treated. In the special case where Z is a preintervention variable, as age was in our discussion of conditional interventions (Section 3.5) we have $Z_{X=1} = Z$, and we need not distinguish between the two. The inequality in (4.8) then turns into an equality.

$$E[Y | do(X = 1), Z = 1] = E[Y_{X=1} | Z = 1]$$

Example of expectation of counterfactuals



The table depicts the counterfactuals associated with the model for X. We replace the equation $X = u$ with the appropriate constant (zero or one) and solving for Y and Z.

Table 4.2 The values attained by $X(u), Y(u), Z(u), Y_0(u), Y_1(u), Z_0(u)$, and $Z_1(u)$ in the model of Eq. (4.7)

		X=u₁	$Z = aX + u_2$	$Y = bZ$				
u_1	u_2	$X(u)$	$Z(u)$	$Y(u)$	$Y_0(u)$	$Y_1(u)$	$Z_0(u)$	$Z_1(u)$
0	0	0	0	0	0	ab	0	a
0	1	0	1	b	b	$(a+1)b$	1	$a+1$
1	0	1	a	ab	0	ab	0	a
1	1	1	$a+1$	$(a+1)b$	b	$(a+1)b$	1	$a+1$

Using the table we can show:

$$E[Y_1|Z=1] = (a+1)b \tag{4.9}$$

$$E[Y_0|Z=1] = b \tag{4.10}$$

$$E[Y|do(X=1), Z=1] = b \tag{4.11}$$

$$E[Y|do(X=0), Z=1] = b \tag{4.12}$$

Despite the fact that Z separates X from Y in the graph we find that X has an effect on Y for those units falling under Z = 1: $E[Y_1 - Y_0|Z = 1] = ab \neq 0$

While the salary of those who have acquired skill level Z = 1 depends only on their skill, not on X, the salary of those who are currently at Z = 1 would have been different had they had a different past.

Example of expectation of counterfactuals

The table depicts the counterfactuals associated with the model for X. We replace the equation $X = u$ with the appropriate constant (zero or one) and solving for Y and Z.

Table 4.2 The values attained by $X(u), Y(u), Z(u), Y_0(u), Y_1(u), Z_0(u)$, and $Z_1(u)$ in the model of Eq. (4.7)

		X=u₁	$Z = aX + u_2$	$Y = bZ$				
u_1	u_2	$X(u)$	$Z(u)$	$Y(u)$	$Y_0(u)$	$Y_1(u)$	$Z_0(u)$	$Z_1(u)$
0	0	0	0	0	0	ab	0	a
0	1	0	1	b	b	$(a+1)b$	1	$a+1$
1	0	1	a	ab	0	ab	0	a
1	1	1	$a+1$	$(a+1)b$	b	$(a+1)b$	1	$a+1$

Using the table we can show:

$$E[Y_1|Z=1] = (a+1)b \tag{4.9}$$

$$E[Y_0|Z=1] = b \tag{4.10}$$

$$E[Y|do(X=1), Z=1] = b \tag{4.11}$$

$$E[Y|do(X=0), Z=1] = b \tag{4.12}$$

The probabilities play a role, however, if we assume a = 1 in the model, since Z = 1 can now occur under two conditions: (u_1 = 0, u_2 = 1) and (u_1 = 1, u_2 = 0). The first occurs with probability $P(u\_1 = 0)P(u\_2 = 1)$, and the second with probability $P(u\_1 = 1)P(u\_2 = 0)$.

$$E[Y_{X=1}|Z=1] = b\left(1 + \frac{P(u_1=0)P(u_2=1)}{P(u_1=0)P(u_2=1) + P(u_1=1)P(u_2=0)}\right)$$

$$E[Y_{X=0}|Z=1] = b\left(\frac{P(u_1=0)P(u_2=0)}{P(u_1=0)P(u_2=1) + P(u_1=1)P(u_2=0)}\right)$$

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.
  - The 3-steps
  - Do operators are limited and Expressing do by counterfactuals
  - The graphical representation of counterfactuals

The Graphical Representation of Counterfactuals

Can we see counterfactual in our causal model's graph?
Yes. Based on the fundamental law of counterfactuals

$$Y_x(u) = Y_{M_x}(u)$$

If we modify model M to obtain the submodel M_x, then the outcome variable Y in the modified model is the counterfactual Y_x of the original model. Since modification calls for removing all arrows entering the variable X, the node associated with the Y variable serves as a surrogate for Y_x

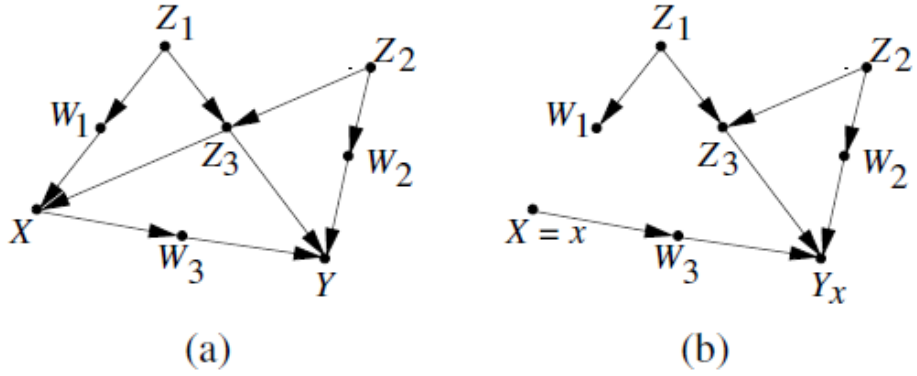


Figure 4.4: Illustrating the graphical reading of counterfactuals. (a) The original model. (b) The modified model M_x in which the node labeled Y_x represents the potential outcome Y predicated on $X = x$

The Graphical Representation of Counterfactuals

- When we ask about the statistical properties of Y_x, we need to examine what would cause Y_x to vary. Statistical variations of Yx are therefore governed by all exogenous variables capable of influencing Y when X is held constant at X=x, that is, when the arrows entering X are removed.
- The set of variables capable of transmitting variations to Y are the parents of Y , (observed and unobserved) as well as parents of nodes on the pathways between X and Y.
- For example, in the figure these parents are {Z_3,W_2,U_3,U_Y }, (U_Y and U_3, the error terms of Y and $W3$, are not shown in the diagram). Any set of variables that blocks a path to these parents also blocks that path to Yx, yield a conditional independence for Y_x. In particular, if we have a set Z that satisfies the backdoor criterion in M, that set also blocks all paths between X and those parents, and consequently, it renders X and Y_x independent for every $Z = z$.

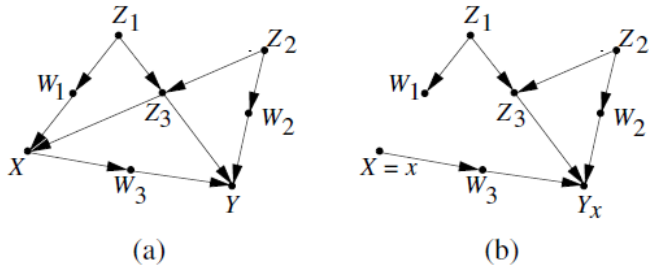


Figure 4.4: Illustrating the graphical reading of counterfactuals. (a) The original model. (b) The modified model M_x in which the node labeled Y_x represents the potential outcome Y predicated on $X = x$

The Graphical Representation of Counterfactuals

Theorem 4.3.1 (Counterfactual Interpretation of Backdoor) *If a set Z of variables satisfies the backdoor condition relative to (X, Y) then, for all x, the counterfactual Y_x is conditionally independent of X given Z*

$$P(Y_x|X, Z) = P(Y_x|Z) \tag{4.15}$$

Theorem 4.3.1 has far-reaching consequences when it comes to estimating the probabilities of counterfactuals from observational studies. In particular, it implies that $P(Y_x = y)$is identifiable by the adjustment formula of Eq. (3.5). To prove this, we conditionalize on Z (as in Eq. (1.9)) and write

$$\begin{aligned} P(Y_x = y) &= \sum_z P(Y_x = y|Z = z)P(z) \\ &= \sum_z P(Y_x = y|Z = z, X = x)P(z) \\ &= \sum_z P(Y = y|Z = z, X = x)P(z) \end{aligned} \tag{4.16}$$

The second line was licensed by Theorem 4.3.1, whereas the third line was licensed by the consistency rule (4.6).

Indeed this is just P(Y|do(x))

The Graphical Representation of Counterfactuals

Having a graphical representation for counterfactuals, we can now explain graphically why a stronger education (X) would have had an effect on the salary (Y) of people who are currently at skill level Z = z, despite the fact that, according to the model, salary is determined by skill only.

Formally, to determine if the effect of education on salary (Y_x) is statistically independent of the level of education, we need to locate Y_x in the graph and see if it is d-separated from X given Z.

We see in the figure that Y_x can be identified with U_2, the only parent of nodes on the causal path from X to Y (and therefore, the only variable that produces variations in Y_x while X is held constant). Indeed Z acts as a collider between X and U_2, So, X and U2 are not d-separated given Z.

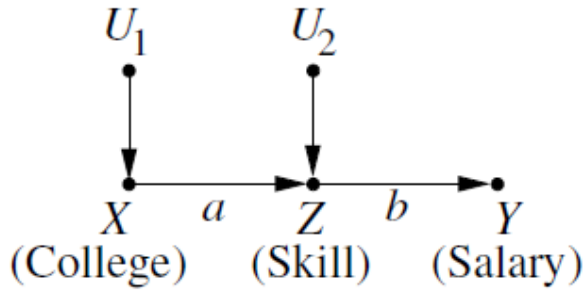
We conclude again:

$$E[Yx|X,Z] \neq E[Yx|Z]$$

despite the fact that

$$E[Y|X,Z] = E[Y|Z]$$

In (and similarly X and Y_x) are not d-separated given Z. :



Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.
  - The 3-steps
  - Do operators are limited and Expressing do by counterfactuals
  - The graphical representation of counterfactuals
  - Counterfactuals in Experimental Settings
  - Practical use of counterfactuals

Counterfactual in Experimental Settings

So we can answer counterfactual question from a fully specified structural model.
But what to do when a model is not available, and we have only a finite sample of observed individuals?

Let's consider again the "encouragement design" model in which we analyzed the behavior of an individual named Joe. Assume that the experimenter observes a set of 10 individuals, with Joe being participant 1. Each, characterized by a distinct vector $U_i = (U_X, U_H, U_Y)$, as shown in the first 3 columns

Table 4.3 Potential and observed outcomes predicted by the structural model of Figure 4.1 units were selected at random, with each U_i uniformly distributed over $[0, 1]$

	Participant characteristics			Observed behavior			Predicted potential outcomes				
Participant	U_X	U_H	U_Y	X	Y	H	Y_0	Y_1	H_0	H_1	$Y_{00}\ldots$
1	0.5	0.75	0.75	0.5	1.50	1.0	1.05	1.95	0.75	1.25	0.75
2	0.3	0.1	0.4	0.3	0.71	0.25	0.44	1.34	0.1	0.6	0.4
3	0.5	0.9	0.2	0.5	1.01	1.15	0.56	1.46	0.9	1.4	0.2
4	0.6	0.5	0.3	0.6	1.04	0.8	0.50	1.40	0.5	1.0	0.3
5	0.5	0.8	0.9	0.5	1.67	1.05	1.22	2.12	0.8	1.3	0.9
6	0.7	0.9	0.3	0.7	1.29	1.25	0.66	1.56	0.9	1.4	0.3
7	0.2	0.3	0.8	0.2	1.10	0.4	0.92	1.82	0.3	0.8	0.8
8	0.4	0.6	0.2	0.4	0.80	0.8	0.44	1.34	0.6	1.1	0.2
9	0.6	0.4	0.3	0.6	1.00	0.7	0.46	1.36	0.4	0.9	0.3
10	0.3	0.8	0.3	0.3	0.89	0.95	0.62	1.52	0.8	1.3	0.3

$$X = U_X$$
$$H = a \cdot X + U_H$$
$$Y = b \cdot X + c \cdot H + U_Y$$
$$\sigma_{U_i U_j} = 0 \quad \text{for all } i, j \in \{X, H, Y\}$$

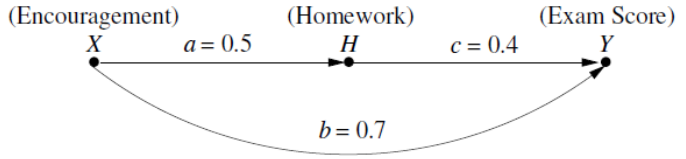


Figure 4.1: A model depicting the effect of Encouragement (X) on student's score

We use the model to fill the data from the U variables.

First item: Y_0 = 0.4 times 1 + 0.75= 1.05

Counterfactual in Experimental Settings

From this synthetic population, one can estimate the probability of every counterfactual query on variables X, Y,Z, assuming, of course, that we are in possession of all entries of the table.

Clearly the table is not available to us in either observational or experimental studies. This was deduced from the fully specified model from which we could infer the defining characteristics $\{U_X, U_H, U_Y\}$ of each participant, given the observations $\{X, H, Y\}$.

Without a parametric model, the observed behavior $\{X, H, Y\}$ tells very little of the potential outcome Y_1 or Y_0.

We know only the consistency rule: that $Y1$ must be equal to Y in case X = 1, and Y_0 must be equal to Y in case X = 0.

Yet we can say much at the population level estimating their probabilities or expectation. We can use
The adjustment formula of (4.16), where we were able to compute $E(Y\_1 - Y\_0)$ using the graph alone as we will see next.

Using Experimental Data

Assume that we have no information whatsoever about the underlying model. All we have are measurements on Y taken in an experimental study in which X is randomized over two levels, $X = 0$ and $X = 1$.

Table 4.4 Potential and observed outcomes in a randomized clinical trial with X randomized over $X = 0$ and $X = 1$

	Predicted potential outcomes		Observed outcomes	
Participant	Y_0	Y_1	Y_0	Y_1
1	1.05	1.95	1.05	■
2	0.44	1.34	■	1.34
3	0.56	1.46	■	1.46
4	0.50	1.40	■	1.40
5	1.22	2.12	1.22	■
6	0.66	1.56	0.66	■
7	0.92	1.82	■	1.82
8	0.44	1.34	0.44	■
9	0.46	1.36	■	1.36
10	0.62	1.52	0.62	■
	True average treatment effect: 0.90		Study average treatment effect: 0.68	

Randomized: participants 1, 5, 6, 8 and 10 assigned to X = 0, and the rest to X = 1. The first two columns give the true potential outcomes (taken from Table 4.3) while the last two columns describe the information available to the experimenter.

The difference between the observed means in the treatment and control groups will converge to the difference of the population averages, $E(Y\_1 - Y\_0) = 0.9$ due to randomization.

Under randomization, the adjustment formula (4.16) is applicable with Z = {empty}, yielding $E[Y\_x] = E[Y \mid X = x]$.
So, Table 4.4 helps us understand what is actually computed when we take sample averages in experimental settings and how those averages are related to the underlying counterfactuals, Y_1 and Y_0.

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.
  - The 3-steps
  - Do operators are limited and Expressing do by counterfactuals
  - The graphical representation of counterfactuals
  - Counterfactuals in Experimental Settings
  - Practical use of counterfactuals

Practical Uses of Counterfactuals

- Recruitment program
- Additive Interventions
- Personal decision making
- Sex discrimination in hiring
- Mediation and path disabling

Recruitment ProgramJ job training helps?

Example 4.4.1 A government is funding a job training program aimed at getting jobless people back into the workforce. A pilot randomized experiment shows that the program is effective; a higher percentage of people were hired among those who finished the program than among those who did not go through the program. As a result, the program is approved, and a recruitment effort is launched to encourage enrollment among the unemployed, by offering the job training program to any unemployed person who elects to enroll.

Enrollment is successful, and the hiring rate among the program's graduates turns out even higher than in the randomized pilot study. Success!!!

Critics say: Those who self-enroll, may be more intelligent, more resourceful, and more socially connected than the eligible who did not enroll and are more likely to have found a job regardless of the training.

The critics claim that what we need to estimate is the differential benefit of the program on those enrolled: the extent to which hiring rate has increased among the enrolled, compared to what it would have been had they not been trained.

Personal Decision Making

Example 4.4.3 Ms. Jones, a cancer patient, is facing a tough decision between two possible treatments: (i) lumpectomy alone, or (ii) lumpectomy plus irradiation. In consultation with her oncologist, she decides on (ii). Ten years later, Ms. Jones is alive, and the tumor has not recurred. She speculates: Do I owe my life to irradiation?
Mrs. Smith, on the other hand, had a lumpectomy alone, and her tumor recurred after a year. And she is regretting: I should have gone through irradiation.
Can these speculations ever be substantiated from statistical data? Moreover, what good would it do to confirm Ms. Jones's triumph or Mrs. Smith's regret?

Sex Discrimination in Hiring

Example 4.4.4 Mary files a law suit against the New York-based XYZ International, alleging discriminatory hiring practices. According to her, she has applied for a job with XYZ International, and she has all the credentials for the job, yet she was not hired, allegedly because she mentioned, during the course of her interview, that she is gay. Moreover, she claims, the hiring record of XYZ International shows consistent preferences for straight employees. Does she have a case? Can hiring records prove whether XYZ International was discriminating when declining her job application?
At the time of writing, U.S. law doesn't specifically prohibit employment discrimination on

Mediation and Path-disabling

Example 4.4.5 A policy maker wishes to assess the extent to which gender disparity in hiring can be reduced by making hiring decisions gender-blind, rather than eliminating gender inequality in education or job training. The former concerns the “direct effect” of gender on hiring, whereas the latter concerns the “indirect effect,” or the effect mediated via job qualification.

Outline

- Overview of last class:
  - Counterfactuals
  - Defining and computing counterfactuals.
  - The tree steps of computing counterfactuals (the deterministic case)
- Nondeterministic counterfactuals.
  - The 3-steps
  - Do operators are limited and Expressing do by counterfactuals
  - The graphical representation of counterfactuals
  - Counterfactuals in Experimental Settings
  - Practical use of counterfactuals