Algorithms for Reasoning with graphical models

Slides Set 12 (part a):

Causal Graphical Models

Rina Dechter

Causal Inference in Statistics, A primer, J. Pearl, M Glymur and N. Jewell

“The book of Why” Pearl

https://www.nytimes.com/2018/06/01/business/dealbook/review-the-book-of-why-examines-the-science-of-cause-and-effect.html

http://bayes.cs.ucla.edu/WHY/

“The book of Why” Pearl

http://bayes.cs.ucla.edu/WHY/

Dog owners are happier

https://www.washingtonpost.com/business/2019/04/05/dog-owners-are-much-happier-than-cat-owners-survey-finds/?utm_term=.db698fed4acb

The science of cause and effect (quotes)

- Causal calculus
- Causal models are all about alternatives, and alternative reality. It is no accident that we developed the ability to think this way, because *Homo sapiens* is a creature of change.

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 advertisement. 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 such questions about actions not taken before. Moreover, most learning machine today do not utilize a representation from which such questions can be answered" (Pearl, position paper, 2016)

Chapter 1,
Preliminaries: Statistical and Causal Models.

- Why study causation? (sec 1.1).
  - To be able to asses the effect of actions on things of interest
  - Examples: The impact of smoking on cancer, the impact of learning on salary, the impact of selecting a president on human rights and well being, war/ peace. Dogs make people happy (NYT 2019)
  - Is causal inference part of statistics?
  - Causation is an addition to statistics and not part of statistics.
  - The language of statistics is not sufficient to talk about the above queries.
  - See The Simpson Paradox
- Simpson Paradox (sec 1.2)
- Probability and Statistics (sec 1.2)
- Graphs (sec 1.4)
- Structural Causal Models (sec 1.5)

The Simpson Paradox

- It refers to data in which a statistical association that holds for an entire population is reversed in every subpopulation.
- (Simpson 1951) a group of sick patients are given the option to try a new drug. Among those who took the drug, a lower percentage recover than among those who did not. However, when we partition by gender, we see that more men taking the drug recover than do men not taking the drug, and more women taking the drug recover than do women not taking the drug! In other words, the drug appears to help men and help women, but hurt the general population.
- Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. 350 patients chose to take the drug and 350 patients did not. We got:

Table 1.1 Results of a study into a new drug, with gender taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

The Simpson Paradox

- Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. 350 patients chose to take the drug and 350 patients did not. We got:

Table 1.1 Results of a study into a new drug, with gender taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

- The data says that if we know the gender of the patient we can prescribe the drug, but if not we should not…. Which is ridiculous.
- So, given the results of the study, should the doctor prescribe the drug for a man? For a woman? Or when gender is unknown?
- The answer cannot be found in the data!! We need to know the story behind the data- the causal mechanism that lead to, or generated the results we see.

The Simpson Paradox

- Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. 350 patients chose to take the drug and 350 patients did not. We got:

Table 1.1 Results of a study into a new drug, with gender taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

- Suppose we know that estrogen has negative recovery on Women, regardless of drugs. Also woman are more likely to take the drug
- So, being a woman is a common cause for both drug taking and failure to recover. So… we should consult the segregated data (not to involve the estrogen impact). We need to control for gender.

The Simpson Paradox

- The same phenomenon with continuous variables. Example: Impact of exercise on Cholesterol for different age groups:

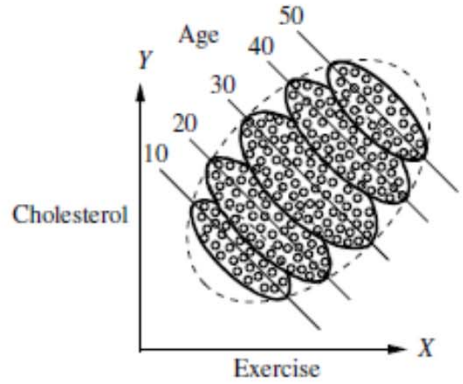


Figure 1.1: Results of the exercise-cholesterol study, segregated by age

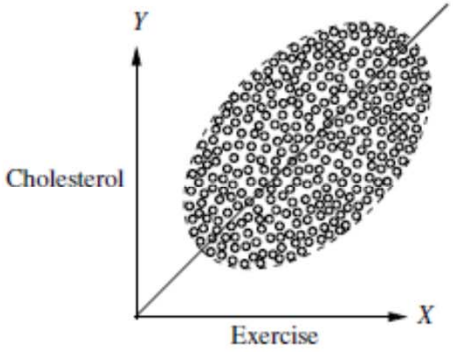


Figure 1.2: Results of the exercise-cholesterol study, unsegregated. The data points are identical to those of Figure 1.1, except the boundaries between the various age groups are not shown

- Because, Age is a common cause of both treatment (exercise) and outcome (cholesterol). So we should look at the age-segregated data in order to compare same-age people, and thereby eliminate the possibility that the high exercisers in each group we examine are more likely to have high cholesterol due to their age, and not due to exercising.

The Simpson Paradox

- Segregated data is not always the right way. What if we record blood (BP) pressure instead of gender?
- We know that drug lower blood pressure but also has a toxic effect.

- Would you recommend the drug to a patient?

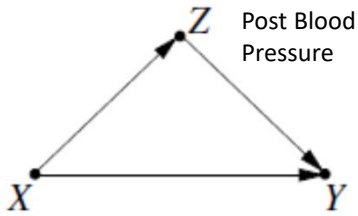
Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

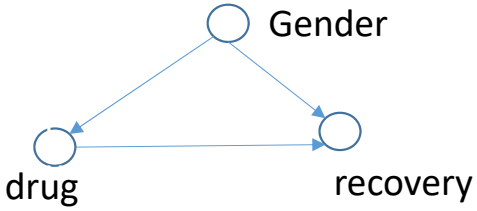
	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

- In the general population, the drug might improve recovery rates because of its effect on blood pressure. But in the subpopulations—the group of people whose post-treatment BP is high and the group whose post-treatment BP is low—we of course would not see that effect; we would only see the drug's toxic effect.
- In this case the aggregated data should be consulted.
- **Same data opposite conclusions!!!**

The Simpson Paradox

- The fact that treatment affect BP and not the opposite was not in the data. Indeed in Statistics it is often stressed that “correlation is not causation”, so there is no statistical method that can determine the causal story from the data alone. Therefore, there is no statistical method that can aid in the decision.





- We can make causal assumptions because we know that drug cannot affect gender. “treatment does not cause sex” cannot be expressed in the data.
- So, what do we do? How can we make causal assumptions and make causal inferences?

The Simpson Paradox SCM (Structural Causal Model)

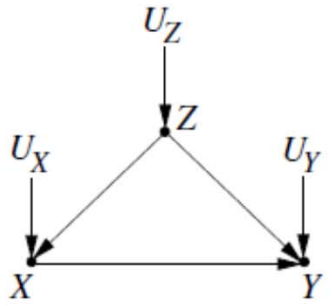


Figure 3.3: A graphical model representing the effects of a new drug, with Z representing gender, X standing for drug usage, and Y standing for recovery

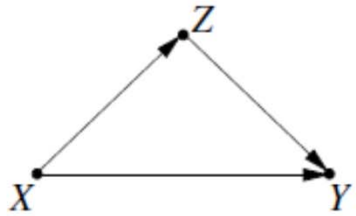


Figure 3.5: A graphical model representing the effects of a new drug, with X representing drug usage, Y representing recovery, and Z representing blood pressure (measured at the end of the study). Exogenous variables are not shown in the graph, implying that they are mutually independent

For Causal Inference We Need:

1. A working definition of “causation”
2. A method by which to formally articulate causal assumptions—that is, to create causal models
3. A method by which to link the structure of a causal model to features of data
4. A method by which to draw conclusions from the combination of causal assumptions embedded in a model and data.

Structural Causal Models (SCM), M

In order to deal with causality we need a formal framework to talk about the causal story

A structural causal model describes how nature assigns values to variables of interest.

- Two sets of variables, U and V and a set of functions f: (U,V,f)
- Each function assigns value to a variable in V based on the values of the other variables.
- Variable X is a direct cause of Y if it appears in the function of Y. X is a cause of Y
- U are exogenous variables (external to the model. We do not explain how they are caused).
- A SCM is associated with **a graphical model**. There is an arc from each direct cause to the node it causes.
- Variables in U have no parents.

SCM 1.5.1 (Salary Based on Education and Experience)

$$U = \{X, Y\}, \quad V = \{Z\}, \quad F = \{f_Z\}$$
$$f_Z : Z = 2X + 3Y$$

Z- salary, X – years in school, Y – years in the profession

X and Y are direct causes for Z

Structural Causal Models (SCM), M

Every SCM is associated with a graphical causal model.

The graphical model G for an SCM M contains one node for each variable in M. If, in M, the function f_X for a variable X contains variable Y (i.e., if X depends on Y for its value), then, in G, there will be a directed edge from Y to X.

We will deal primarily with SCMs that are acyclic graphs (DAGs).

A graphical definition of causation:
If, in a graphical model, a variable X is the child of another variable Y , then Y is a direct cause of X; if X is a descendant of Y , then Y is a potential cause of X .

Structural Causal Models (SCM)

SCM 1.5.2 (Basketball Performance Based on Height and Sex)

$$
\begin{aligned}
V = \quad & \{\text{Height, Sex, Performance}\}, \quad U = \{U_1, U_2, U_3\}, \quad F = \{f1, f2\} \\
& \text{Sex} = U_1 \\
& \text{Height} = f_1(\text{Sex}, U_2) \\
& \text{Performance} = f_2(\text{Height, Sex}, U_3)
\end{aligned}
$$

U are unmeasured terms that we do not care to name. Random causes we do not care about.
U are sometime called error terms

The graphical causal model provides lots of information about what is going on: X causes Y and Y causes Z

A study question

Study question 1.5.1

Suppose we have the following SCM. Assume all exogenous variables are independent and that the expected value of each is 0.

$$
\begin{aligned}
V &= \{X, Y, Z\}, \quad U = \{U_X, U_Y, U_Z\}, \quad F = \{f_X, f_Y, f_Z\} \\
& f_X : X = u_X \\
& f_Y : Y = \tfrac{X}{3} + U_Y \\
& f_Z : Z = \tfrac{Y}{16} + U_Z
\end{aligned}
$$

(a) *Draw the graph that complies with the model.*

(b) *Determine the best guess of the value (expected value) of* Z*, given that we observe* $Y = 3$.

(c) *Determine the best guess of the value of* Z*, given that we observe* $X = 3$.

(d) *Determine the best guess of the value of* Z*, given that we observe* $X = 1$ *and* $Y = 3$.

(e) *Assume that all exogenous variables are normally distributed with zero means and unit variance, that is,* $\sigma = 1$.

(i) *Determine the best guess of* X*, given that we observed* $Y = 2$.

(ii) *(Advanced) Determine the best guess of* Y*, given that we observed* $X = 1$ *and* $Z = 3$. *[Hint: You may wish to use the technique of multiple regression, together with the fact that, for every three normally distributed variables, say* X*,* Y*, and* Z*, we have* $E[Y|X = x, Z = z] = R_{YX \cdot Z}x + R_{YZ \cdot X}z$.*]*

Outline (chapter 3)

- The semantic of Intervention in Structural Causal Models
- The do operators
- How to determine P(Y|do(x)) given an SCM
- The back door criterion and the adjustment formula
- The front door criterion and its adjustment formula

Target: to Determine the Effect of Interventions

- "Correlation is no causation", e.g., Increasing ice-cream sales is correlated with more crime, still selling more ice-cream will not cause more violence. Hot weather is a cause for both.
- **Randomized controlled experiments** are used to determine causation: all factors except a selected one of interest are kept static or random. So the outcome can only be influenced by the selected factor.
- Randomized experiments are often not feasible (we cannot randomize the weather), so how can we determine cause for wildfire?
- **Observational studies** must be used. But how we untangle correlation from causation?

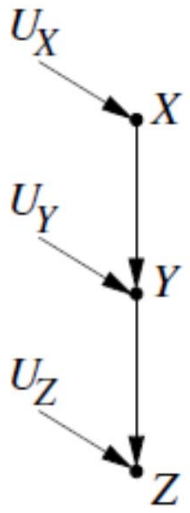
Effect of Interventions (intuition)

A recent University of Winnipeg study that showed that heavy text messaging in teens was correlated with “shallowness.” Media outlets jumped on this as proof that texting makes teenagers more shallow. (Or, to use the language of intervention, that intervening to make teens text less would make them less shallow.) The study, however, proved nothing of the

- It might be the case that shallowness makes teens more drawn to texting.
- It might be that both shallowness and heavy texting are caused by a common factor—a gene, perhaps—and that intervening on that variable, if possible, would decrease both.

Intervention vs Conditioning: When we **intervene** on a variable in a model, we fix its value. We change the system, and the values of other variables often change as a result. When we **condition** on a variable, we change nothing; we merely narrow our focus to the subset of cases in which the variable takes the value we are interested in. What changes, then, is our perception about the world, not the world itself.

Structural Causal Models (SCM)



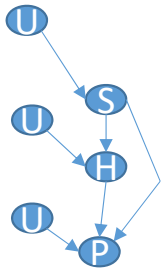
SCM 1.5.2 (Basketball Performance Based on Height and Sex)

$$
\begin{aligned}
V = \quad & \{\text{Height, Sex, Performance}\}, \quad U = \{U_1, U_2, U_3\}, \quad F = \{f1, f2\} \\
& \text{Sex} = U_1 \\
& \text{Height} = f_1(\text{Sex}, U_2) \\
& \text{Performance} = f_2(\text{Height, Sex}, U_3)
\end{aligned}
$$

U are unmeasured terms that we do not care to name. Random causes we do not care about.
U are sometime called error terms.

The graphical causal model provides lots of information about what is going on: X causes Y and Y causes Z

Structural Causal Models (SCM)



SCM 1.5.2 (Basketball Performance Based on Height and Sex)

$$
\begin{aligned}
V = \quad & \{\text{Height, Sex, Performance}\}, \quad U = \{U_1, U_2, U_3\}, \quad F = \{f1, f2\} \\
& \text{Sex} = U_1 \\
& \text{Height} = f_1(\text{Sex}, U_2) \\
& \text{Performance} = f_2(\text{Height, Sex}, U_3)
\end{aligned}
$$

U are unmeasured terms that we do not care to name. Random causes we do not care about.
U are sometime called error terms.

The graphical causal model provides lots of information about what is going on: X causes Y and Y causes Z

Rule of product decomposition

For any model whose graph is acyclic, the joint distribution of the variables in the model is given by the product of the conditional distributions $P(child|parents)$ over all the "families" in the graph. Formally, we write this rule as

$$P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i|pa_i) \tag{1.29}$$

Intervention vs. Conditioning, The Ice-Cream Story

Conditioning P(X=x|Y=y)
Intervening P(X=x| do(Y=y))

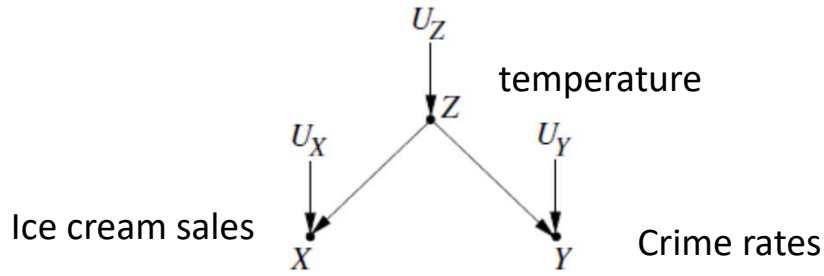
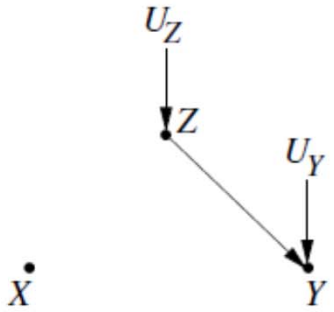


Figure 3.1: A graphical model representing the relationship between temperature (Z), ice cream sales (X), and crime rates (Y)



When we intervene to fix a value of a variable,
We curtail the natural tendencies of the variable to vary
In response to other variables in nature.

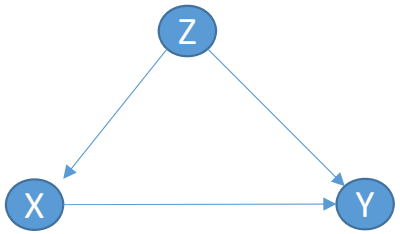
- **This corresponds to a surgery of the model**
- i.e. varying Z will not affect X
- intervention is different than conditioning.
- **Intervention depends on the structure of the graph.**

Figure 3.2: A graphical model representing an intervention on the model in Figure 3.1 that lowers ice cream sales

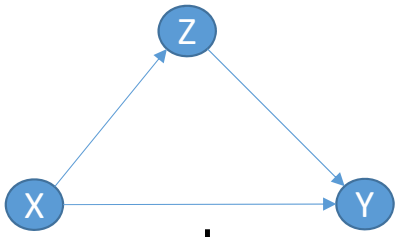
Intervention vs Conditioning, The Surgery Operation

Conditioning P(Y=y|X=x)

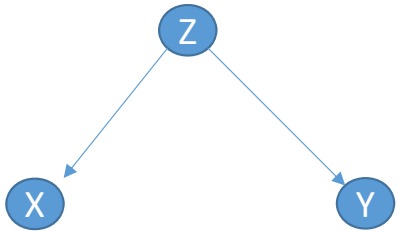
The Simpson story



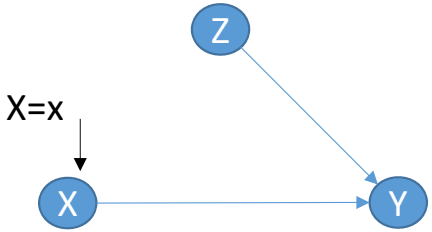
The blood pressure story

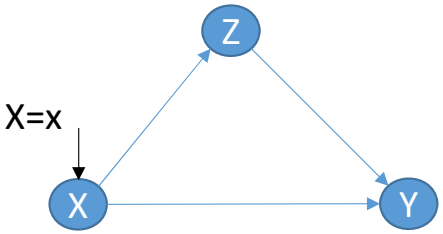


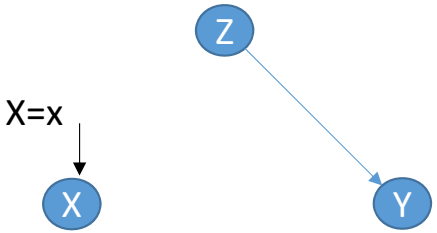
The ice-cream story



Intervening P(Y=y| do(X=x))







Intervention vs. Conditioning…

In notation, we distinguish between cases where a variable X takes a value x naturally and cases where we fix $X = x$ by denoting the latter $do(X = x)$. So $P(Y = y|X = x)$ is the probability that $Y = y$ conditional on finding $X = x$, while $P(Y = y|do(X = x))$ is the probability that $Y = y$ when we intervene to make $X = x$. In the distributional terminology, $P(Y = y|X = x)$ reflects the population distribution of Y among individuals whose X value is x. On the other hand, $P(Y = y|do(X = x))$ represents the population distribution of Y if *everyone in the population* had their X value fixed at x. We similarly write $P(Y = y|do(X = x), Z = z)$ to denote the conditional probability of $Y = y$, given $Z = z$, in the distribution created by the intervention $do(X = x)$.

Do operation and graph surgery can help determine causal effect

We make an assumption that intervention has no side-effect. Namely, assigning a variable by intervention does not affect other variables in a direct way.

The Adjustment Formula

To find out how effective the drug is in the population, we imagine a hypothetical intervention by which we administer the drug uniformly to the entire population and compare the recovery rate to what would obtain under the complementary intervention, where we prevent everyone from using the drug.

We want to estimate the "causal effect difference," or "average causal effect" (ACE).

$$P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0)) \quad (3.1)$$

We need a causal story articulated by a graph (for the Simpson story):

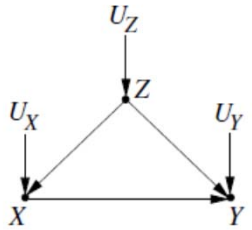
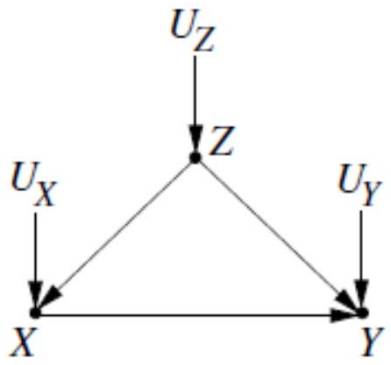


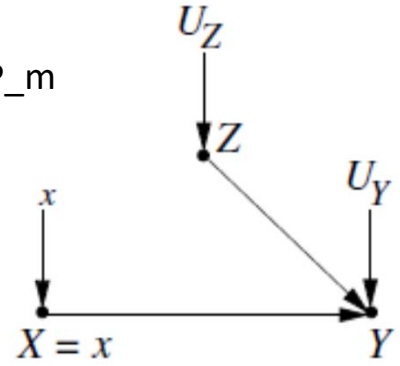
Figure 3.3: A graphical model representing the effects of a new drug, with Z representing gender, X standing for drug usage, and Y standing for recovery

Definition of Intervention and Graph Surgery: The Adjustment Formula



- We simulate the intervention in the form of a graph surgery.
- The causal effect $P(Y = y|do(X = x))$ equals to the conditional probability $P_m(Y = y|X = x)$ that prevails in the manipulated model of the figure below

P_m



Important: the random functions for Z and Y remain invariant

$$P_m(Y = y|Z = z, X = x) = P(Y = y|Z = z, X = x) \quad \text{and} \quad P_m(Z = z) = P(Z = z)$$

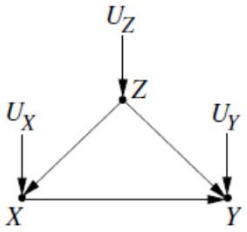


Figure 3.3: A graphical model representing the effects of a new drug, with Z representing gender, X standing for drug usage, and Y standing for recovery

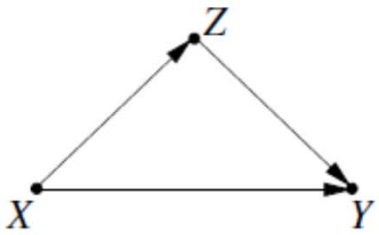
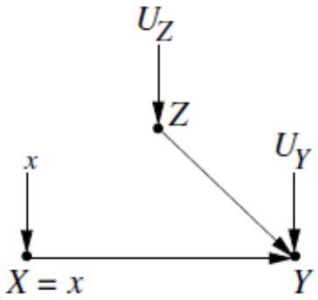


Figure 3.5: A graphical model representing the effects of a new drug, with X representing drug usage, Y representing recovery, and Z representing blood pressure (measured at the end of the study). Exogenous variables are not shown in the graph, implying that they are mutually independent

The Adjustment Formula



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

$$= P_m(Y = y|X = x) \qquad \text{(by definition)} \tag{3.2}$$

$$= \sum_z P_m(Y = y|X = x, Z = z)P_m(Z = z|X = x) \tag{3.3}$$

$$= \sum_z P_m(Y = y|X = x, Z = z)P_m(Z = z) \tag{3.4}$$

Equation (3.3) is obtained from Bayes' rule by conditioning on and summing over all values of $Z = z$ (as in Eq. (1.19)), while (Eq. 3.4) makes use of the independence of Z and X in the modified model.

Finally, using the invariance relations, we obtain a formula for the causal effect, in terms of preintervention probabilities:

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z) \tag{3.5}$$

Equation (3.5) is called the *adjustment formula* and as you can see, it computes the association between X and Y for each value z of Z, then averages over those values. This procedure is referred to as "adjusting for Z" or "controlling for Z."

The Adjustment Formula

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z) \qquad (3.5)$$

The right hand-side can be estimated from the data since it has only conditional probabilities.

If we had a randomized controlled experiments on X (taking the drug) we would not need adjustment
Because the data is already generated from the manipulated distribution. Namely it will yield P(Y=y|do(x))
From the data of the randomized experiment.

In practice adjustment is sometime used in randomized experiments to reduce sampling variations (Cox 1958)

In the Simpson example:

Table 1.1 Results of a study into a new drug, with gender taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

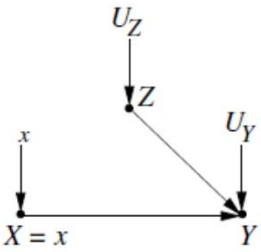
$$P(Y = 1|do(X = 1)) = P(Y = 1|X = 1, Z = 1)P(Z = 1) + P(Y = 1|X = 1, Z = 0)P(Z = 0)$$

Substituting the figures given in Table 1.1 we obtain

$$P(Y = 1|do(X = 1)) = \frac{0.93(87 + 270)}{700} + \frac{0.73(263 + 80)}{700}$$ =0.832

while, similarly,

$$P(Y = 1|do(X = 0)) = \frac{0.87(87 + 270)}{700} + \frac{0.69(263 + 80)}{700}$$ =0.7818



We get that the Average Causal Effect (ACE):

$$ACE = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0)) = 0.832 - 0.7818 = 0.0502$$

A more informal interpretation of ACE is that it is the difference in the fraction of the population that would recover if everyone took the drug compared to when no one takes the drug.

The Blood Pressure Example

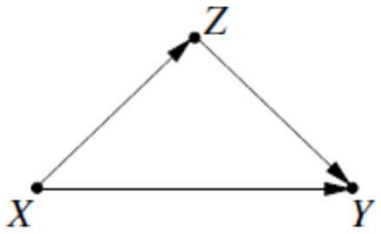


Figure 3.5: A graphical model representing the effects of a new drug, with X representing drug usage, Y representing recovery, and Z representing blood pressure (measured at the end of the study). Exogenous variables are not shown in the graph, implying that they are mutually independent

P(Y=y | do(X=x) = ? Here the "surgery on X changes nothing. So,

This means that no surgery is required; the conditions under which data were obtained were such that treatment was assigned "as if randomized." If there was a factor that would make subjects prefer or reject treatment, such a factor should show up in the model; the absence of such a factor gives us the license to treat X as a randomized treatment.

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

To Adjust or not to Adjust?

Rule 1 (The Causal Effect Rule) *Given a graph G in which a set of variables PA are designated as the parents of X, the causal effect of X on Y is given by*

$$P(Y = y|do(X = x) = \sum_z P(Y = y|X = x, PA = z)P(PA = z) \qquad (3.6)$$

where z ranges over all the combinations of values that the variables in PA can take.

So, the causal graph helps determine the parents PA!

But, in many cases some of the parents are unobserved so we cannot perform the calculation.

Luckily we can often adjust for other variables substituting for the unmeasured variables in PA(X), and this Can be decided via the graph.

Multiple Interventions, the Truncated Product Rule

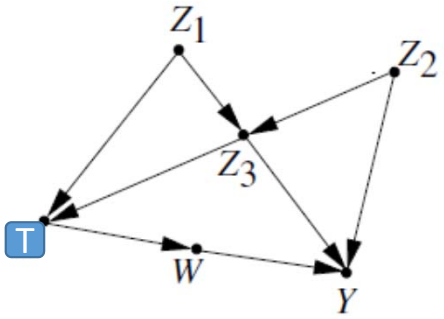
Often we have multiple interventions that may not correspond to disconnected variables.
We will use the product decomposition. We write **the product truncated formula**

$$P(x_1, x_2, \ldots, x_n | do(x)) = \prod_i P(x_i | pa_i) \qquad \text{for all } i \text{ with } X_i \text{ not in } X.$$

Example:

$$P(z_1, z_2, w, y | do(T = t, Z_3 = z_3)) = P(z_1)P(z_2)P(w|t)P(y|w, z_3, z_2)$$

where we have deleted the factors $P(t|z_1, z_3)$ and $P(z_3|z_1, z_2)$ from the product.



Multiple Interventions and the Truncated Product Rule

preintervention distribution in the model of Figure 3.3 is given by the product

$$P(x, y, z) = P(z)P(x|z)P(y|x, z) \tag{3.8}$$

whereas the postintervention distribution, governed by the model of Figure 3.4 is given by the product

$$P(z, y|do(x)) = P_m(z)P_m(y|x, z) = P(z)P(y|x, z) \tag{3.9}$$

with the factor $P(x|z)$ purged from the product, since X becomes parentless as it is fixed at $X = x$. This coincides with the adjustment formula, because to evaluate $P(y|do(x))$ we need to marginalize (or sum) over z, which gives

$$P(y|do(x)) = \sum_z P(z)P(y|x, z)$$

in agreement with (3.5).

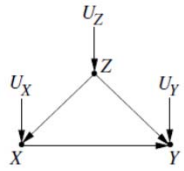


Figure 3.3: A graphical model representing the effects of a new drug, with Z representi
gender, X standing for drug usage, and Y standing for recovery

Multiple Interventions and the Truncated Product Rule

It is interesting to note that combining (3.8) and (3.9), we get a simple relation between the pre- and postintervention distributions:

$$P(z, y|do(x)) = \frac{P(x, y, z)}{P(x|z)} \tag{3.10}$$

It tells us that the conditional probability $P(x|z)$ is all we need to know in order to predict the effect of an intervention $do(x)$ from nonexperimental data governed by the distribution $P(x, y, z)$.

3.3 The Backdoor Criterion

Under what conditions does a causal story permit us to compute the causal effect of one variable on another, from data obtained by passive observations, with no interventions? Since we have decided to represent causal stories with graphs, the question becomes a graph-theoretical problem: Under what conditions is the structure of the causal graph sufficient for computing a causal effect from a given data set?

3.3 The Backdoor Criterion

Definition 3.3.1 (The Backdoor Criterion) *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*

If a set of variables Z satisfies the backdoor criterion for X and Y, then the causal effect of X on Y is given by the formula

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$$

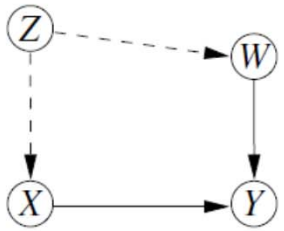
Rationale:

1. We block all spurious paths between X and Y.
2. We leave all directed paths from X to Y unperturbed.
3. We create no new spurious paths.

Rationale:

When trying to find the causal effect of X on Y, we want the nodes we condition on to block any "backdoor" path in which one end has an arrow into X, because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X, and if we do not block them, they will confound the effect that X has on Y. We condition on backdoor paths so as to fulfill our first requirement. However, we don't want to condition on any nodes that are descendants of X. Descendants of X would be affected by an intervention on X and might themselves affect Y; conditioning on them would block those pathways. Therefore, we don't condition on descendants of X so as to fulfill our second requirement. Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y. The requirement of excluding descendants of X also protects us from conditioning on children of intermediate nodes between X and Y (e.g., the collision node W in Figure 2.4.) Such conditioning would distort the passage of causal association between X and Y, similar to the way conditioning on their parents would.

Examples for Backdoors



P(Y|do(X))?

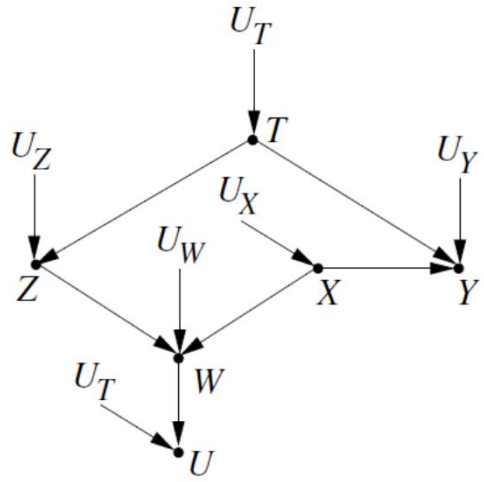
Figure 3.6: A graphical model representing the relationship between a new drug (X), recovery (Y), weight (W), and an unmeasured variable Z (socioeconomic status)

W is a backdoor. Therefore we can compute:

$$P(Y = y|do(X = x)) = \sum_{w} P(Y = y|X = x, W = w)P(W = w)$$

Examples

P(Y|do(X))?



No backdoors between X and Y and therefore: P(Y|do(X))= P(Y|X)

What if we adjust for W? … wrong!!!

But what if we want to determine P(Y|do(X),w)? What do we do with the spurious path $X \rightarrow W \leftarrow Z \leftrightarrow T \rightarrow Y$?

if we condition on T, we would block the spurious path $X \rightarrow W \leftarrow Z \leftrightarrow T \rightarrow Y$. We can compute:

$$P(Y = y|do(X = x), W = w) = \sum_t P(Y = y|X = x, W = w, T = t)P(T = t|W = w)$$

Example: W can be post-treatment pain

Adjusting for Colliders?

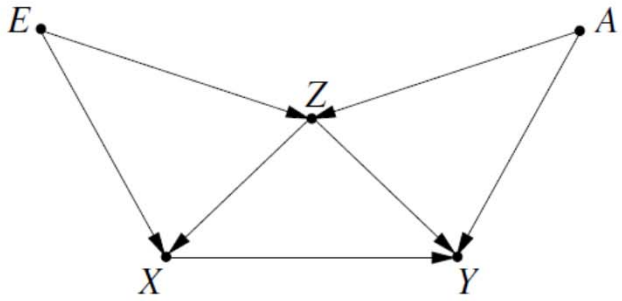


Figure 3.7: A graphical model in which the backdoor criterion requires that we condition on a collider (Z) in order to ascertain the effect of X on Y

There are 4 backdoor paths. We must adjust for Z, and one of E or A or both

The Front Door Criterion

When we don't have a backdoor path, we may still have a front door path

Consider the century-old debate on the relation between smoking and lung cancer. In the years preceding 1970, the tobacco industry has managed to prevent antismoking legislation by promoting the theory that the observed correlation between smoking and lung cancer could be explained by some sort of carcinogenic genotype that also induces an inborn craving for nicotine.

A graph depicting this example is shown in Figure 3.10(a) This graph does not satisfy

Causal effect not identifiable here

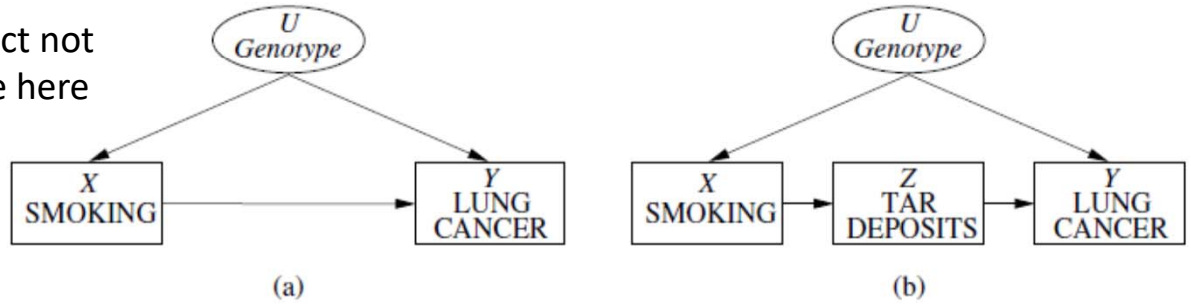


Figure 3.10: A graphical model representing the relationships between smoking (X) and lung cancer (Y), with unobserved confounder (U) and a mediating variable Z

Front Door…

We cannot satisfy the backdoor criterion since we cannot measure U. But consider the model in (b).
It does not satisfy the backdoor criterion, but we can measure the tar level, Z, which will allow identifiability of P(Y|do(X)),

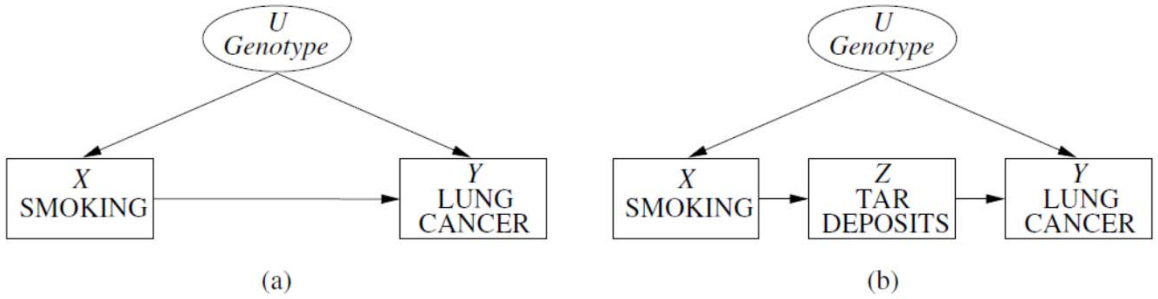


Figure 3.10: A graphical model representing the relationships between smoking (X) and lung cancer (Y), with unobserved confounder (U) and a mediating variable Z

Example

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

	Tar 400		**No tar 400**		**All subjects 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%)

Tobaco industry:
Only 15% of smoker developed cancer while 90% from the non-smoker

Antismoke lobbyist:
If you smoke you have 95% tar vs no smokers (380/400 vs 20/400)

Table 3.2 Reorganization of the dataset of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	SMOKERS 400		**NON-SMOKERS 400**		**ALL SUBJECTS 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%)	(9%)	(81%)

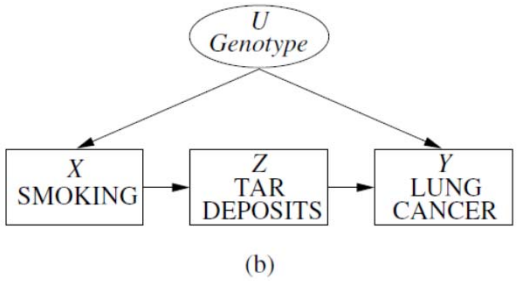
If you have more tar, you increase the chance of cancer in both smoker (from 10% to 15%) and non-smokers (from 90% To 95%).

The graph of Figure 3.10(b) enables us to decide between these two groups of statisticians. First, we note that the effect of X on Z is identifiable, since there is no backdoor path from X to Z. Thus, we can immediately write

$$P(Z = z|do(X = x)) = P(Z = z|X = x) \tag{3.12}$$

Next we note that the effect of Z on Y is also identifiable, since the backdoor path from Z to Y, namely $Z \leftarrow X \leftarrow U \rightarrow Y$, can be blocked by conditioning on X. Thus we can write

$$P(Y = y|do(Z = z)) = \sum_x P(Y = y|Z = z, X = x) \text{ P(x)} \tag{3.13}$$



(b)

We are now going to chain together the two partial effects to obtain the overall effect of X on Y. The reasoning goes as follows: If nature chooses to assign Z the value z, then the probability of Y would be $P(Y = y|do(Z = z))$. But the probability that nature would choose to do that, given that we choose to set X at x, is $P(Z = z|do(X = x))$. Therefore, summing over all states z of Z we have

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|do(Z = z))P(Z = z|do(X = x)) \tag{3.14}$$

The terms on the right hand side of (3.14) were evaluated in (3.12) and (3.13), and we can substitute them to obtain a *do*-free expression for $P(Y = y|do(X = x))$. We also distinguish between the x that appears in (3.12) and the one that appears in (3.13), the latter of which is merely an index of summation and might as well be denoted x'. The final expression we have is

$$P(Y = y|do(X = x)) =$$
$$\textstyle\sum_z \sum_{x'} P(Y = y|Z = z, X = x')P(X = x')P(Z = z|X = x) \tag{3.15}$$

Equation (3.15) is known as the *front-door formula*.

Definition 3.4.1 (Front-Door)
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 unblocked backdoor path from X to Z.*
3. *All backdoor paths from Z to Y are blocked by X.*

Theorem 3.4.1 (Front-Door Adjustment)
If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x'). \tag{3.16}$$

The Do-Calculus

Theorem 3.4.1 (Front-Door Adjustment)
If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x'). \tag{3.16}$$

The conditions stated in Definition 3.4.1 are overly conservative; some of the backdoor paths excluded by conditions (ii) and (iii) can actually be allowed provided they are blocked by some variables. There is a powerful symbolic machinery, called the *do-calculus*, that allows analysis of such intricate structures. In fact, the *do*-calculus uncovers *all* causal effects that can be identified from a given graph. Unfortunately, it is beyond the scope of this book (see Pearl 2009 and Shpitser and Pearl 2008 for details). But the combination of the adjustment formula, the backdoor criterion, and the front-door criterion covers numerous scenarios. It proves the enormous, even revelatory, power that causal graphs have in not merely representing, but actually discovering causal information.

From chapter 1 in popular book

At the time, I was so intoxicated with the power of probabilities that I considered causality to be a subservient concept, merely a convenience or a mental shorthand for expressing probabilistic dependencies, and for distinguishing the relevant from the irrelevant. In my 1988 book *Probabilistic Reasoning in Intelligent Systems*, I wrote, "Causation is a language with which one can talk efficiently about certain structures of relevance relationships." The words embarrass me today, because "relevance" is so obviously a rung 1 notion. Even by the time the book was published, I knew in my heart that I was wrong. To my fellow computer scientists, my book was the bible of reasoning under uncertainty, but I was already feeling like an apostate.

Consider the graph in Figure 3.8:

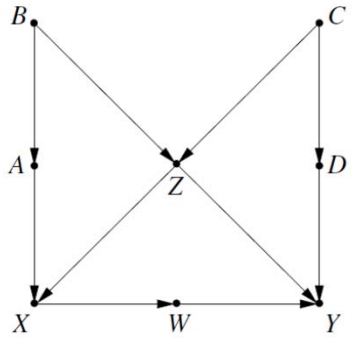


Figure 3.8: Causal graph used to illustrate the backdoor criterion in the following study questions

(a) *List all of the sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y.*

(b) *List all of the minimal sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y (i.e., any set of variables such that, if you removed any one of the variables from the set, it would no longer meet the criterion).*

(c) *List all minimal sets of variables that need be measured in order to identify the effect of D on Y. Repeat, for the effect of $\{W, D\}$ on Y.*

Study question 3.4.1

Assume that in Figure 3.8, only X, Y, and one additional variable can be measured. Which variable would allow the identification of the effect of X on Y? What would that effect be?

Conditional intervention

Assume a policy x= g(Z) when Z is a random variable(Z can be age. And we may give a drug conditiononed on Z>z_0)
We are interested to asses P(Y| do (X=g(Z)). We can often get it through z-specific effect of P(Y|do(X=x),Z=z)

Rule 2 *The z-specific effect $P(Y = y|do(X = x), Z = z)$ is identified whenever we can measure a set S of variables such that $S \cup Z$ satisfies the backdoor criterion. Moreover, the z-specific effect is given by the following adjustment formula*

$$P(Y = y|do(X = x), Z = z)$$
$$= \sum_s P(Y = y|X = x, S = s, Z = z)P(S = s)$$

Conditional Intervention

We now show that identifying the effect of such policies is equivalent to identifying the expression for the z-specific effect $P(Y = y|do(X = x), Z = z)$.

To compute $P(Y = y|do(X = g(Z)))$, we condition on $Z = z$ and write

$$\begin{aligned} & P(Y = y|do(X = g(Z))) \\ &= \sum_z P(Y = y|do(X = g(Z)), Z = z)P(Z = z|do(X = g(Z))) \\ &= \sum_z P(Y = y|do(X = g(z)), Z = z)P(Z = z) \end{aligned} \tag{3.17}$$

The equality

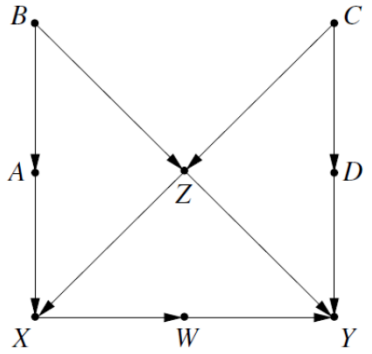
$$P(Z = z|do(X = g(Z))) = P(Z = z)$$

stems, of course, from the fact that Z occurs before X; hence, any control exerted on X can have no effect on the distribution of Z. Equation (3.17) can also be written as

$$\sum_z P(Y = y|do(X = x), z)|_{x=g(z)} P(Z = z)$$

which tells us that the causal effect of a conditional policy $do(X = g(Z))$ can be evaluated directly from the expression of $P(Y = y|do(X = x), Z = z)$ simply by substituting $g(z)$ for x and taking the expectation over Z (using the observed distribution $P(Z = z)$).

Conditional Intervention



Study question 3.5.1

Consider the causal model of Figure 3.8.

(a) Find an expression for the c-specific effect of X on Y.

(b) Identify a set of four variables that need to be measured in order to estimate the z-specific effect of X on Y, and find an expression for the size of that effect.

(c) Using your answer to part (b), determine the expected value of Y under a Z-dependent strategy where X is set to 0 when Z is smaller or equal to 2 and X is set to 1 when Z is larger than 2. (Assume Z takes on integer values from 1 to 5.)