BME I5100: Biomedical Signal Processing

Causal inference
introduction

Lucas C. Parra
Biomedical Engineering Department
City College of New York

CCNY

Recommended reading

Accessible introductions

Pearl, J., and Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect (Basic Books).

Angrist, J.D., and Pischke, J.-S. (2014). Mastering 'Metrics: The Path from Cause to Effect (Princeton University Press).

Textbooks

Pearl, J. (2009). Causality (Cambridge: Cambridge University Press).

VanderWeele, T. (2015). Explanation in Causal Inference: Methods for Mediation and Interaction (Oxford University Press).

Papers on special topics

Pearl, J. (2013). Linear Models: A Useful "Microscope" for Causal Analysis. Journal of Causal Inference 1, 155–170.

Cinelli, C., Forney, A., and Pearl, J. (2020). A Crash Course in Good and Bad Controls (Rochester, NY: Social Science Research Network)

Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association 91, 444–455.

Causal graphs vs associative graphs

Causal graphs have direction, unlike associative networks*

$$P(Y|do(X))$$

Causal: Distribution of Y that results if we do fix value of X

$$P(Y,X)$$

$P(X|Y)$ $P(Y|X)$

Associative: Observed distribution of Y and X

Unidirectional because changing X will change Y, but adjusting Y does not change X.

Example: When it rains, the grass gets wet, but poring water on the grass does not make it rain.

Bidirectional because the observed value of Y is predictive of observed value of X, and vise versa, even if they don't directly affect each other.
Example: The weight of one sibling predicts weight of the other (genes, culture, etc), but dieting will not affect the weight of the sibling (unless they copy each other).

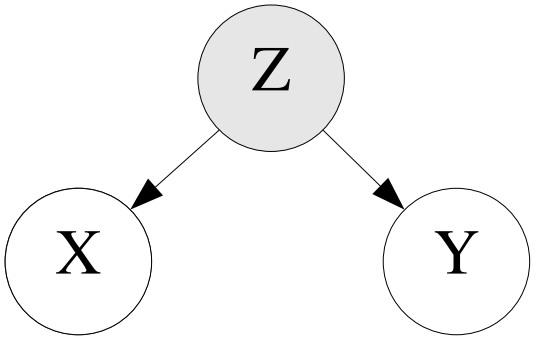
* Bayes networks are associative. Direction, can be turned around with Bayes rule.

Can not be inferred from data alone

Best to think of it in terms of process how the data was generated.

Example: Direct causal effect of X on Y

Example: Association of X with Y due to unobserved common cause Z



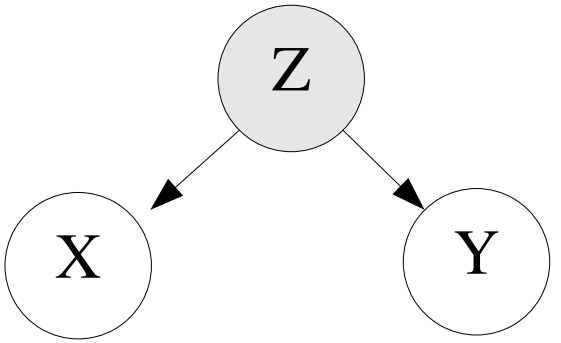
In code (using ← instead of = to make assignment clear):

```
X← randn(N)
Y← 3*X+randn(N)
```

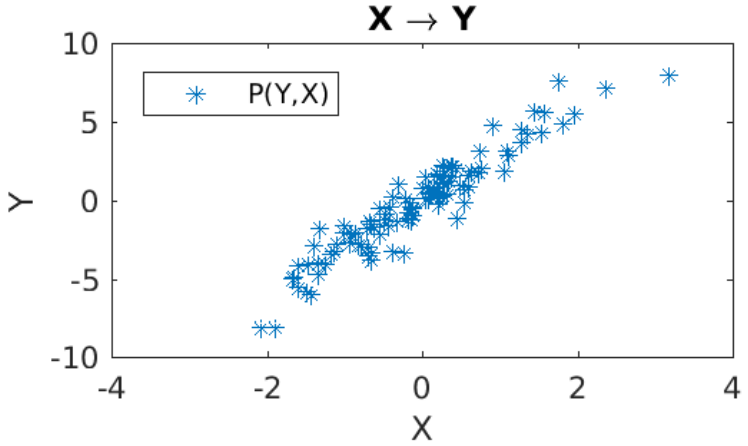
```
Z ← randn(N)  % unobserved
X ← 1*Z
Y ← 3*Z + randn(N)
```

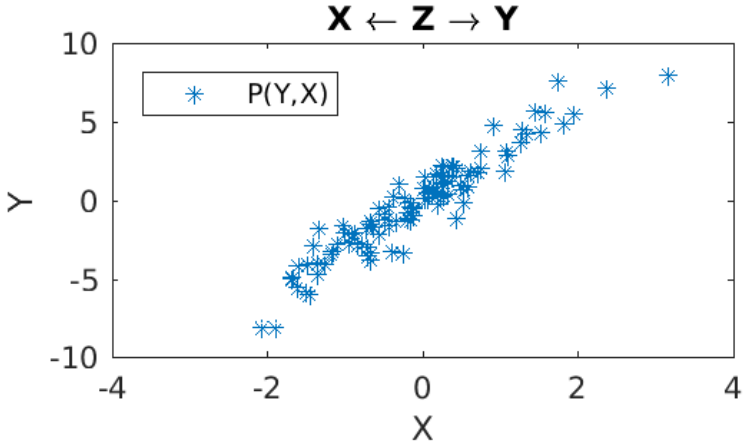
These two processes can generate identical data ...

Can not be inferred from data alone

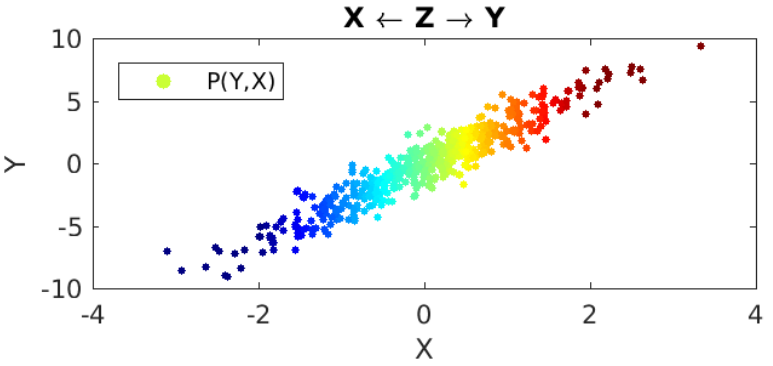


From just observing X and Y, the two models are indistinguishable:





Unless of course you get to see Z (here color-coded):



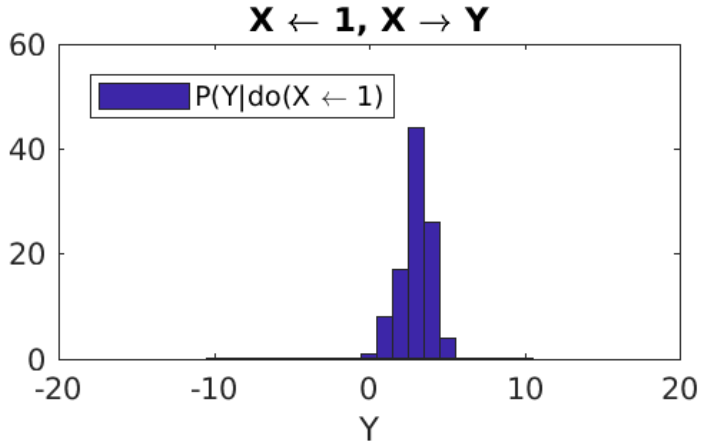
Intervention – do(x)

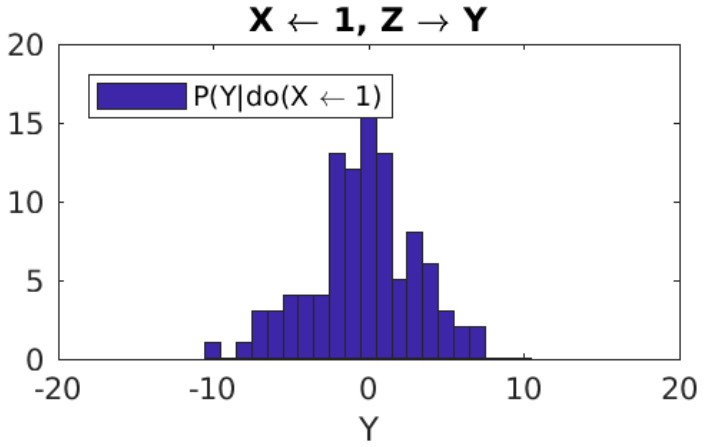
Don't need to see Z if we intervene to do ($X \leftarrow 1$) …

$X \leftarrow 1$
$Y \leftarrow 3+\text{randn(N)}$

$Z \leftarrow \text{randn(N)}$
$X \leftarrow 1$
$Y \leftarrow 3*Z + \text{randn(N)}$

… then the two processes will result in different observed $P(Y|\text{do}(x))$

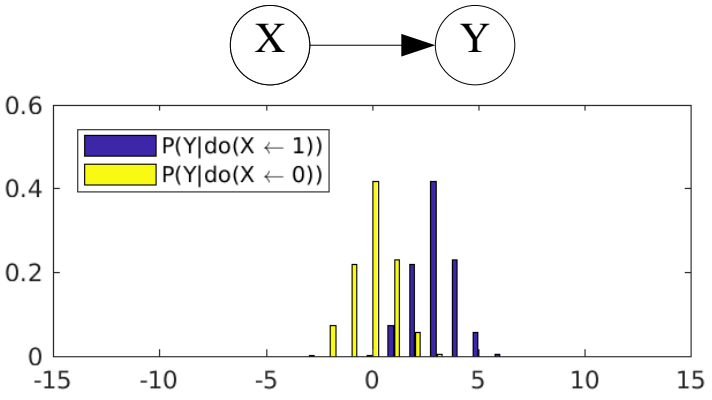




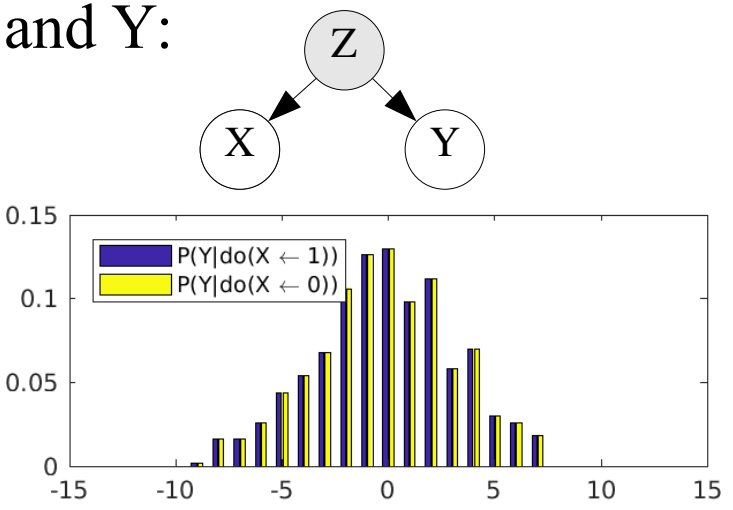
In conclusion, $P(Y|\text{do}(x))$ captures how the distribution of Y is causally affected by forcing $X \leftarrow x$.

Intervention – do(x)

$P(Y|\text{do}(X \leftarrow x))$ will depend on what value we pick for x, but only when there is a causal link between X and Y:

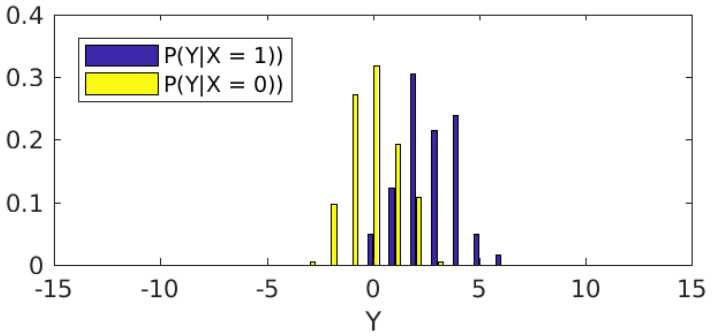


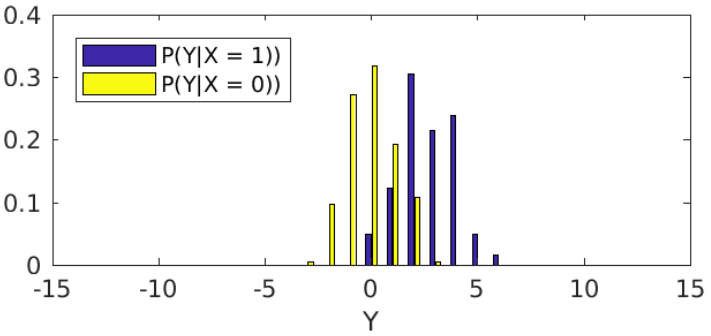
=> X causes Y



=> X does not cause Y

In contrast, $P(Y|X)$ only captures the association, which may be due to cause and effect, or some unobserved common cause.

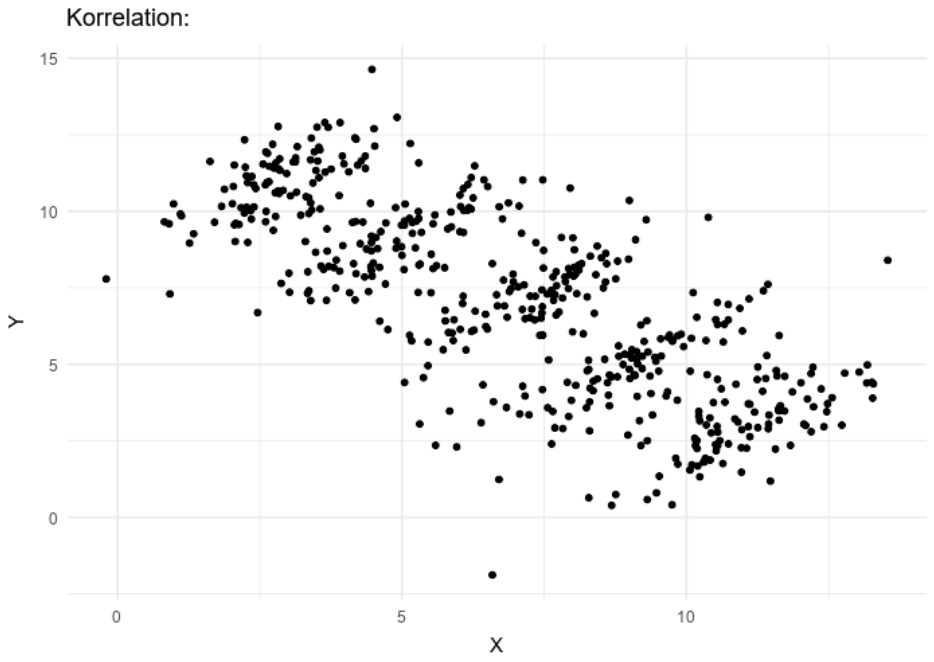


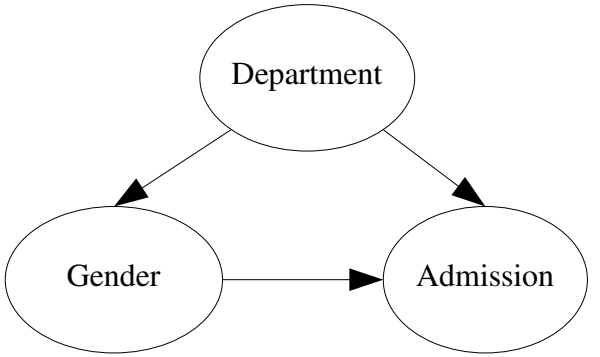


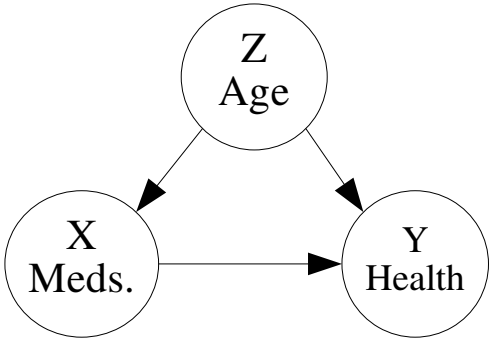
P(Y|X) can not tell the difference.

Simpson's Paradox

Example – Gender bias: Graduate admissions at Berkley in 1973 shows a bias in favor of males University wide, but a bias in favor of women in most individual department. How is that possible?







https://en.wikipedia.org/wiki/Simpson%27s_paradox

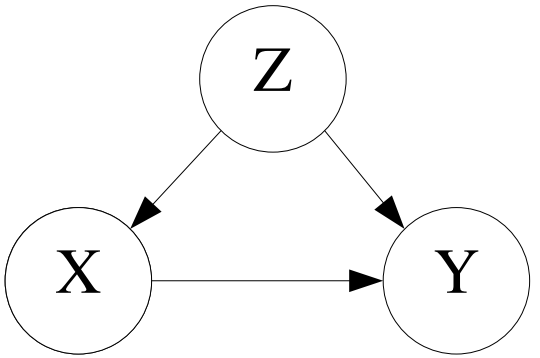
Conclusion:

- Need to "control" for common cause.

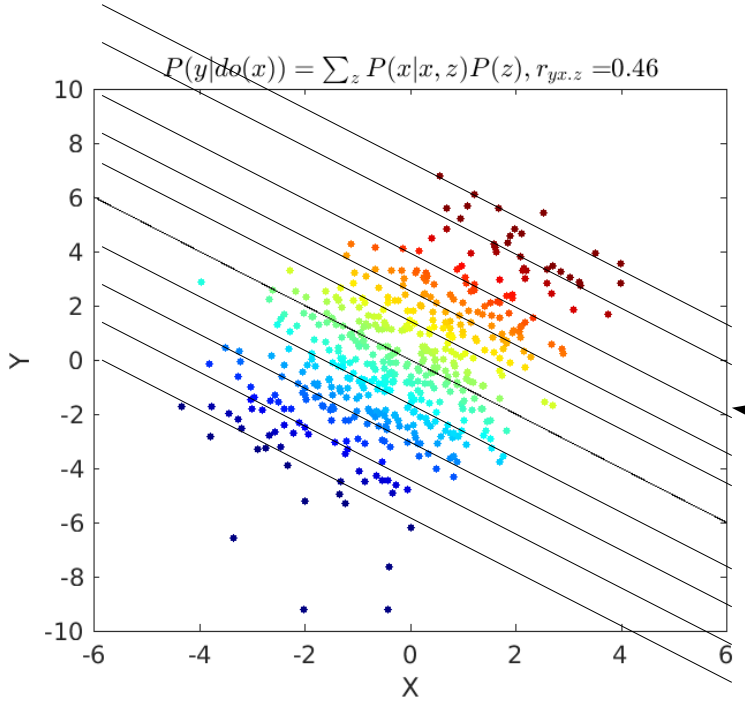
Adjustment formula

do(x) while controlling for Z :

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



This motivates the name "Average Causal Effect" (ACE) for $P(Y|\text{do}(X))$.

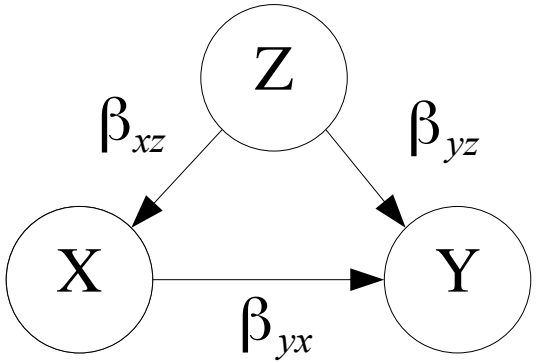


$Z = z$

Linear models

Making causal model relationship explicit in structural equations. For example, assume that:

X affects Y, but both are affected by Z:



Structural equations are linear:

$$Z \leftarrow U_z$$
$$X \leftarrow \beta_{xz} Z + U_x$$
$$Y \leftarrow \beta_{yx} X + \beta_{yz} Z + U_y$$

Parameters β reflect the causal effects we want to estimate. For example, β_{yx} is the size of the effect that a unit increase in X causes on Y.

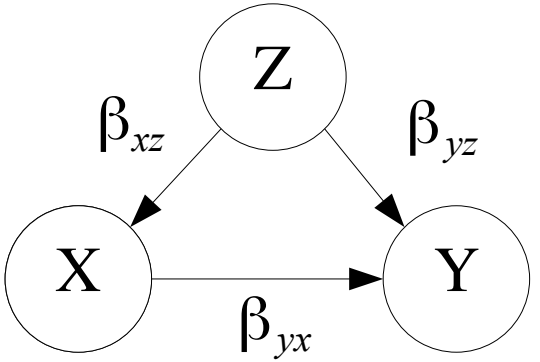
U_z and U_y are unobserved source of individual variation, or "omitted factors". If they are normally distributed we can estimate the ACE with ordinary least squares from observations X_i, Y_i, Z_i. i=1..N

Confounding

The effect of X on Y is confounded by the common cause Z. We can "control" for Z by estimating parameters with ordinary least squares (OLS)*:

$$y = \beta_{yx} x + \beta_{yz} z + e_y$$

β_{yx}, β_{yz} are called "partial regression coefficients", because each explains a part of the variance of y. They are denoted $r_{yx.z}, r_{yz.x}$ to make explicit their dependence on the other variable.



With OLS we can also estimate a few other associations, e.g.

$$y = r_{yx} x + e_y \qquad z = r_{zx} x + e_z$$

Well-established property of OLS reveals the bias introduced by Z*:

$$r_{yx} = \beta_{yx} + \beta_{yz} r_{zx}$$

Association = Direct effect + Bias
X→Y X→Y X←Z→Y

Z was "controlled for" by adding as regressor to the OLS.

* A good summary on linear models for causal inference is Pearl, J. (2013).

Assignment

Derive this relationship

$$r_{yx} = \beta_{yx} + \beta_{yz} r_{zx}$$

By correlating the following equations with x

$$y = \beta_{yx} x + \beta_{yz} z + e$$

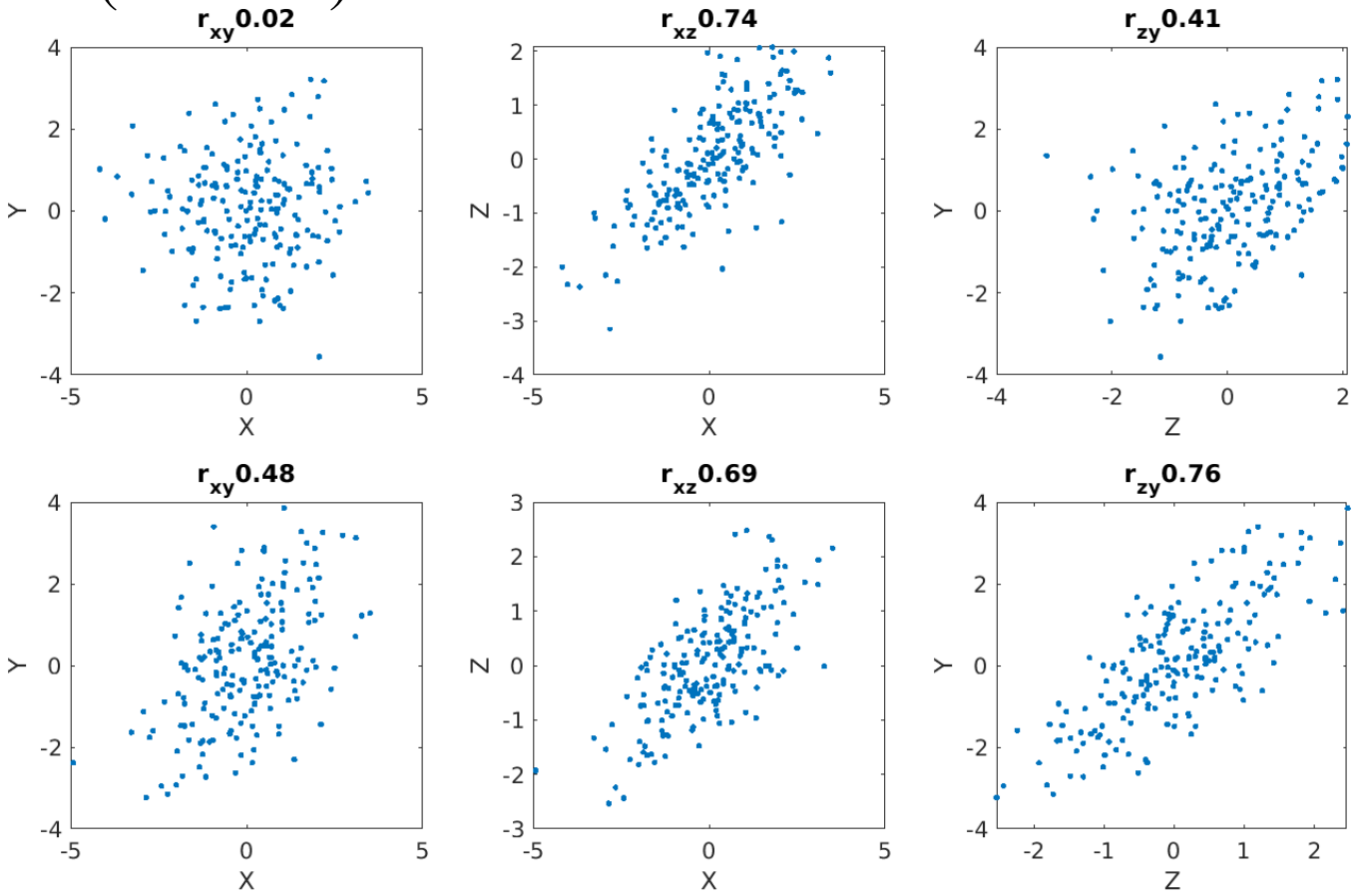
$$y = r_{yx} x + e$$

$$z = r_{zx} x + e$$

And solving for β_{yx}, β_{yz}, r_{yx}, r_{yz}. You can assume that the correlation of x with the error terms e are all zero, and that all variables are zero mean.

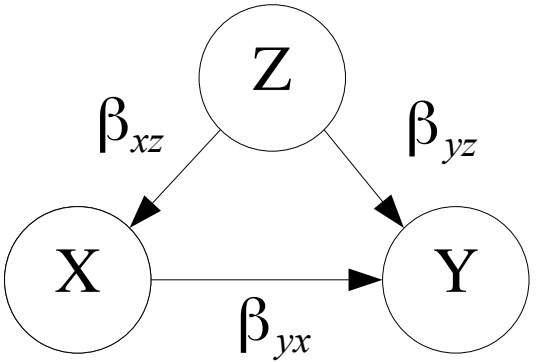
Assignment

Look at the two datasets of variables X, Y, Z (on website cause_or_confound.mat). Answer the question, does X causally affect Y assuming that Z may be a common cause for X and Y, i.e. prior knowledge model assumes: X←Z → Y, X→Y)? Use OLS to estimate all regression parameters with Z as potential confound. Compute all effect sizes (all betas)



Confounding – example in code

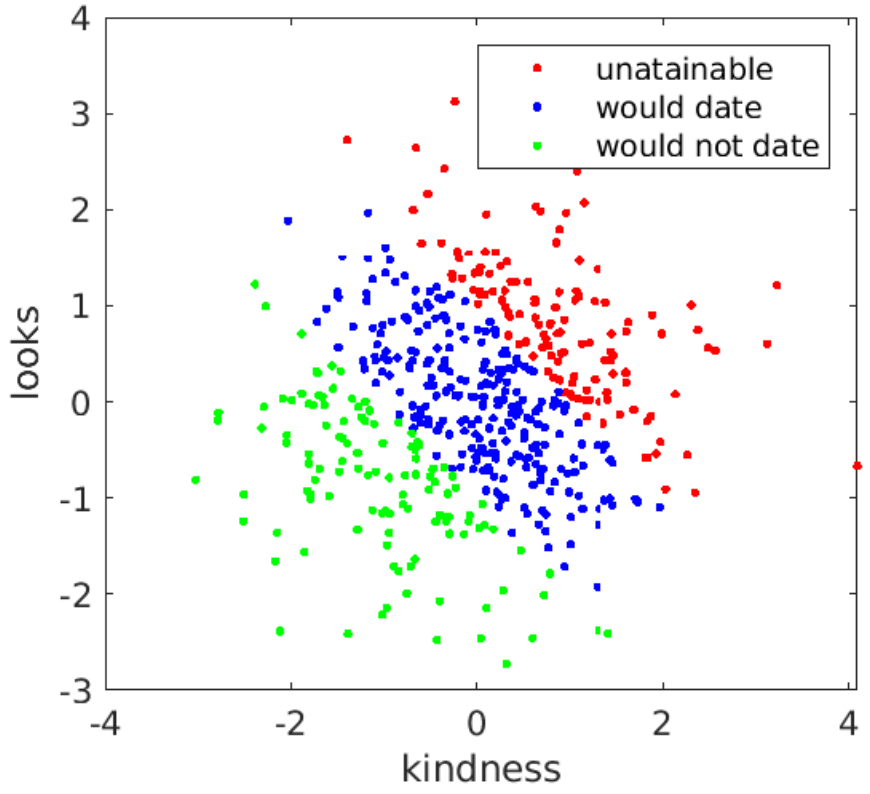
```
% emulate some data that fits model assumption exactly:
N=200;
beta_yx = 3;
beta_xz = 1;
beta_yz = 1;

% emulate individual variation
U_z = randn(N,1); % confounder variation
U_y = randn(N,1); % outcome variation
U_x = randn(N,1); % treatment variation

% emulate linear data-generation process
Z = U_z;
X = beta_xz*Z + U_x;
Y = beta_yz*Z + beta_yx*X + U_y;

% fit the data with linear model while "adjusting", "controlling" for Z
m = fitlm([X Z],Y);
r_yx_z = m.Coefficients.Estimate(2);
r_yz_x = m.Coefficients.Estimate(3);

% fit the naive associations
m = fitlm([X],Y); r_yx = m.Coefficients.Estimate(2);
m = fitlm([X],Z); r_zx = m.Coefficients.Estimate(2);

% show results:
r_yx_z       % estimate of true effect of X→Y
r_yz_x*r_zx  % estimate of bias
r_yx         % estimate of total association
```



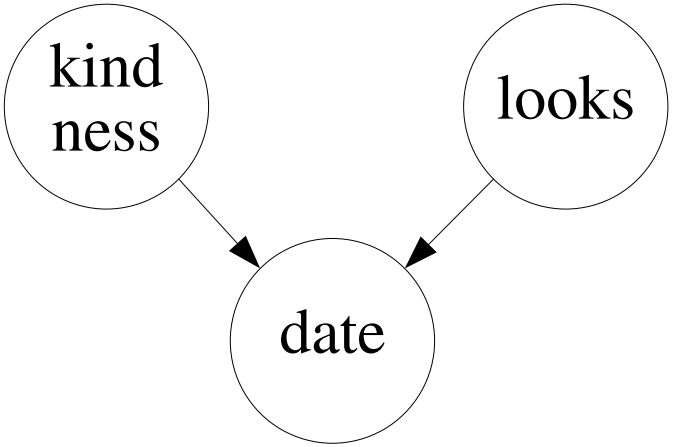
Output on one sample run of this code:
2.9987 estimate of β_{yx} =3
0.5093 estimate of bias
3.5080 total association is exactly the sum

When not to “control”: Berkson's paradox

Example - Selective dating: Looks and kindness are reasonable selection criteria for dating. If you exclude ugly and mean candidates (green), and/or if good looking nice candidates are beyond your reach (red), you will find in the remaining candidates (blue) that looks and kindness are negatively correlated.



collider or selection bias



Conclusion:

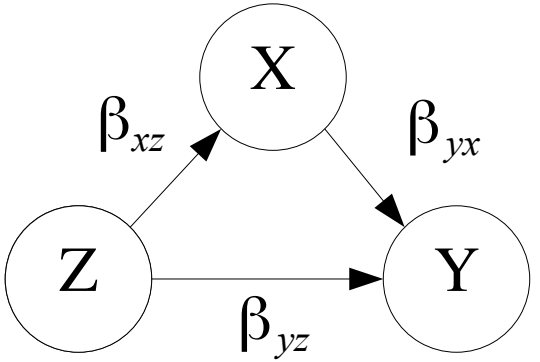
- Controlling for a collider introduces spurious correlation.
- Any kind of selection has a risk of bias.

Mediation

The diagram for mediation is the same as confounding. In this view X mediates indirect effect of Z on Y. Parameters can be found with two OLS:

$$x = \beta_{xz} z + e_x$$

$$y = \beta_{yx} x + \beta_{yz} z + e_y$$



With LS we can also estimate another association

$$y = r_{yz} z + e_y$$

Which reveals the total association (proof: correlate with z):

$$r_{yz} = \beta_{yz} + \beta_{yx}\beta_{xz}$$

Total effect = Direct effect + Indirect effect
Z→Y Z→Y Z→X→Y

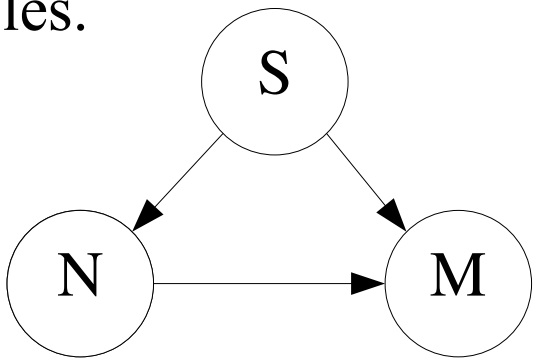
The total effect is the "path integral" from cause to outcome – concept pioneered by Sewall Wright (PNAS, 1920) The Relative Importance of Heredity and Environment in Determining the Piebald Pattern of Guinea-Pigs: "The correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients along all of the paths by which they are connected.

Adjustment formula with binary outcome

Example: A salient stimulus is easy to memorize and might elicit stronger neural activity. Say we have a measure of salience S, strength of neural response N and memory performance M, i.e. correct and incorrect recall. The questions is, does stronger neural activity itself cause better memory, or is it just the results of salience driving both?

We can not $do(n)$, but we can measure all three variables.

$$P(M|do(n)) = \sum_s P(M|n,s)\,P(s)$$



M is binary (correct vs incorrect) so $P(M|n,s)$ could be a logistic model:

$$P(M|n,s) = logistic(\beta_n n + \beta_s s + \beta_0)$$

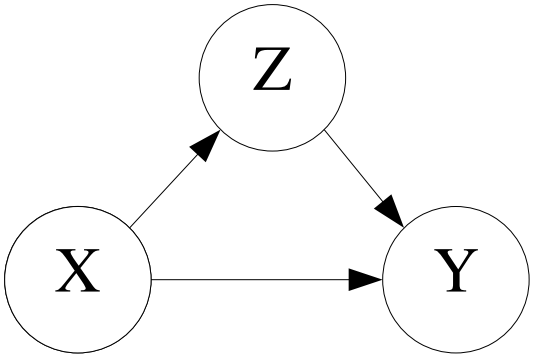
$$logistic(x) = e^x/(1+e^x)$$

i.e. logistic regression, an example of "generalized linear model".

Binary outcome

For a binary outcome Y we use a logistic model:

$$P(Y|X,Z) = logist(\beta_{yx}X + \beta_{yz}Z + \beta_0)$$



The "log odds" is the log of the odds ratio of the two outcomes Y=1 vs Y=0:

$$\log\left(\frac{P(Y)}{1-P(Y)}\right) = logit(P(Y)) = \beta_{yx}x + \beta_{yz}z + \beta_0$$

Coefficients β_{yx} and β_{yz} indicate how much the log-odds ratio of Y increases with an unit increase of x or z.

If z depends linearly on x with a corresponding linear model:

$$z = \beta_{zx}x + e_z$$

then $\exp(\beta_{yx})$ is the direct effect of an unit increase in x, $\exp(\beta_{yz}\beta_{zx})$ is the indirect effect mediated by z, and $\exp(\beta_{yz}\beta_{zx}+\beta_{yx})$ is the total effect on the odds ratio of outcome y. Can use `fitglm()` for both models.

Mediation

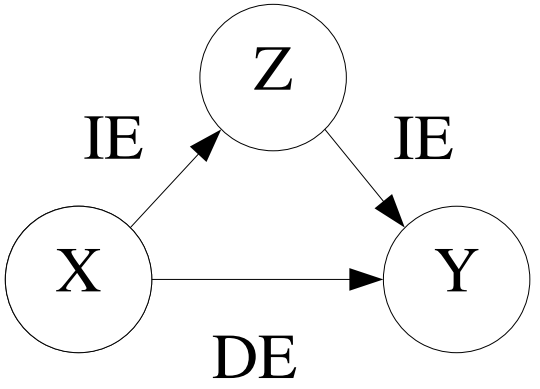
For the general non-linear case, assuming the mediator Z is not confounded then the direct and indirect effects from X → Y are in terms of expected values E(Y):*, **

$$DE = \sum_z \left[E(Y|z, X=1) - E(Y|z, X=0) \right] P(z|X=0)$$

$$IE = \sum_z E(Y|z, X=0) \left[P(z|X=1) - P(z|X=0) \right]$$

And the total effect X → Y is the sum:

$$TE = DE + IE$$



* Pearl, 2009, p. 132
** Treatments X=1 and X=0 can be replaced in these formulas by any two values x and x'.

Comments on adjustment and regression

Quick rules on when to "adjust" or "control" and when not to adjust*

- Adjust for common causes to remove bias. See Simpson's paradox.
- Control for variables that inject variation, even if they do not introduce bias, controlling does improve precision.
- On the other hand, if control variable does not explain any variance, then the extra parameter reduced statistical power, i.e. fitting noise. So only add control variables that are expected to have an effect.
- Do not control for colliders as this can introduce bias, see Berkson's paradox.
- Note that if there is an effect X→Y, the statistical test can also find an effect Y→X. The direction can not be established from the data alone and comes insted from prior knowledge!

Comments on regression with binary variables:

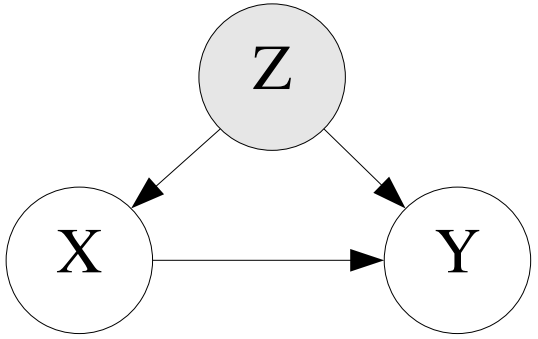
- Regression with binary regressors is the same as taking the difference.
- LS still gives correct β if treatment is binary, but not if the output is binary. Use generalized linear models, e.g. logistic regression in that case.
- In case of binary variable, the p-values from regression are the same as a p-values from 2 sample t-test and effect size β is the same as Cohen's d-prime.

* Cinelli, C., Forney, A., and Pearl, J. (2020). A Crash Course in Good and Bad Controls.

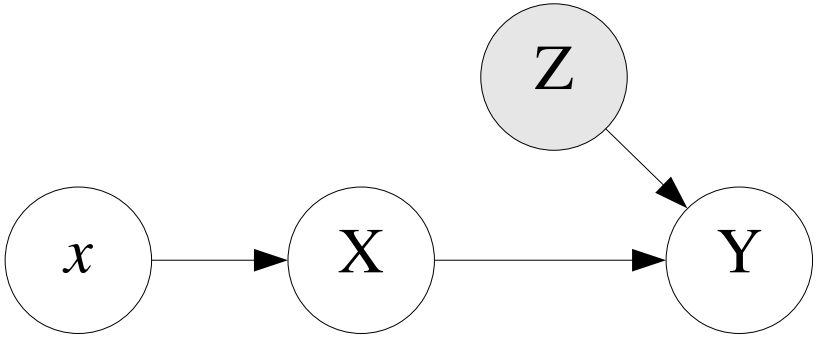
Intervention – do(x)

In a graph, the do(X← x) operation removes the links to X and instead sets the values with some external value x.

Model:



do(x) in Model:



The resulting distribution $P(Y|\text{do}(x))$, which is a function of x, reflects the causal effect of X on Y, *regardless of other effects!*

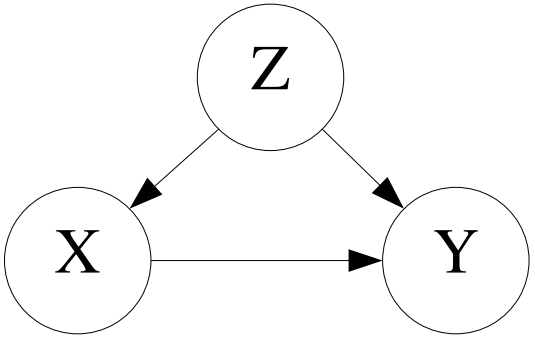
How can we estimate the causal effect $P(Y|\text{do}(x))$?

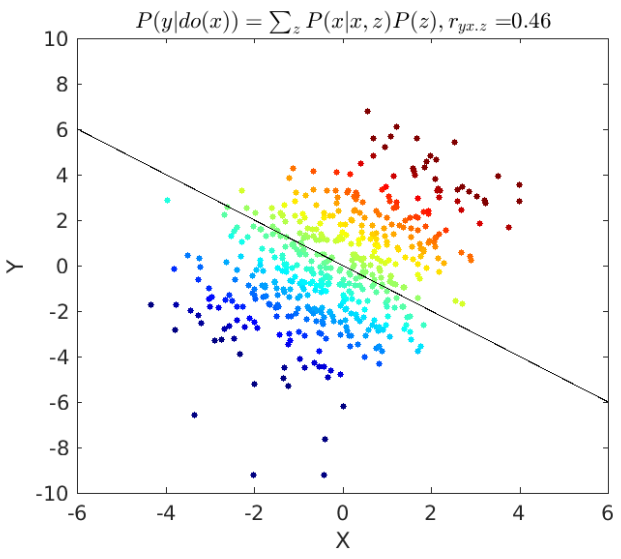
The simplest is to actually set $X \leftarrow x$ for various values of x , i.e. to control x in a "controlled trial". But for some models, that is not needed.

Intervention – Randomized Controlled Trial

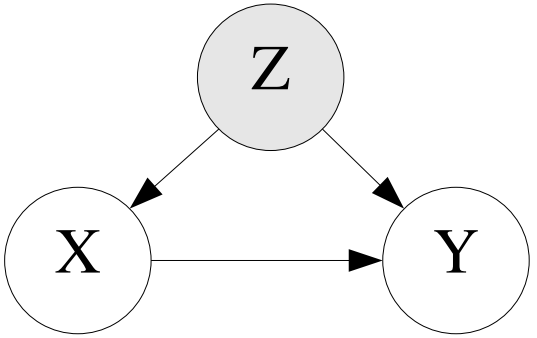
In a RCT variable X is controlled by setting it to a random values.
Example: Say X, Y, Z are all continuous with Z "confounding" the causal effect of X on Y which in this example we set to $\beta_{yx} = -1$

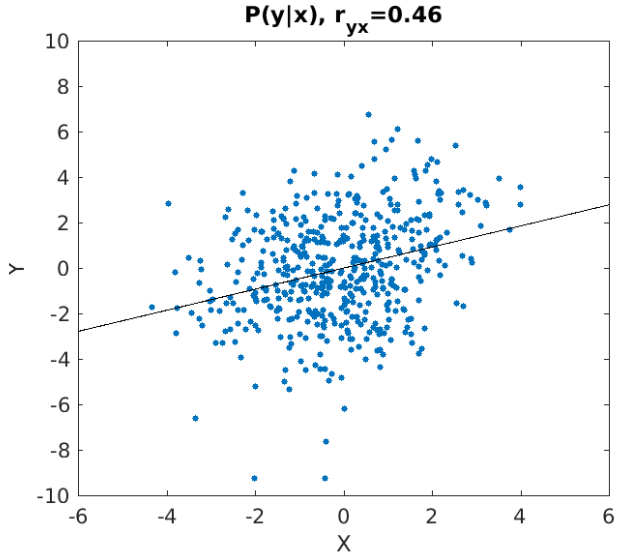
Observed confounder
(easy, use LS)



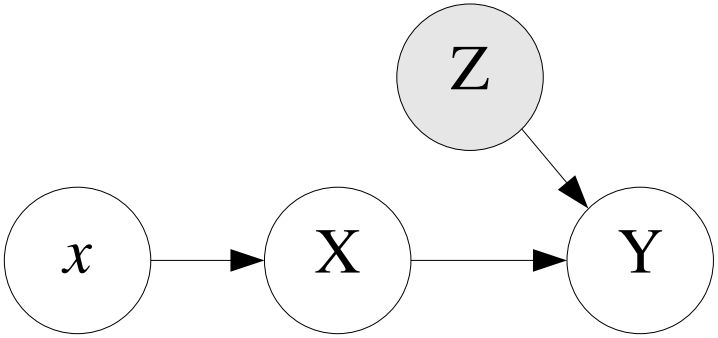


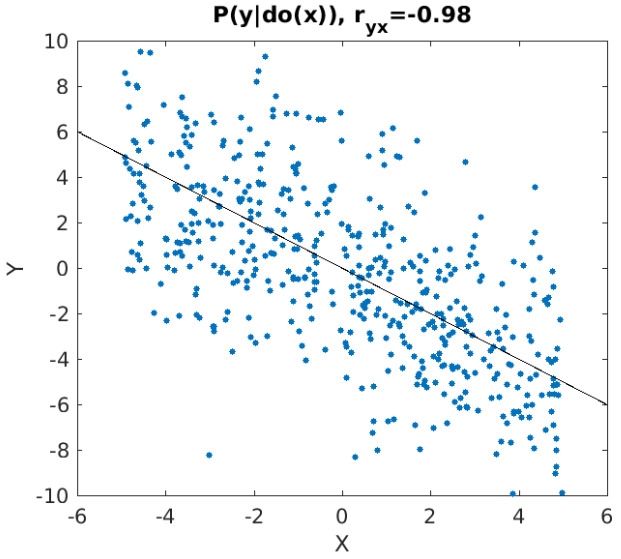
Unbserved confounder
(can't do anything)





RCT to break influence of
all(!) confounders on X

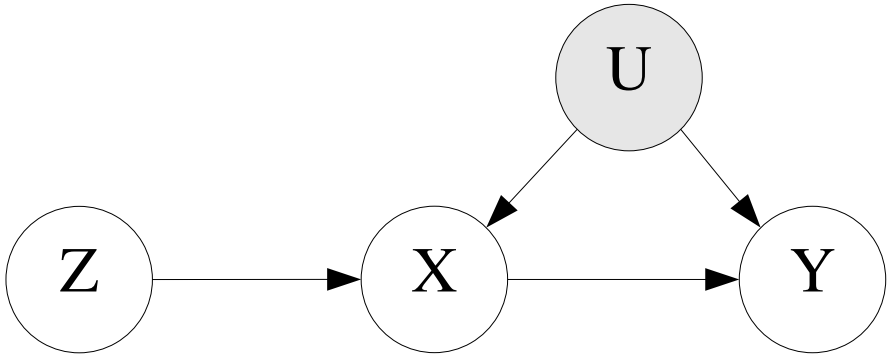




Instrumental variable (natural experiment)

What to do when one can not really fix X. e.g. when the hypothesis is that the intervention causes damage, say smoking. Also, in RCT when participants don't comply with instructions.

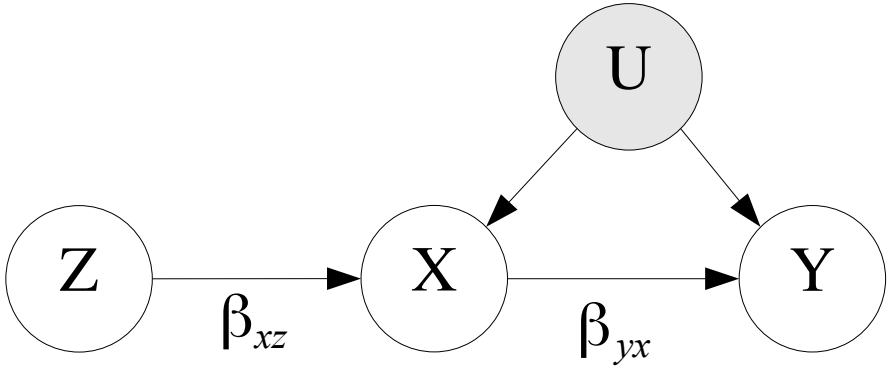
We can use a "instrumental variable" Z



Examples are natural experiments. Imagine Z is assignment to treatment or incentive (school lottery). X is treatment (attending school). Y is outcome (future salary). Problem is that confounders could affect who actually chose/complies with treatment (wealth affects private school enrollment and job prospects).

Instrumental variable – linear case

IV is the case of mediation with no direct effect of Z on Y ("exclusion" assumption) and a confounder U.



The linear case is special. It allows estimate of average causal effect $X \rightarrow Y$ with the same mediation formula (zero direct effect) and is called the Instrumental Variable Estimate*:

$$IVE = \beta_{yx} = \frac{r_{yz}}{r_{xz}}$$

It is still the correct ACE estimate, $P(Y|\text{do}(X))$, despite unobserved confounder U**

* this is the mediation formula for z→x→y with no direct effect of z→y.
**Pearl 2009, Causality, Chapter 5. IVE dates back to ~1928 Philip and Sewall Wright (father and son)

Instrumental variable (natural experiment)

If not linear, then the story is more complicated. An important nonlinear case is that of natural experiment or RCT with incomplete compliance. Individuals are assigned (binary Z) to treatment (binary X) and my or may not comply with the assignment.

We could be asking for a few different things:

- ACE = Average causal effect of treatment X on outcome Y: $P(Y|\text{do}(x))$
- ETT = Effect of Treatment on the Treated: $P(Y|\text{do}(x), X=1)$. This is a conterfactual: We need to know what would have happened for $X=0$ on subjects who actually received treatment $X=1$.
- ITT = Intention To Treat analysis. $P(Y|\text{do}(z))$. Everyone is analyzed even if they did not comply with assignment. This captures effect of assignment on outcome, rather than effects of treatment. Effect on general population will vary with the incentives.*
- PP = Per Protocol analysis, $P(Y|X)$: PP=E[Y|X=1]-E{Y|X=0], i.e. diffence in outcome between those that got treatment and those that did not, regardless of assignment.*

* ITT analysis is generally thought to be less susceptible to bias from non-complicance and dropout as compared to PP analysis. However, bias can go either way: Say drug makes some participants sick. If these drop out, then the PP analysis will make drug look better. If they stay in study, but stop taking the drug as prescribed, then the ITT analysis will look better for the drug.

Instrumental variable (natural experiment)

If both Z and X are binary (X no longer linear with Z), the formula reads

$$IVE = \frac{r_{yz}}{r_{xz}} = \frac{E\left(Y|z=1\right) - E\left(Y|z=0\right)}{E\left(X=1|z=1\right) - E\left(X=1|z=0\right)}$$

Expression to the right is called local average treatment effect (LATE)**. In words:

$$\text{LATE} = \frac{\text{Effect of assignment on outcome (i.e. ITT)}}{\text{Fraction that followed the assignment}}$$

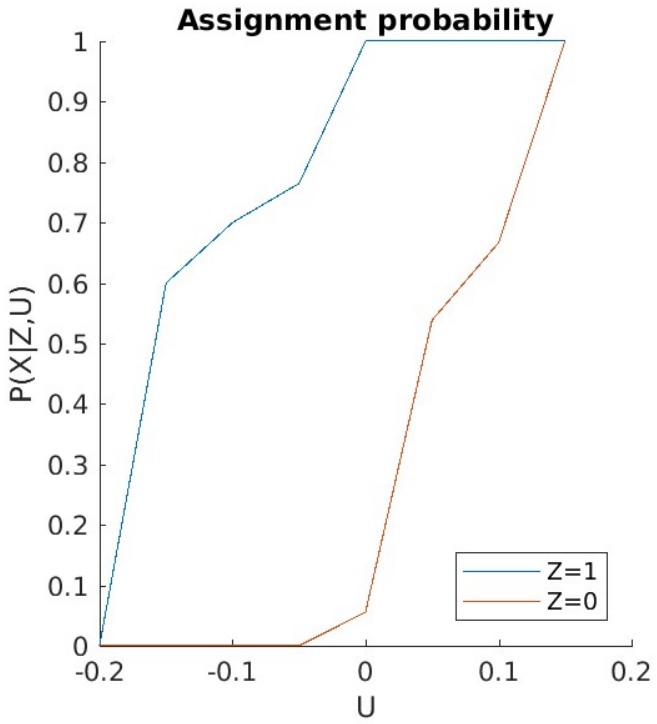
It is numerically identical to the *IVE*. The neat thing about using regression coefficients formulation is that you can add other observed covariates W:

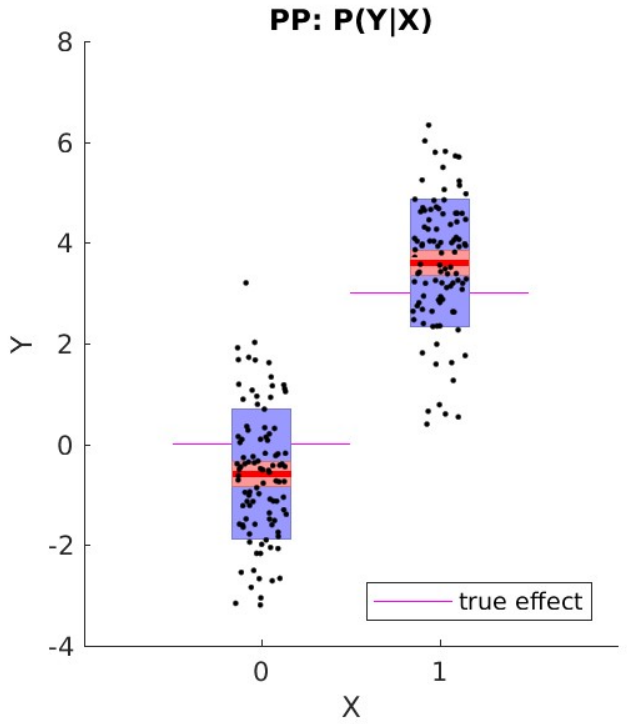
$$IVE = \frac{r_{yz.w}}{r_{xz.w}}$$

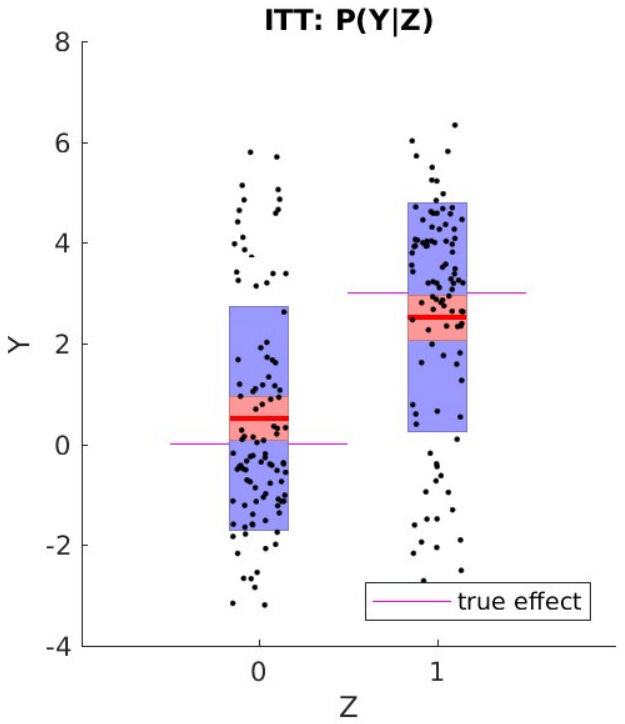
** Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996)

Partial compliance example

Example, say X is private school attendance, Y future earnings, Z school admissions, U is household income. Simulated data following model in slide 15...







N=200
True Average Causal Effect=3
Intention to Treat analysis=2.0025 (under estimates effect of schooling)
Per Protocol analysis=4.1906 (over estimate effect of schooling)
LATE estimate=3.2825 (closer, but not guaranteed to be exact, even for large N)
ACE Regression estimate=3.2457 (used U as covariate – perfect knowledge)

Natural experiment theoretical results

Theoretical results and a few special cases:

LATE = ETT if no intruders i.e. $P(x=1|z=0) = 0$ *

LATE = ITT if no intruders and no dropouts i.e. $P(x=1|z=1) = 1$

LATE = ACE if no defiers ("monotonicity" assumption), i.e. Defier is a person that always does the opposite of assignment.**, +

ACE can not be estimated, in general, with any more accuracy than rate of noncompliance: $P(X=1|Z=0)+P(X=0|Z=1)$***

Angrist and Imbens received Nobel price in economics 2021 on this topic.

* Pearl, Causality, 2009, Equation 8.20
** Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996).
*** Pearl, Causality, 2009, Equation 8.17
+ Critique from Pearl is that assumption can not be verified based on observations.

Noncompliance in controlled experiments

Observed behavior

	Assigned to treatment	Assigned to control
Treated	Compliant	Intruder
Not treated	Dropout	Compliant

Type of participants directly observable in the data.

Hypothetical behavior

	Assigned to Treatment	Assigned to control
Complier	Treated	Not treated
Always taker	Treated	Treated
Never taker	Not treated	Not treated
Defier	Not treated	Treated

Type of participants can not be determined from the data. Only if we assume there are no defiers, can one estimate the fraction of the other three.

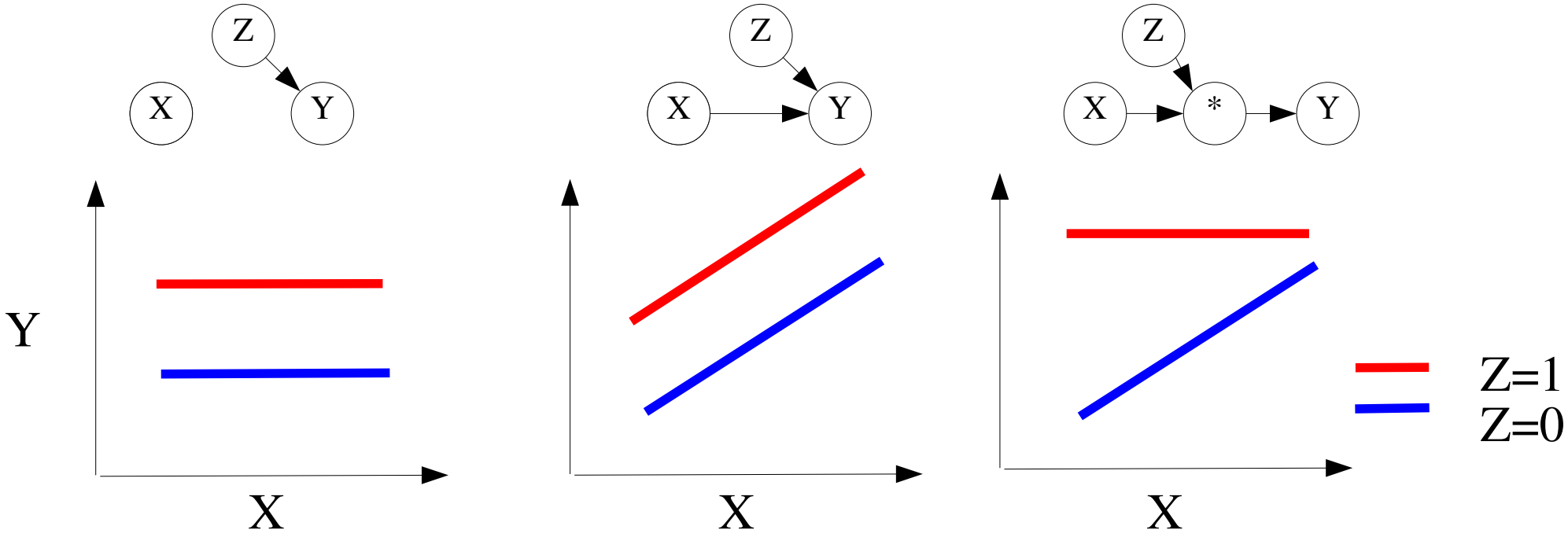
Hypothetical behavior requires "counterfactual" or "potential outcome" reasoning, which is integral to causal inference, but we will leave out of this lecture.

Assignment

Estimate LATE, ITT estimate, PP estimate, and ACE from variables X, Y, Z, U in 'LATE_example.mat' and make figures as in slide 27. Assume the causal graph is as in slide 23. This data was generated the same way with the same effect sizes, but will not give the exact same values due to random sampling.

Interaction

So far the effect of x on y, was independent of z. Interaction is when it depends on the value of z



Single linear model able to identify all (provided enough data)

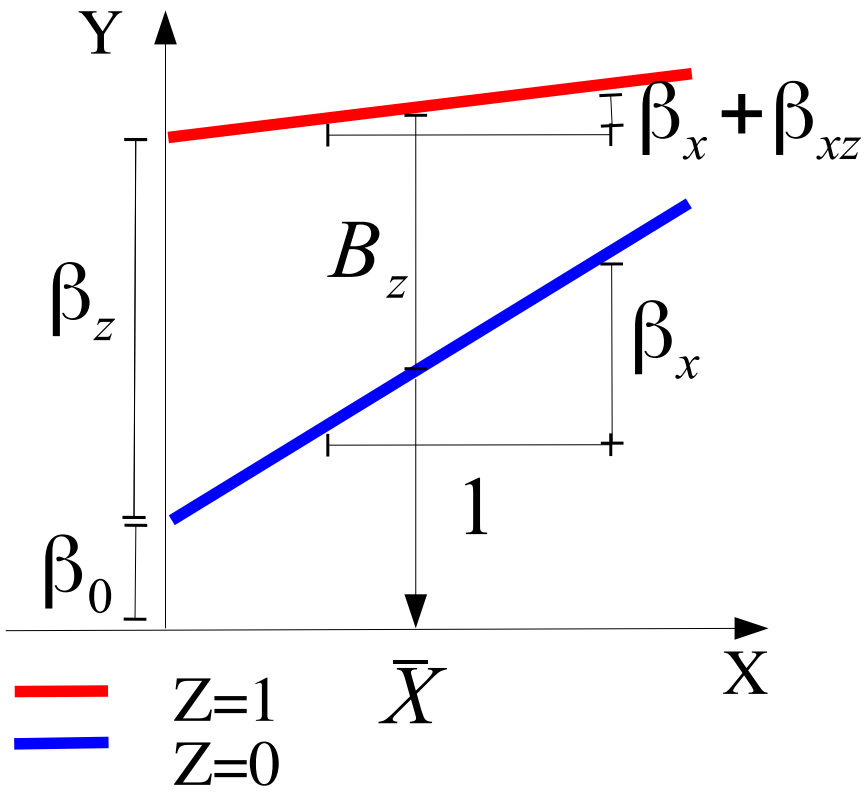
$$y=\beta_o+\beta_x x+\beta_z z+\beta_{xz} x*z+e_y$$

$$\beta_z\neq 0 \qquad \beta_x\neq 0,\ \beta_z\neq 0 \qquad \beta_x\neq 0,\ \beta_z\neq 0,\ \beta_{xz}\neq 0$$

If Z is a mediator (X→Z) it may also interacts with X and one can used the same ordinary least squares (See VanderWeele, 2015, section 2.2)

Interaction interpretation

$$y=\beta_o+\beta_x x+\beta_z z+\beta_{zx} x*z+e_y$$



In the case of a significant interaction, non-zero "simple main effect" β_z does not mean that z has an effect in the actual range that is observed for x. The size of the treatment z on the average person (with average $x=\bar{x}$) is

$$B_z=\beta_z+\beta_{zx}\bar{x}$$

This is sometimes called the "overall main effect".

If we subtract the mean of x prior to running the model then $B_z=\beta_z$ And we can interpret p-value for the simple main effect as the significance of the effect of treatment z on the average person.

When can we just use observations?

Is there a formal way to figure all of this out? What about non-linear relationships. Yes, rules of probability plus rules of do Calculus*

Rule 1 (insertion/deletion of observation)

$$P(y|do(x),z,w)=P(y|do(x),w)\,if\,(Y\perp Z|X,W),G_{\overline{X}}$$

Rule 2 (action/observation exchange)

$$P(y|do(x),do(z),w)=P(y|do(x),z,w)\,if\,(Y\perp Z|X,W),G_{\overline{X}Z}$$

Rule 3 (insertion/deletion of action)

$$P(y|do(x),do(z),w)=P(y|do(x),w)\,if\,(Y\perp Z|X,W),G_{\overline{X}\overline{Z}}$$

$G_{\overline{X}}$ - Graph where arrows into X have been removed
$G_{\underline{X}}$ - Graph where arrows out of X have been removed
$(Y\perp Z|X),G$ - Means that Y is independent of Z given X in graph G, i.e. conditional independence.

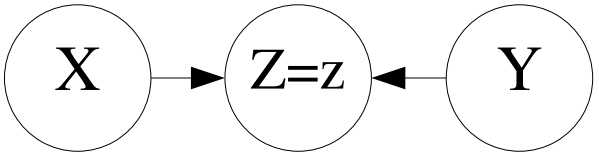
*Pearl, Causality, 2009 page 85

Conditional independence

"Y independent of Z given X in graph G" What does that mean?

Conditioning is like freezing at Z=z, so Z will no longer cause a variation in X or Y, and they will no longer co-vary, so they become independent upon conditioning. True for "forks" and "chains"

However, the opposite will happen for a "collider":



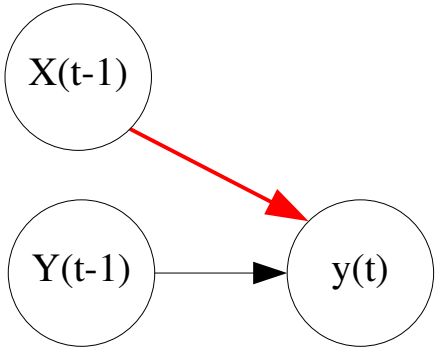
When we freeze Z=z, what was independent now becomes dependent, because any change in X needs to be compensated by a change in Y so that Z remains unchanged at Z=z.

Pearl suggests to think of conditioning as a valve for information flow, opening if for colliders and closing it for chains and forks.

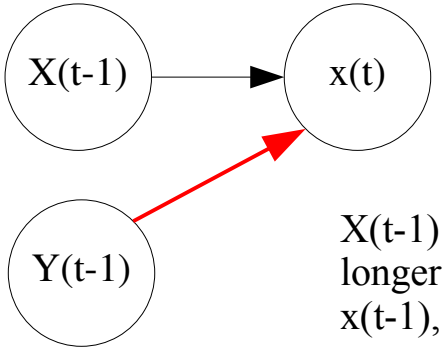
Granger "Causality" – temporal precedence

Same rationale as linear model, but now taking history into account. The causal assumptions are captured by this graph*:

X "Granger causes" Y



Y "Granger causes" X



X(t-1) here can stand for longer history with Q delays: x(t-1), x(t-2) … x(t-Q).

The word "causes" is a misnomer. Granger himself prefers to say X is temporally associated Y. Better terminology is X forecasts Y. Because we have a history, we can say that X forecasts Y, or Y forecasts X, or both. So we can establish a direction of the effect, which we could not do without history.

* Granger received Nobel Prize in 2003.

Thanks to Behtash Babadi for a lot of help with this material on Granger Causality! All errors mine.

Granger "Causality" – test for significance

The test measures if the red arrow is non-zero, by fitting a two linear model, e.g. for X→ Y fit the "full" a "reduced" models

$$y(t) = a\,Y(t-1) + b\,X(t-1) + e_f$$

$$y(t) = a\,Y(t-1) + e_r$$

In system identification these are called ARX and AR model, respectively (here with filter length Q=1). Parameters *a,b* can be found with ordinary least squares. Various significance test exists, eg. Wald Test or log-likelihood ratio (LR) test. For $e(t)$ normally distributed this is the log of the sum of squares of the residual (sum over T samples), or *Deviance* (defined as 2*LR):

$$D = T \log\left(\sum_t e_r^2(t) / \sum_t e_f^2(t) \right)$$

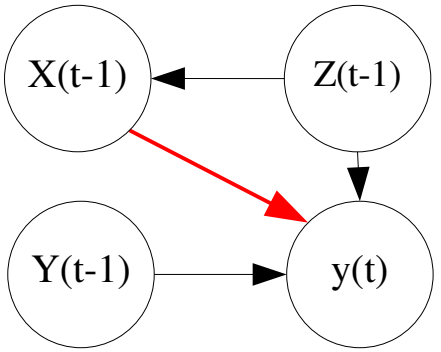
For large T, this D follows a Chi-square distribution with Q degrees of freedom (here the length of filter $\boldsymbol{b}$). So the p-value for significance is:

$$p = 1 - chi2cdf(D, Q)$$

This test assumes that innovation $e(t)$ is normal i.i.d. Independence is why $Y(t\text{-}1)$ is included in the model. Identical implies that it is stationary. In total, it requires that signals are normal and WSS. For small T this is not a good test.

Granger "Causality" – confounders

In case of a common-cause confounder Z(t-1) …



… one can simply add it to the regression as before.

$$y(t) = a\,Y(t-1) + c\,Z(t-1) + b\,X(t-1) + e_f$$

$$y(t) = a\,Y(t-1) + c\,Z(t-1) + e_r$$

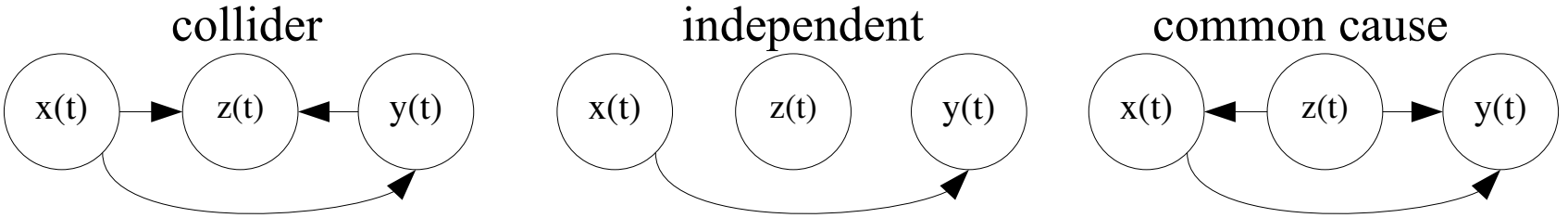
The statistical test remains the same.

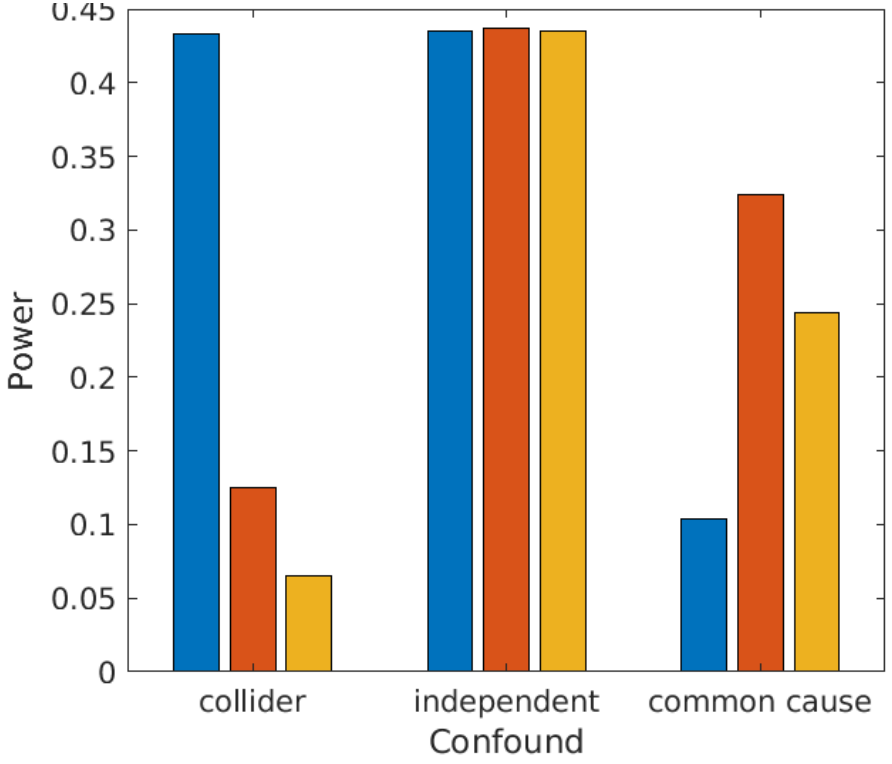
Granger "Causality" – let's get real

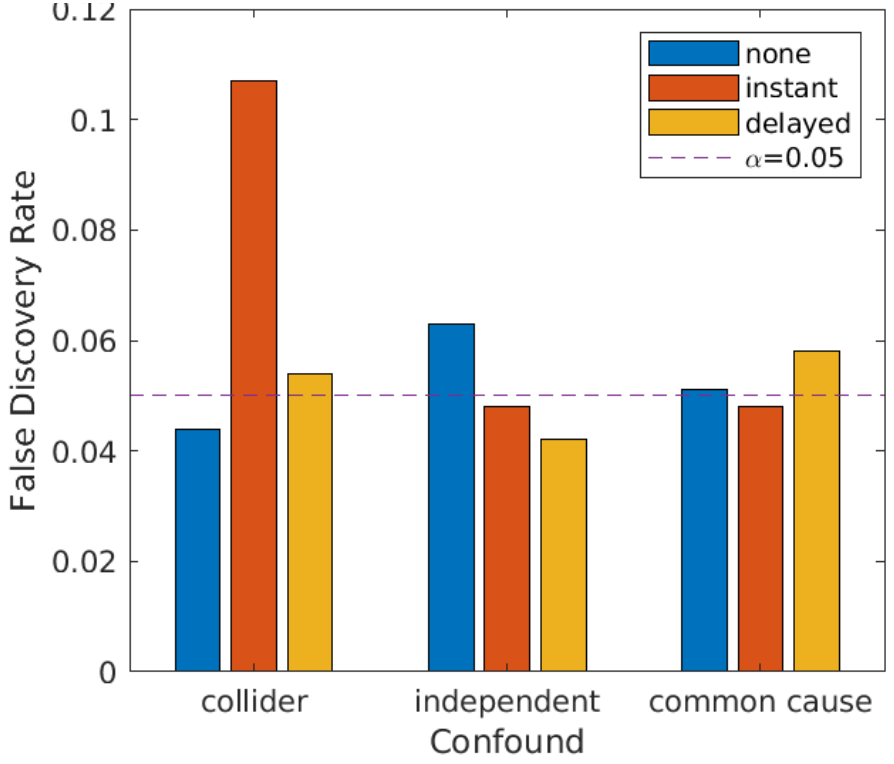
```
function pval = granger_test(x,y,Q,z,control)
% Tests if time series Y can be linearly predicted from Q preceding
% values in time series X while controlling for Z.
% (c) Lucas Parra, May 9, 2022. Use at your own risk.

X = toeplitz(x(Q:end-1),x(Q:-1:1));
Y = toeplitz(y(Q:end-1),y(Q:-1:1));
switch control
    case 'instant'
        Z = toeplitz(z(Q+1:end),z(Q+1:-1:2));
    case 'delayed'
        Z = toeplitz(z(Q:end-1),z(Q:-1:1));
    case 'none'
        Z = [];
end
bf = [X Y Z]\y(Q+1:end); % full model
br = [  Y Z]\y(Q+1:end); % reduced model
SSf = sum((y(Q+1:end)-[X Y Z]*bf).^2);
SSr = sum((y(Q+1:end)-[  Y Z]*br).^2);
[T,P] = size(X); % making sure numbers are correct for stats
D = T*log(SSr/SSf); % Deviance (2*log-likelihood ratio)
pval = 1-chi2cdf(D,P);
```

Granger – risk/benefit of controlling for z

Effects of different ways of "adjusting" for variable Z(t) (none, instant, delay). "instant" includes the current sample Z(t) as regressor, "delay" only includes the history Z(t-1). In the data generation process (equation error model) the variable Z had one of 3 roles:







Granger "Causality" – comments and caveats

If there is an instant effect, the direction of GC can not be established. Both directions will show up in the result.

In case that there is a collider, conditioning on it with instant delay can introduce strong collider bias, i.e. find GC where there is none. Conditioning on the history of the collider is safe.

Flip side, if there is an instant common cause as confound, conditioning on it improves power to detect true GC. So, in general, use instant de-confounding only if you are very sure about cause and effect of the de-confounding variable.

For smaller T, rate of false discovery of GC goes up i.e. the test finds a GC link where there is none. Can partially fix this by using $T'=T-P$ in Deviance definition. Trust GC only for large T.

Don't trust the test if the data generating model does not match the ARX model, e.g. when the dynamic variables are not directly observed, as in the output-error model.

The difference of GC modeling over conventional linear modeling (fitlm() in matlab) is that the likelihood ratio test of GC does a single test for the entire history, whereas linear modeling tests each delay. Thus, GC test provides higher statistical power.

All bets are off if the effect is non-linear or the error is not normally distributed, or signals are not wide sense stationary, e.g. heteroscedastic, as in the stock market.

Assignment

Determine which causal (forecast) graphs are consistent with the data in ‘granger_example.mat’. Variables there are a, b, c. Compare with the truth information reflecting how the data was generated. Are the Granger estimates consistent, are there false discoveries, are there missing links?