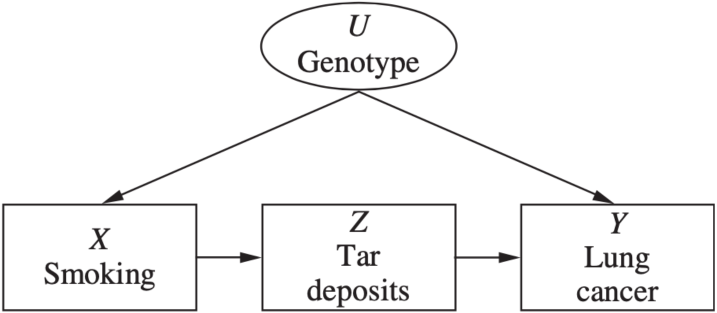
Estimating causal effects

In the previous lecture, we saw that

- the causal effect $P(Y \mid \operatorname{do} X = x)$ is the distribution of Y if we *change* the system so that X is always x
  Note: data from this modified system is called experimental data
- In practice, we may have only observational data (from the unmodified system), because doing the experiment would be too costly, unethical, or simply impossible
- Because the two systems are related, sometimes we can still use the observational data to calculate $P(Y \mid \operatorname{do}(X = x))$
  1. Find a set of variables to adjust for (e.g. satisfying backdoor criterion)
  2. Apply adjustment formula

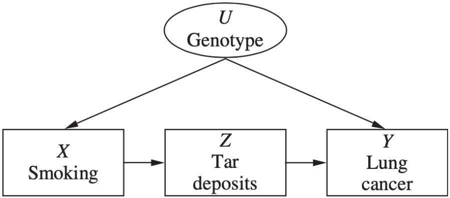
More complex case

We'll take a look at a more complex case, where the causal effect is not of the form of the adjustment formula



U is unmeasured (also called latent), X, Z, Y are measured

- Can't condition on unmeasured parent U for causal effect rule
- Can't block spurious path $X \leftarrow U \rightarrow Y$ for backdoor criterion



What we can do:

- Estimate $P(Z \mid \text{do}(X = x))$ directly (no adjustment necessary: no open backdoor paths between X and Z)
- Estimate $P(Y \mid \text{do}(Z = z))$ using the backdoor criterion, adjusting for X

That should be enough to find $P(Y \mid \text{do}(X = x))$, using that

$$P(Y = y \mid \text{do}(X = x)) \\ = \sum_z P(Y = y \mid \text{do}(Z = z)) P(Z = z \mid \text{do}(X = x))$$

Substituting

$$P(Z = z \mid \mathrm{do}(X = x)) = P(Z = z \mid X = x)$$
$$P(Y = y \mid \mathrm{do}(Z = z)) = \sum_x P(Y = y \mid Z = z, X = x)P(X = x)$$

into

$$P(y \mid \mathrm{do}(x)) = \sum_z P(y \mid \mathrm{do}(z))P(z \mid \mathrm{do}(x))$$

gives

$$\begin{aligned} &P(Y = y \mid \mathrm{do}(X = x)) \\ &= \sum_z P(Z = z \mid X = x) \sum_{x'} P(Y = y \mid Z = z, X = x')P(X = x') \end{aligned}$$

(Note that the value x we sum over in $P(Y = y \mid \mathrm{do}(Z = z))$ is unrelated to the actual x in $P(Y = y \mid \mathrm{do}(X = x))$, so we renamed it to x')

Data example

(Not real data for this problem!)

	Smokers 400		Nonsmokers 400	
	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)

Front-door criterion

General case:

Front-door criterion

For finding the causal effect of X on Y, a set Z satisfies the front-door criterion if

- Z blocks all directed paths from X to Y
- there is no backdoor path from X to Z
- X blocks all backdoor paths from Z to Y

Covariate-specific effects

So far we have seen $\text{do}(\cdot)$ appear by itself on the right of the conditioning bar. It can also appear in together with ordinary conditioning: $P(Y \mid \text{do}(X = x), Z = z)$: the *z-specific effect* of X on Y

This means that

- We look at the modified graph where X is intervend on
- For the distribution from that graph, we condition on $Z = z$

Order doesn't matter:

$$P(Y \mid \text{do}(X = x), Z = z) = P(Y \mid Z = z \, \text{do}(X = x))$$

So if X is a cause of Z, remember that the conditioning also refers to the modified Z

Conditional interventions

If X is *not* a cause of Z:
We can use $P(Y \mid \mathrm{do}(X = x), Z = z)$ to talk about the conditional intervention

$$P(Y = y \mid \mathrm{do}(X = g(Z)))$$

E.g. if a doctor bases the choice of treatment X on some other variable Z, such as the patient's temperature

Conditional interventions

If X is *not* a cause of Z:
We can use $P(Y \mid \text{do}(X = x), Z = z)$ to talk about the conditional intervention

$$P(Y = y \mid \text{do}(X = g(Z)))$$

E.g. if a doctor bases the choice of treatment X on some other variable Z, such as the patient's temperature

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

where in the final equation, we know the actual value of $g(Z)$ because we are conditioning on $Z = z$
(In a way, here the order of intervening and conditioning is switched)

Inverse probability weighting: introduction

So far, we have seen expressions for causal effects in terms of the observational distribution

In the practical problem that we don't know the exact observational distribution but have data from it instead, what can we do?

Propensity scores

Suppose we have a valid adjustment formula (e.g. from the backdoor criterion), and we want to estimate it based on data
Rewriting trick for the adjustment formula:

$$
\begin{aligned}
P(Y = y \mid \mathrm{do}(X = x)) &= \sum_z P(Y = y \,|\, X = x, Z = z) P(Z = z) \\
&= \sum_z \frac{P(Y = y \,|\, X = x, Z = z) P(Z = z) P(X =}{P(X = x \,|\, Z = z)} \\
&= \sum_z \frac{P(X = x, Y = y, Z = z)}{P(X = x \,|\, Z = z)}
\end{aligned}
$$

The thing we divide by, $P(X = x \,|\, Z = z)$ is called the propensity score

Weighting by propensity scores

$$P(Y = y \mid \mathrm{do}(X = x)) = \sum_z \frac{P(X = x, Y = y, Z = z)}{P(X = x \mid Z = z)}$$

Here,

- $P(X, Y, Z)$ is exactly the distribution represented by our data samples
- The samples are weighted by some number, namely $1/P(X = x \mid Z = z)$

Compare to conditioning, where we might estimate $P(Y = y \mid X = x)$ by weighting our data samples by either 0 (if $x_i \neq x$) or some constant c (if $x_i = x$)

Getting the propensity scores

The traditional way of getting the propensity scores is by *logistic regression*

- Good choice, because that method works by estimating probabilities of data

Other machine learning methods can also be used, in particular:

- Neural networks: with a softmax activation in the output layer, these also estimate probabilities
- Random forests and variations work well too

Support vector machines do *not* work, because they don't estimate a probability (though variations exist that do)

Inverse probability weighting: pros and cons

Advantages:

- If Z has many possible values (e.g. if it consists of many variables), the computation time depends not on the space of possible values, but on the number of data points

Disadvantages:

- If the estimated probability is small, we multiply by a very large number that could easily have been smaller/larger; this can make our overall estimation unstable

Mediation

TODO :o