DUKE COMPUTER SCIENCE

CompSci 590.01

Causal Inference in Data Analysis with Applications to Fairness and Explanations

Lecture 4:
Front-door Criterion
&
Counterfactuals

Sudeepa Roy

Reading

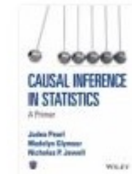
10. Causal inference in statistics : a primer

Judea Pearl, Madelyn Glymour, Nicholas Jewell.

Pearl, Judea, author

Chichester, West Sussex, UK : John Wiley & Sons Ltd, 2016. / Book / Online

View Online



- Primer book - Chapter 3.4, Chapter 4

Acknowledgement (big thanks!):

Many slides are by Amir-Hossein Karimi that have been modified here.

Announcements

- Possible paper presentation topics are posted on the google doc
- Some may be more suitable to some students than the others based on your background and interests
  - AI, ML, Database, Data mining, Theory, Stat, Systems, Applications, something else
- Happy to discuss before you finalize
- Sudeepa’s office hour: Tuesdays after class 3-4 pm, LSRC D325 (you can walk with me!)

Announcements

Timeline:

- After today's class: Star talking to your fellow students – in person or on Ed
- Presentation topic & date due: Thursday 2/2
  - Ping Sudeepa on Ed copying your teammate once you have chosen one or more topics, so that we can decide on one – can be before 2/2
  - Please talk to Sudeepa if you need help with choosing a topic
- Initial project ideas & teammates' names due: Tuesday 2/7
  - Please share on Google doc or Overleaf (latex)
- Project proposal due: Tuesday 2/14

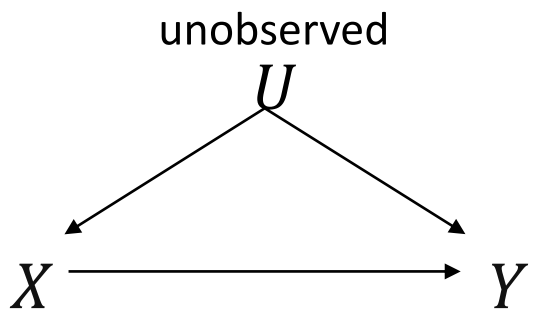
Introduction!

- Please see the spreadsheet link on Ed for presentation/project sign ups and add this info there too
- Please tell us
  - Your name
  - Undergraduate / MS/ PhD
  - Department / Major
  - Your interests: AI, ML, Database, Data mining, Theory, Stat, Systems, Applications, something else

Front-door criterion

Recall: Backdoor is only “sufficient” not “necessary” for adjustment –sometimes it may not work at all

Front-door Criterion



No backdoor criterion satisfied because U is unobservable, and open back path.

Causal effect not identifiable

One cannot measure what correlation between X & Y is due to U
And what is causative

Evaluate the effect of X on Y:

X: Smoking

Y: Lung cancer

U: “carcinogenic genotype”

1970: Tobacco industry managed to prevent antismoking legislation promoting the theory that U is the reason that also induces an inborn craving for nicotine

Front-door Criterion

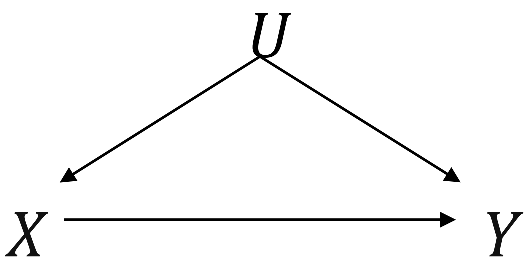
Evaluate the effect of X on Y:

X: Smoking

Y: Lung cancer

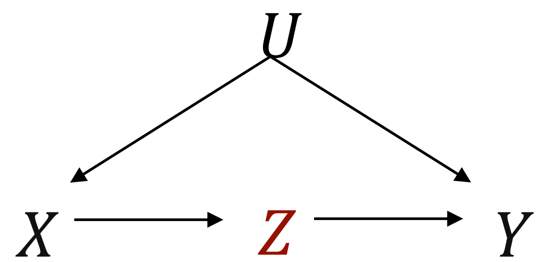
U: “carcinogenic genotype”

Z: Amount of “tar deposits” in patients’ lungs



No backdoor criterion satisfied because U is unobservable, and open back path.

Causal effect not identifiable.



No backdoor criterion satisfied because there is still an open back path.

Causal effect identifiable through two consecutive applications of the backdoor criterion.

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar		No tar		All subjects	
	400		400		800	
	Smokers	Nonsmokers	Smokers	Nonsmokers	Smokers	Nonsmokers
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323	1	18	38	341	39
	(85%)	(5%)	(90%)	(10%)	(85%)	(9.75%)
Cancer	57	19	2	342	59	361
	(15%)	(95%)	(10%)	(90%)	(15%)	(90.25%)

Disclaimer: Hypothetical data contrary to popular observations to illustrate the point

Beneficial effect of smoking ?!?

15% smokers have cancer vs. 90.25% non-smokers
Within each group – tar and no tar, smokers have a low percentage of cancer than non-smoker

Tables from the PRIMER book

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers		Nonsmokers		All subjects	
	400		400		800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323	18	1	38	324	56
	(85%)	(90%)	(5%)	(10%)	(81%)	(19%)
Cancer	57	2	19	342	76	344
	(15%)	(10%)	(95%)	(90%)	(19%)	(81%)

Disclaimer: Hypothetical data contrary to popular observations to illustrate the point

Beneficial effect of smoking ?!? – May be not! Smoking increases risk of lung cancer

- Smokers have higher chance of building up tar deposits (95% vs. 5%)
- See the (harmful) effect of tar-deposits for smoker and non-smoker groups
  - Smokers: cancer rate increased from 10% to 15%
  - Non-smokers: cancer rate increased from 90% to 95%
- Whether or not one has natural craving for nicotine, avoid tar deposit, and therefore avoid smoking!

Tables from the PRIMER book

Front-door Criterion

Evaluate the effect of X on Y:

X: Smoking

Y: Lung cancer

U: “carcinogenic genotype”

Z: Amount of “tar deposits” in patients’ lungs

1. Measure the effect of X on Z (no backdoor path)

$P(Z = z|do(X = x)) = P(Z = z|X = x).$

2. Measure the effect of Z on Y (backdoor blocked by conditioning on X)

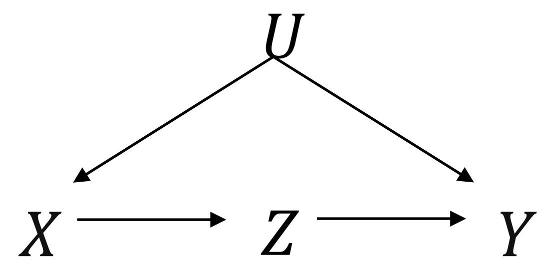
$$P(Y = y|do(Z = z)) = \sum_{x'} P(Y = y|Z = z, X = x')P(X = x').$$

3. Chain together the two partial effects

$P(Y = y|do(X = x)) = \sum_{z} P(Y = y|do(Z = z))P(Z = z|do(X = x)$

(If nature chooses to assign Z = z,
prob. of Y = P(Y = y | do(Z) = z).
The probability of nature doing that
If we choose to set X = x is P(Z = z | do(X) = x)
Then sum over all states z

$= \sum_{z}\sum_{x'} P(Y = y|Z = z, X = x')P(X = x')P(Z = z|X = x)$



No backdoor criterion satisfied beca
there is still an open back path.

Causal effect identifiable through two consecutive applications of the backdoor criterion.

Front-door Criterion

A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X,Y) if:

1. Z intercepts all directed paths from X to Y.
2. There is no backdoor path from X to Z.
3. All backdoor paths from Z to Y are blocked by X.

$$P(Y = y|do(X = x)) = \sum_{z}\sum_{x'} P(Y = y|X = x', Z = z)P(X = x')P(Z = z|X = x)$$

if Z satisfies the front-door criterion relative to (X,Y) and P(x, z)>0.

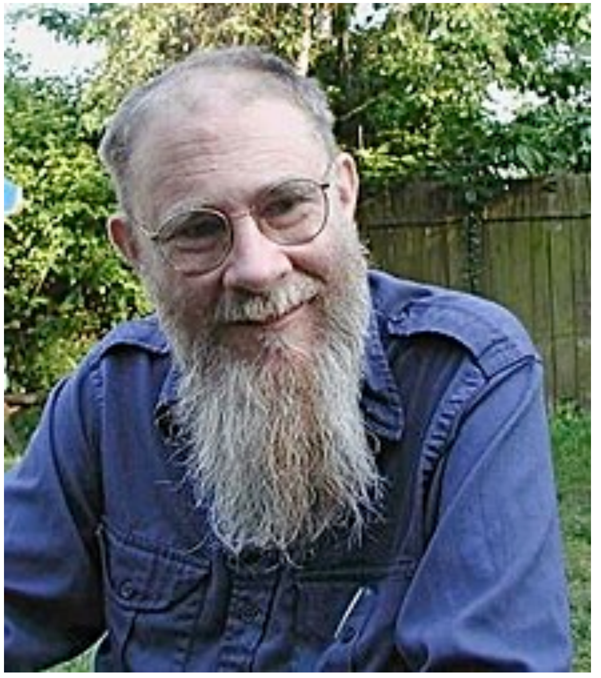
Note: still a sufficient condition – some paths that are unblocked in #2 and #3 can be blocked
By other variables == "do-calculus" is the exact method
– not covered in lecture – possible presentation topic!

Counterfactuals

Lewis's Counterfactual Analysis

"We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects—some of them, at least, and usually all—would have been absent as well."

David Lewis, Journal of Philosophy (1973)



Pearl’s Ladder of Causation

$$P(Y_{X=x'} = y | X = x, Y = y)$$

$$P(Y = y | \mathrm{do}(X = x))$$

$$P(Y = y | X = x)$$

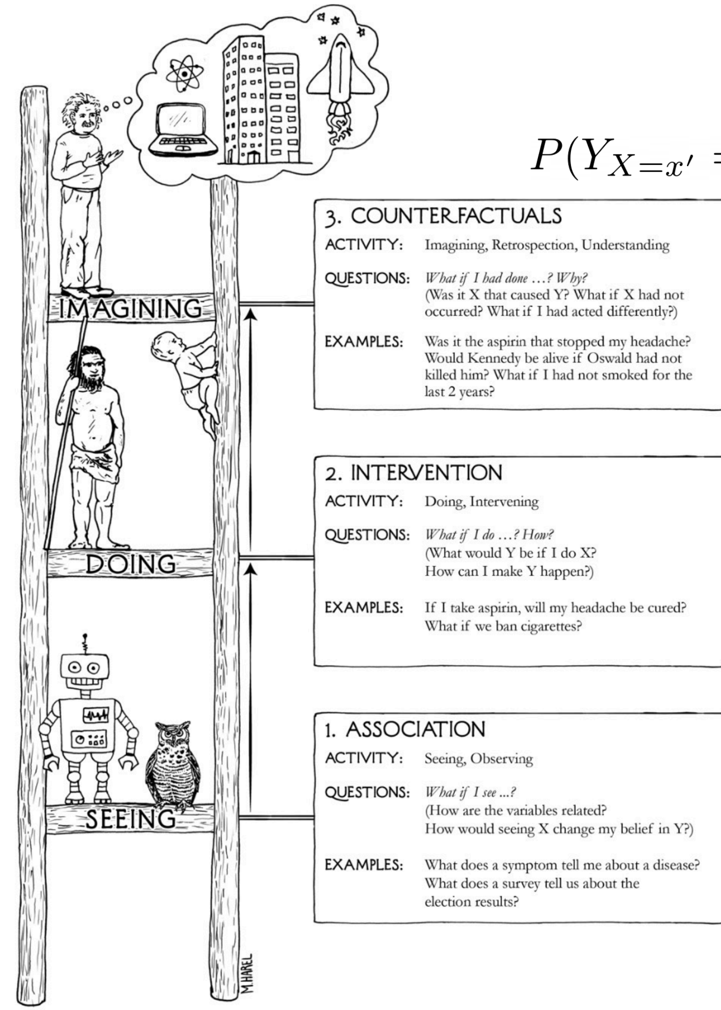


FIGURE 1.2. The Ladder of Causation, with representative organisms at each level. Most animals, as well as present-day learning machines, are on the first rung, learning from association. Tool users, such as early humans, are on the second rung if they act by planning and not merely by imitation. We can also use experiments to learn the effects of interventions, and presumably this is how babies acquire much of their causal knowledge. Counterfactual learners, on the top rung, can imagine worlds that do not exist and infer reasons for observed phenomena. (*Source:* Drawing by Maayan Harel.)

Figure from “Pearl - The book of Why”

Freeway or not?

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."

== if I had taken the freeway I would have gotten home earlier
== E(driving time | do(freeway), driving time = 1 hour)??

Oops – not quite – talking about two worlds – actual and hypothetical driving time

With do operator we can only express:
E(driving time | do(freeway) vs. E(driving time | do(Sepulveda)

Quote from the PRIMER book, figure from the internet

Counterfactuals

- "If X was set to x, what would have been the value of Y"

  $Y_{X=x}$ (or Y_x)
- An "if" statement where the if-portion is not true (counterfactual or hypothetical or retrospective estimate)
- $E(Y \mid do(X = x))$: predicts the effect of intervention

- $E(Y \mid do(X = x)) = E(Y_{X=x})$
- $E(Y \mid do(X = x), Z = z) = E(Y_{X=x} \mid Z = z)$

When all variables are from the same world: the modified distribution created by do(X = x)

$E(Y_{X=1} \mid Y_{X=0}=y')$: About two different worlds – cannot be expressed as do-operator or intervention, and cannot be expressed from experiments alone. need counterfactuals & structural equations

Recap: Structural, Graphical, Probabilistic Causal Model

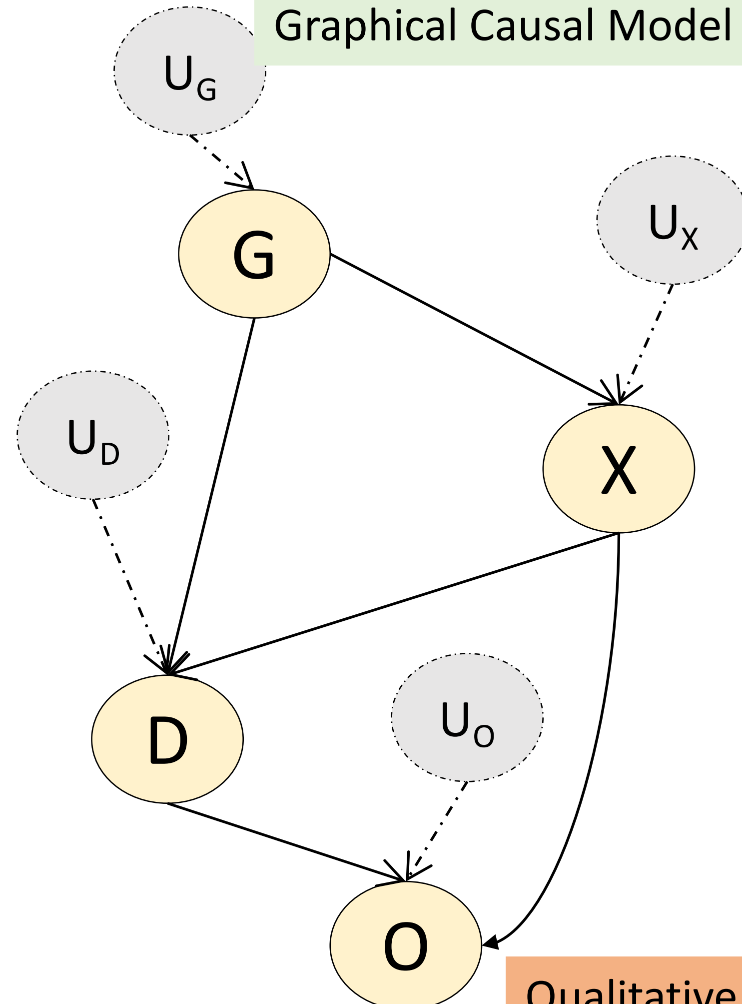
- $M = \langle U, V, F \rangle$
- Endogenous (observable) variables $V = \{G, X, D, O\}$
- Exogenous (noise) variables $U = \{U_G, U_X, U_D, U_o\}$
- Structural equations F:

$$\{G = F_G(U_G),$$
$$X = F_X(U_X, G),$$
$$D = F_D(U_D, G, X),$$
$$O = F_O(U_O, X, D)\}$$

Can be linear, exp, …

Quantitative

Corresponding Graphical Causal Model



Qualitative

Probabilistic == Distribution on U-s

Counterfactuals allow us to do individual analysis

Interventions

$$\mathbb{E}(\ Y_{X=x}\ ) = \mathbb{E}(\ Y \mid \mathrm{do}(X = x)\ )$$

Counterfactuals

$$Y_{X=x}(U = u)$$

Give values to U == Fully specified Deterministic model

- $M = \langle U, V, F \rangle$
- Endogenous (observable) variables $V = \{G, X, D, O\}$
- Exogenous (noise) variables $U = \{U_G, U_X, U_D, U_o\}$
- Structural equations F:

$$\{G = F_G(U_G),$$
$$X = F_x(U_x, G),$$
$$D = F_D(U_D, G, X),$$
$$O = F_O(U_O, X, D)\}$$

Every assignment U = u to the exogenous variables, uniquely determines the values of all endogenous variables in V, corresponding to a single member of, or "unit" in a population, or to a "situation" in nature.
E.g., an individual, agriculture land, etc.

For example, if U = u stands for the defining characteristics of an individual named Joe, and X stands for a variable named "salary," then X(u) stands for Joe's salary.

Counterfactual == Minimal changes in model

The U values are invariant to hypothetical actions

Y would be y had X been x in Situations U = u

$Y_x(u) = y$

- "Had X been x" == Make a minimal change to the model to establish X = x
- i.e., Replace X with the constant x (like do(X = x))

Example

Given model

$X = aU$

$Y = bX + U$

$Y_x(u) = y$ "Had X been x"

$X = x$

$Y = bX + U$

Substitute $U = u$ and solve for Y

$Y_x(u) = bx + u$

Suppose U takes three values 1, 2, 3
And $a = b = 1$

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

Table from Pearl Book

The Fundamental Law of Counterfactuals

Generalize the concept of counterfactuals to any structural model M

Consider any arbitrary two variables X and Y, not necessarily connected by a single equation.

Let M_x stand for the modified version of M, with the equation of X replaced by X = x. Then

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

Consistency rule: $\text{if } X = x \implies X(U = u) = x \implies Y_{X=x} = Y(U = u) = y$

For binary X: $Y = XY_1 + (1 - X)Y_0$

Example: From population data to individual behavior

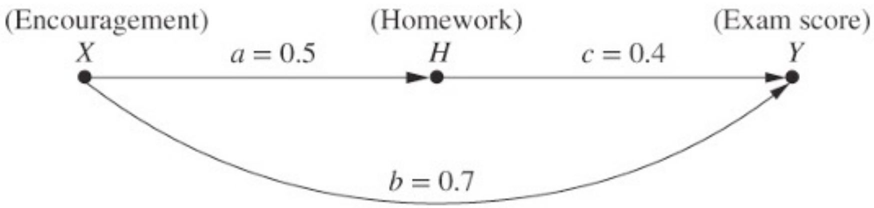


Figure 4.1 A model depicting the effect of Encouragement (X) on student's score

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

$$a = 0.5, \quad b = 0.7, \quad c = 0.4$$

$$X = 0.5$$
$$H = 1$$
$$Y = 1.5$$

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

Figure and example from PRIMER book

Example: From population data to individual behavior

Step 1 (abduction)

(Use evidence to find U variables)

$$U_X = 0.5,$$
$$U_H = 1 - 0.5 \cdot 0.5 = 0.75, \text{ and}$$
$$U_Y = 1.5 - 0.7 \cdot 0.5 - 0.4 \cdot 1 = 0.75$$

Step 2 (action)

(simulate Joe's study by replacing H's equation by H = 2)

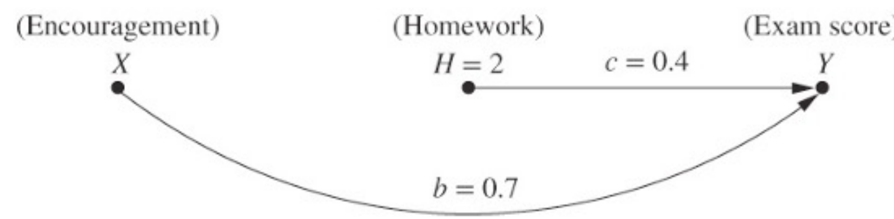


Figure 4.2 Answering a counterfactual question about a specific student's score, predicated on the assumption that homework would have increased to $H = 2$

Step 3 (prediction)

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

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

Counterfactual answer: Joe's score, had he doubled his homework, would have been 1.9 instead of 1.5.

The three steps in computing counterfactuals

	Deterministic $Y_x(u)$	**Probabilistic** $\mathbb{E}(\ Y_x \mid E = e\ )$
Step 1 (abduction)	Use evidence E = e to determine the value of U	Use evidence E = e to update P(U) as P(U | E = e)
Step 2 (action)	Modify the model, M, replacing the structural equation for X, to obtain the modified model, M_x	Modify the model, M, replacing the structural equation for X, to obtain the modified model, M_x
Step 3 (prediction)	Using M_x and U, compute the value of Y	Using M_x and P(U | E = e), compute the expectation of Y

Personalized Decision Making With Counterfactuals

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?

Y = Remission (tumor did not recur)
X – Decision to undergo irradiation

Probabilities of Causation

Probability of Necessity

$$PN = P(Y_0 = 0 | X = 1, Y = 1)$$

Remission would not have occurred
Had Ms. Jones not gone through irradiation

Personalized Decision Making With Counterfactuals

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?

Y = Remission (tumor did not recur)
X – Decision to undergo irradiation

Probabilities of Causation

Probability of Sufficiency

$$PS = P(Y_1 = 1 | X = 0, Y = 0)$$

Remission would have occurred
Had Mrs. Smith gone through irradiation

Personalized Decision Making With Counterfactuals

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?

Probability of Sufficiency

$$PS = P(Y_1 = 1 | X = 0, Y = 0)$$

Probability of Necessity

$$PN = P(Y_0 = 0 | X = 1, Y = 1)$$

In general, not estimable from observed or experimental data
Estimable under certain conditions when both observational and experimental data are available