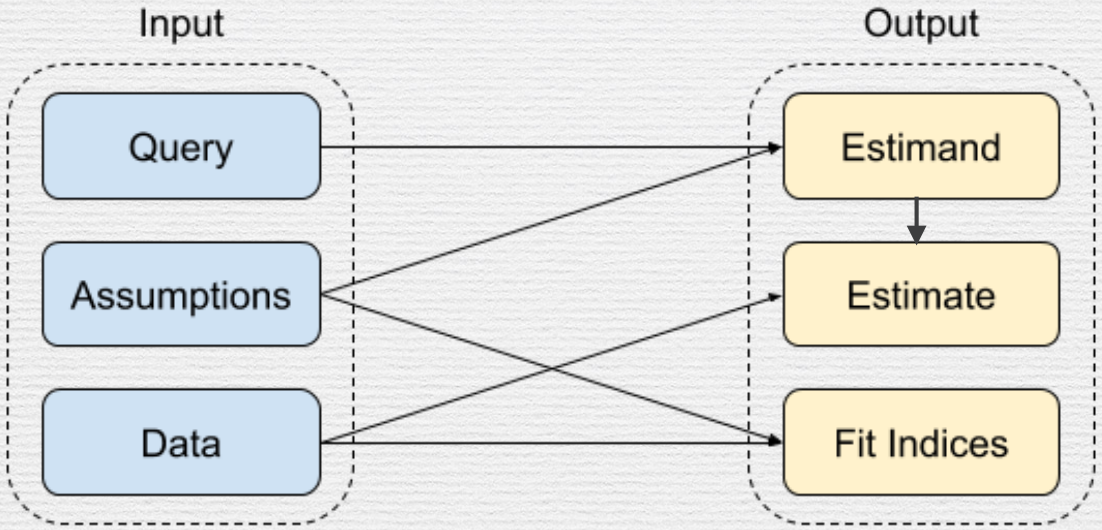
Lec 3

Causal Inference

- The easiest level involves questions about prediction (**How many shoppers buy pencils and pens?**)
- The next level involves questions about interventions (**How many shoppers would buy a pen if I doubled the price of a pencil?**)
- The hardest questions are counterfactuals (**How much profit would I have made yesterday on pencils, if I had doubled the price of pens?**)

Causal Inference engine

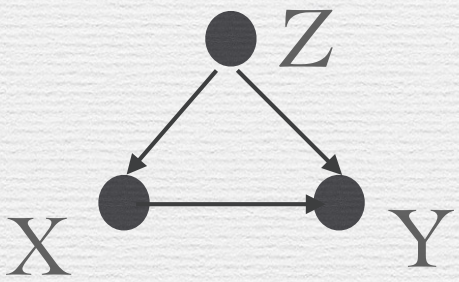
The process of discovering causal relationships between variables. This step can also be thought of as a method of codifying assumptions about the system.



- **Estimand**, which is a recipe or formula to answer the query based on assumptions
- **Estimate**, which is the answer to the query. This is computed from the estimand and after receiving the Data
- **Fit indices**, which give a measure of compatibility between the data and the assumptions
- Graphical models serve as a language for representing what we know about the world, counterfactuals help us to articulate what we want to know, while structural equations serve to tie the two together in a solid semantics

Example

- Query: causal effect of X on Y written Q = P(Y| do(X))



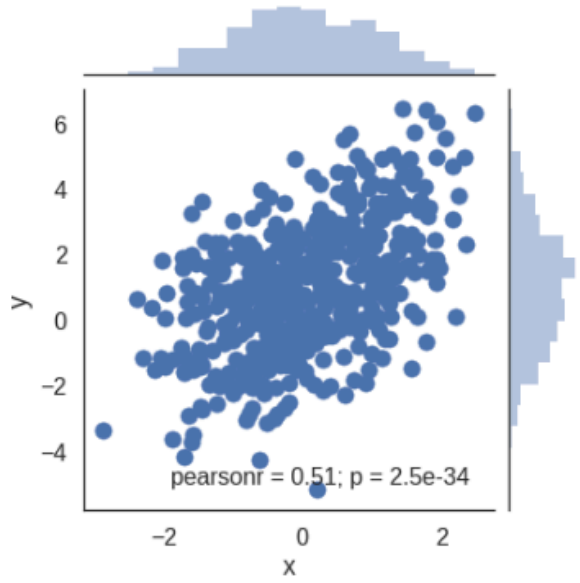
- Assumption graph, Z affects X and Y
- Data sampled from P(X,Y,Z)
- Estimand E = \sum_z P(Y|X,Z) P(Z) It defines a property of P(X,Y,Z)that, if estimated, would provide a correct answer to our Query.
- Estimate, can be produced by any number of techniques that produce a consistent estimate of E, from finite samples of P(X,Y,Z). For example, the sample average (of Y) over all cases satisfying the specified X and Z conditions, would be a consistent estimate.

Details

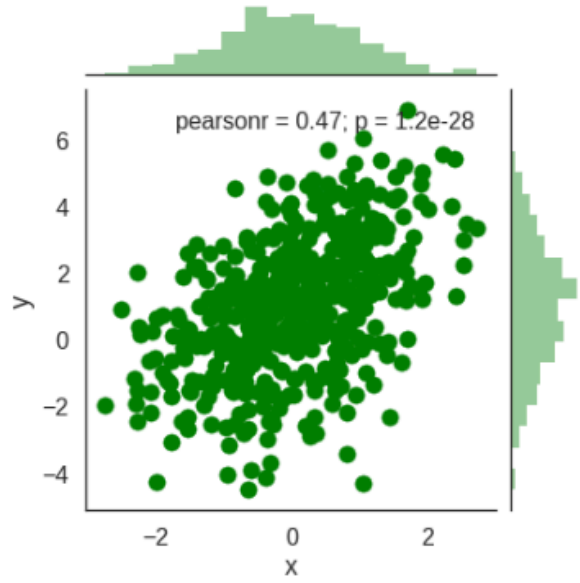
- observational p(y|x): What is the distribution of Y given that I **observe** variable X takes value x. It is a conditional distribution which can be calculated from $p(x,y,z,\ldots)$
- Interventional p(y|do(x)) What is the distribution of y if I were to **set** the value of X to x. This describes the distribution of Y I would observe if I intervened in the data generating process by artificially forcing the variable X, but otherwise simulating the rest of the variables according to the original process that generated the data. (note that the data generating procedure is NOT the same as the joint distribution p(x,y,z…) and this is an important detail).

Toy Example

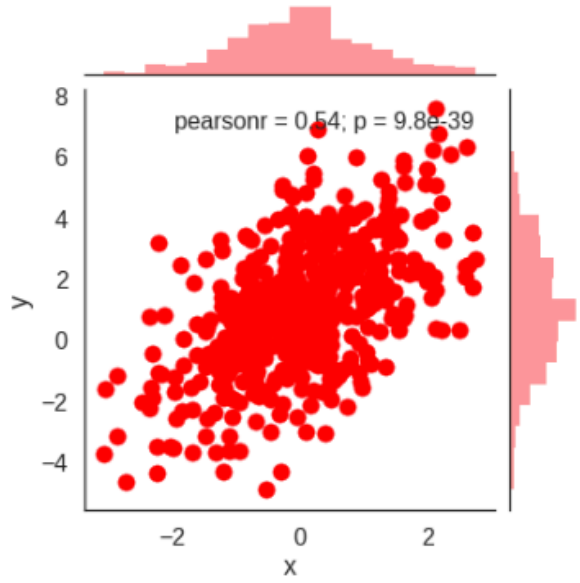
x = randn()
y = x + 1 + sqrt(3)*randn()



y = 1 + 2*randn()
x = (y-1)/4 + sqrt(3)*randn()/2



z = randn()
y = z + 1 + sqrt(3)*randn()
x = z

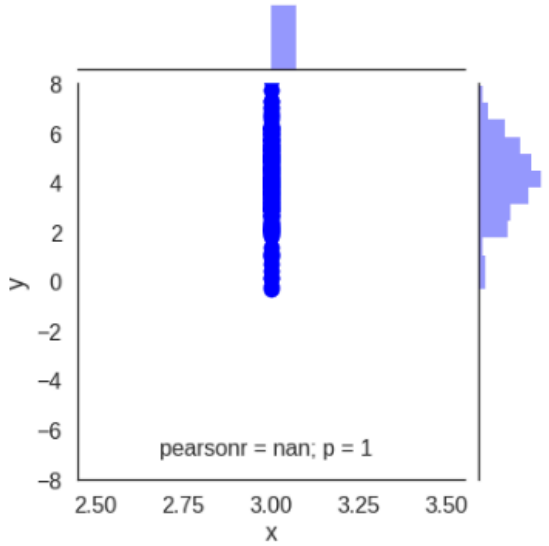


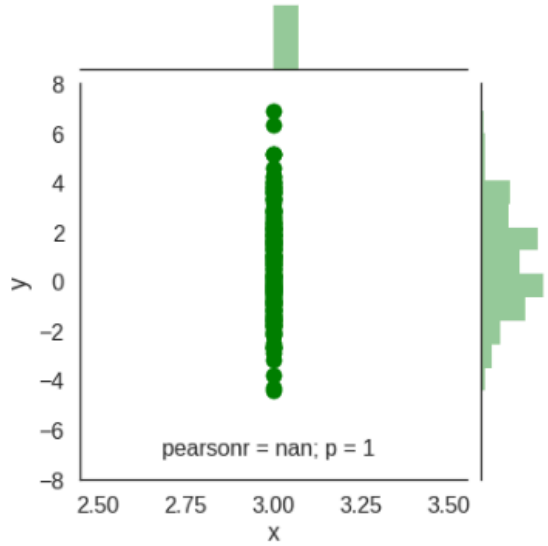
Intervention — do()

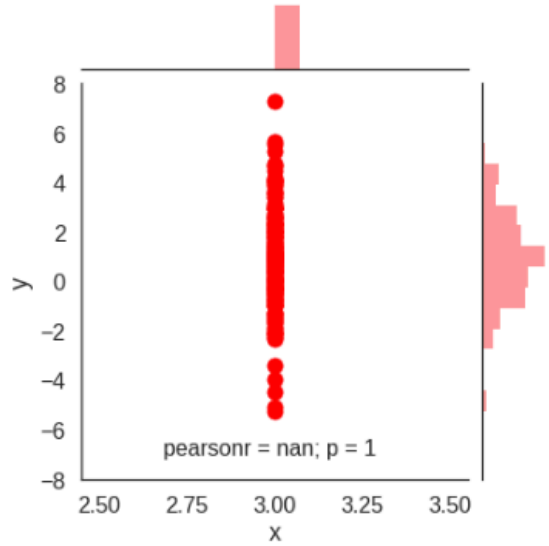
x = randn()
x = 3
y = x + 1 + sqrt(3)*randn()
x = 3

y = 1 + 2*randn()
x = 3
x = (y-1)/4 + sqrt(3)*randn()/2
x = 3

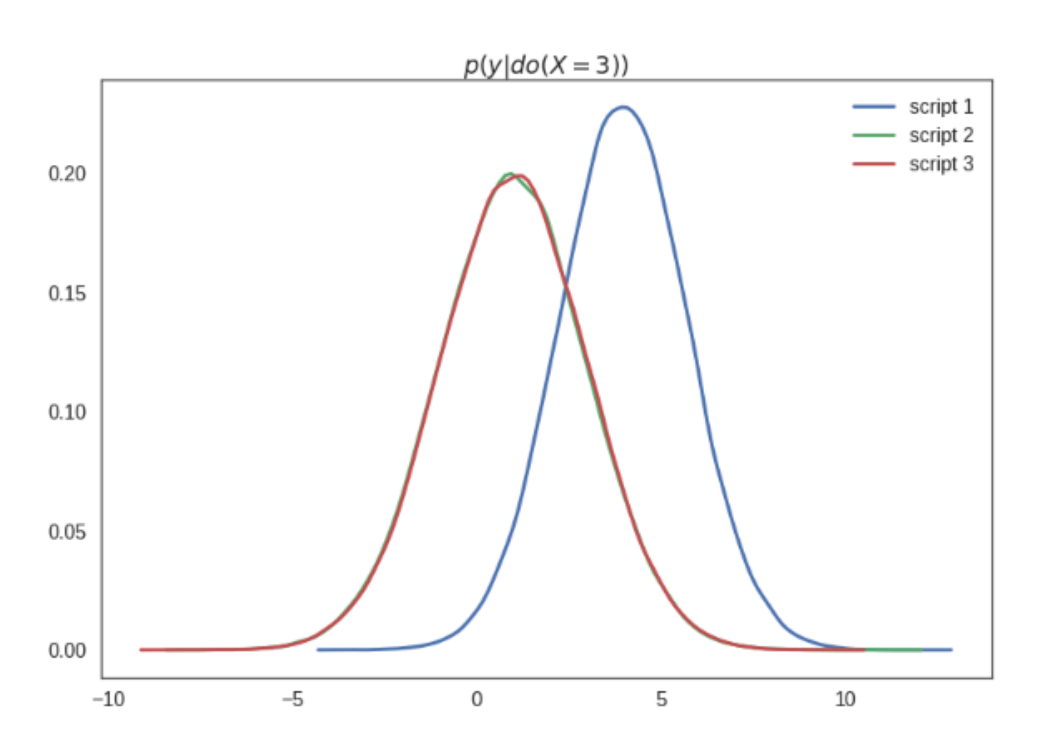
z = randn()
x = 3
x = z
x = 3
y = z + 1 + sqrt(3)*randn()
x = 3

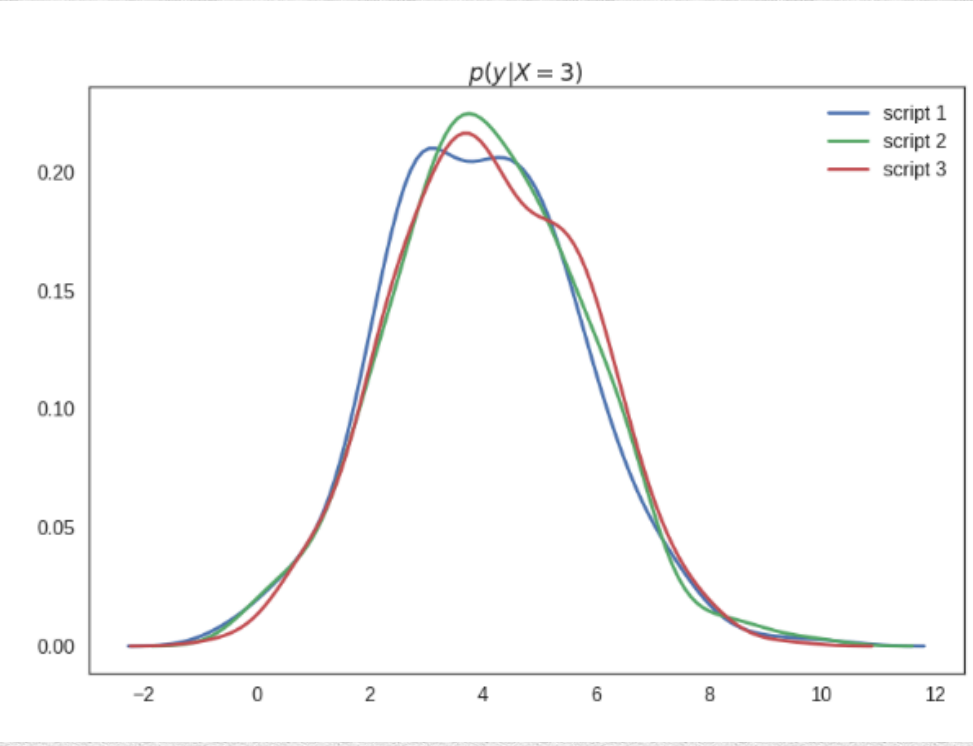






Scripts behave differently under intervention

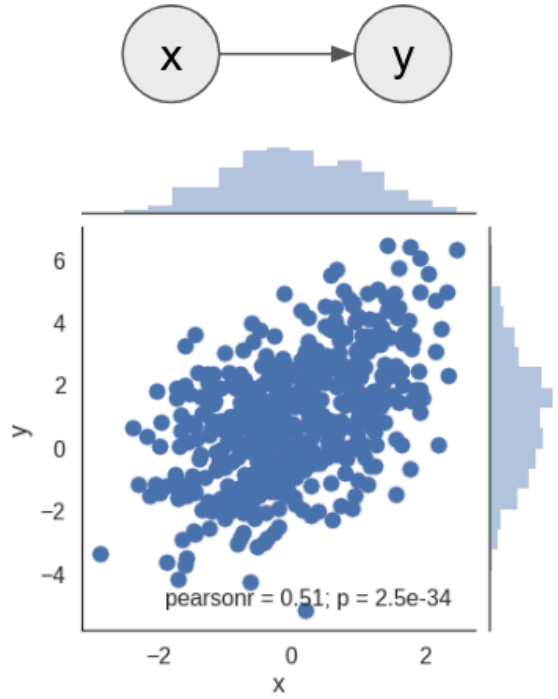




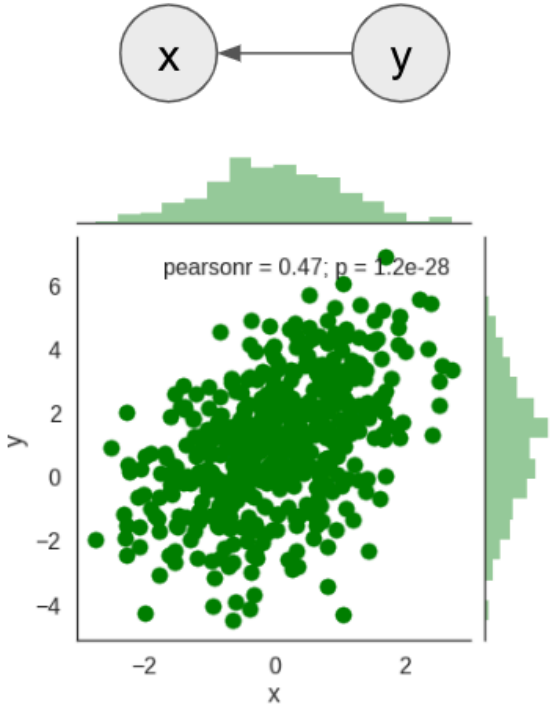
scripts are indistinguishable when you only look at the joint distribution of the samples they produce, yet they behave differently under intervention.

Causal Diagrams

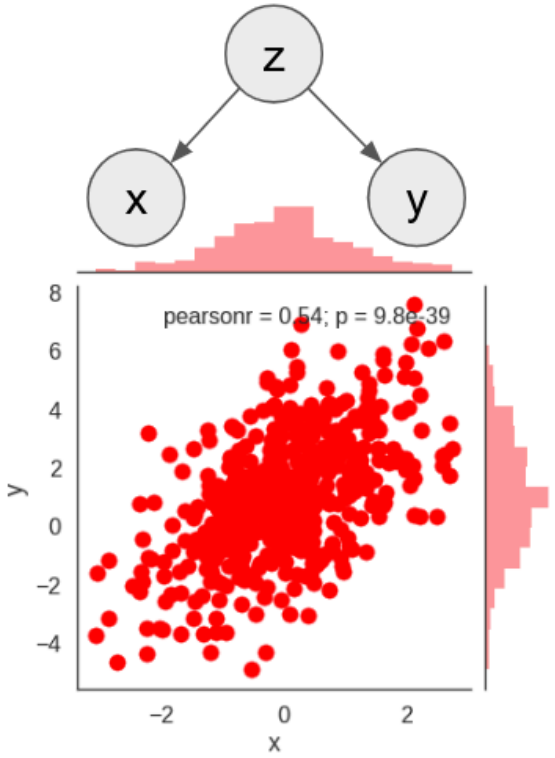
x = randn()
y = x + 1 + sqrt(3)*randn()



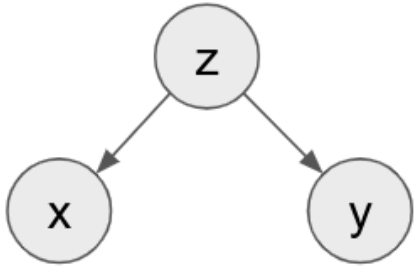
y = 1 + 2*randn()
x = (y-1)/4 + sqrt(3)*randn()/2

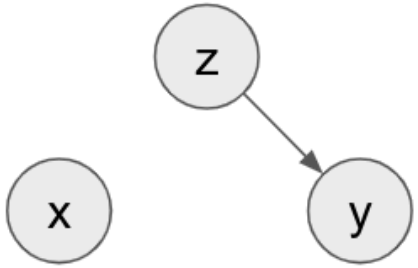


z = randn()
y = z + 1 + sqrt(3)*randn()
x = z



Intervention Causality



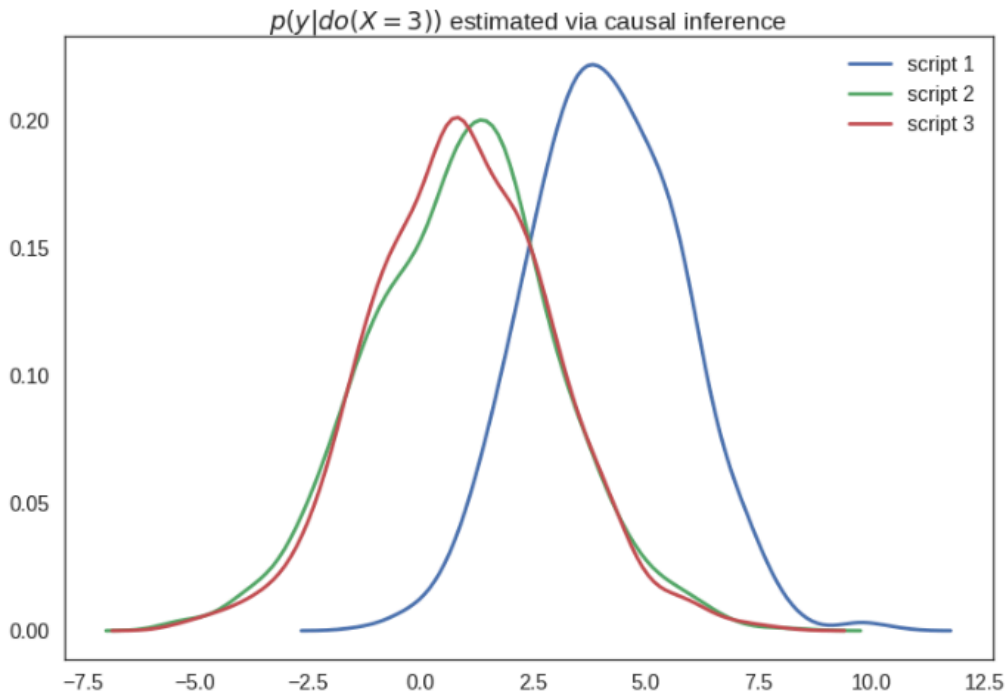


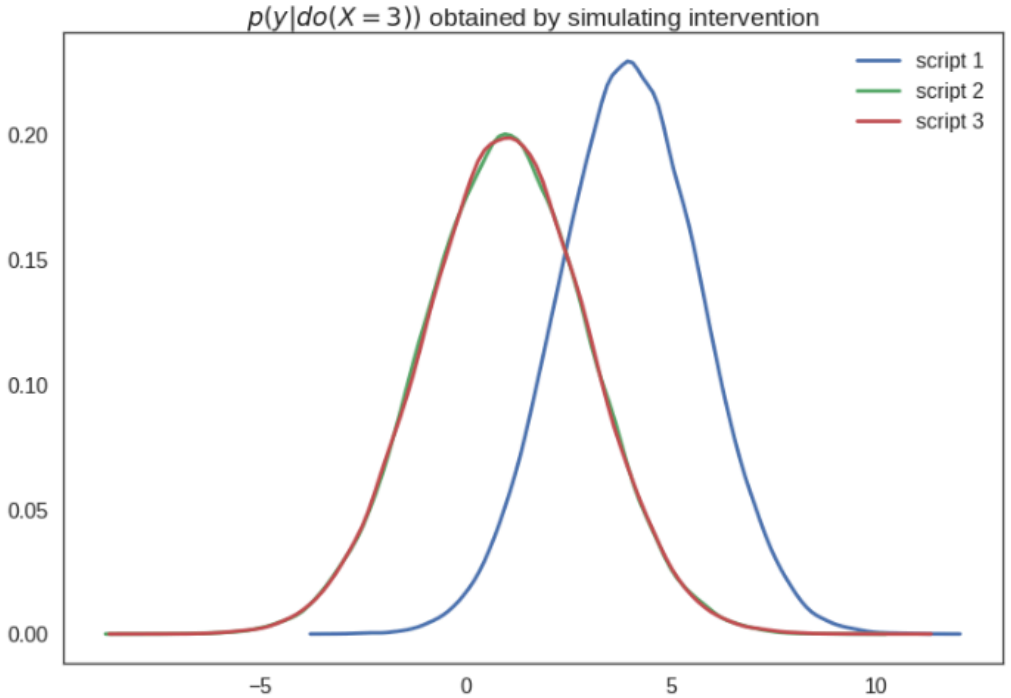
$P(y|do(X)) = p(y|x)$

$P(y|do(X)) = p(y)$

$P(y|do(X)) = p(y)$

- The causal diagram allows us to predict how the models will behave under intervention, without carrying out the intervention

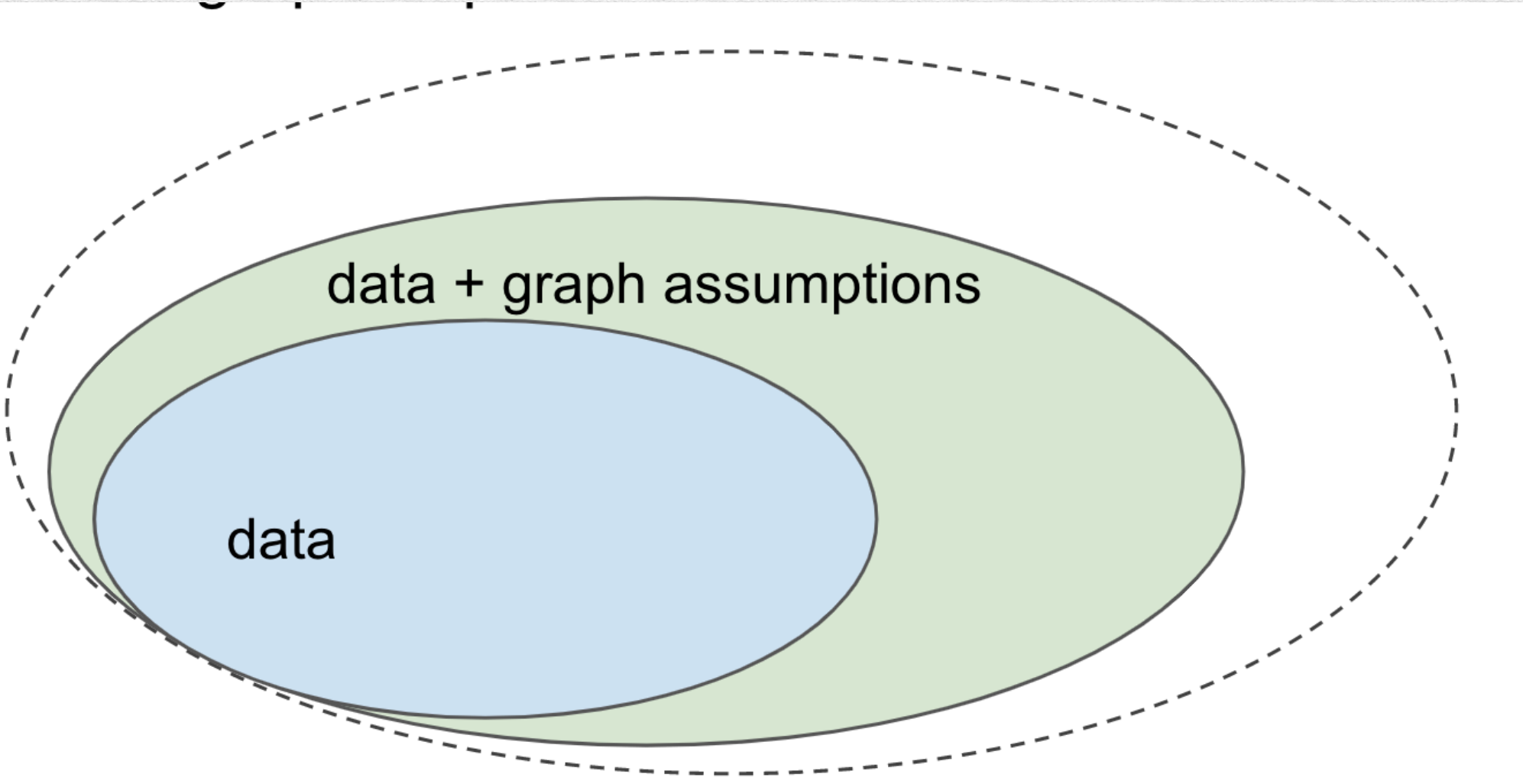




estimated from joint + causal diagram

actually running the experiment

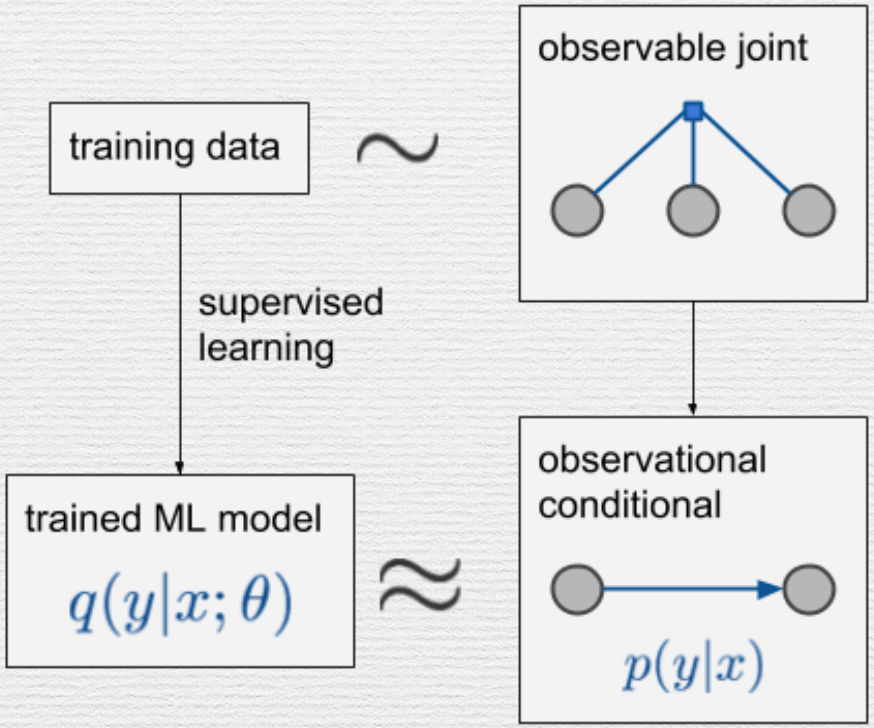
- Data iid samples from joint distribution like images in a dataset
- if you want to make predictions about how the system you study would behave under certain interventions or perturbations, you typically won't be able to make such infernces based on the data you have.
- However, if you complement your data with causal assumptions encoded in a causal diagram - a directed acyclic graph where nodes are your variables - you can exploit these extra assumptions to start answering these questions, shown by the green area.

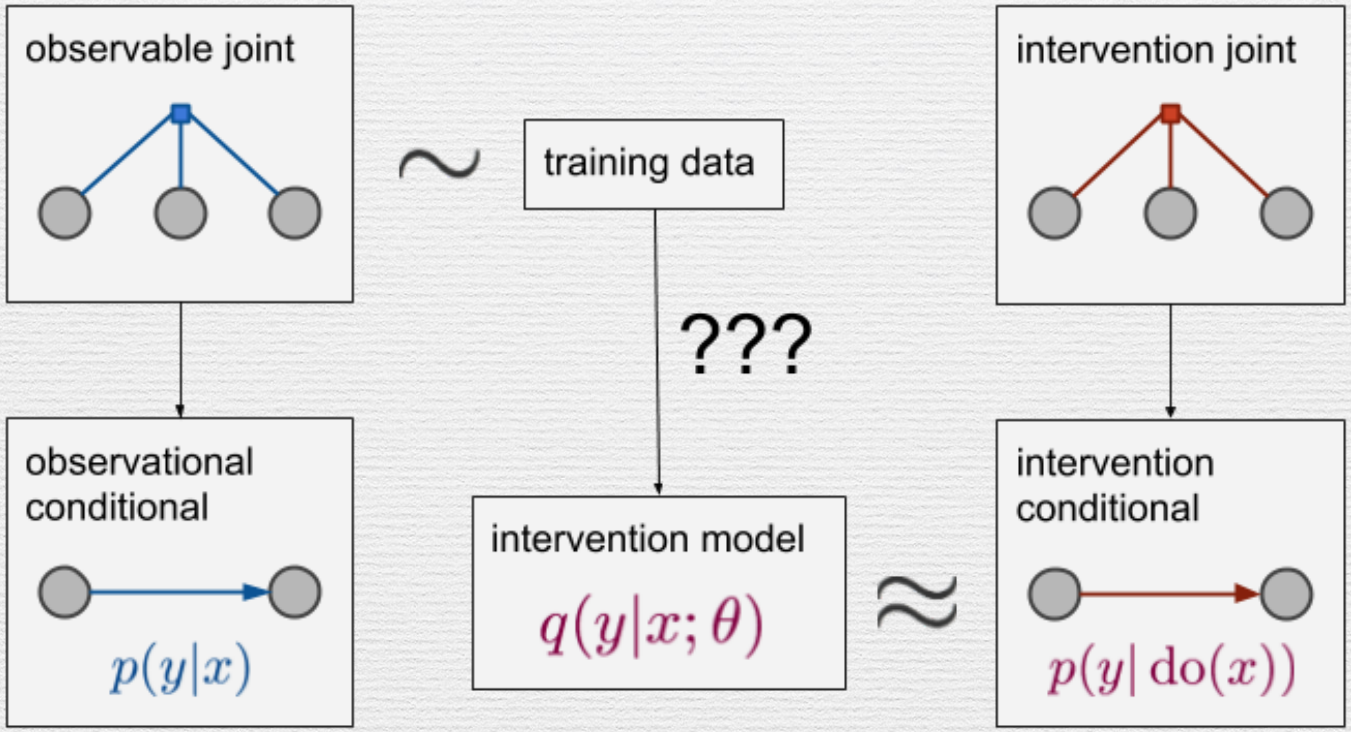


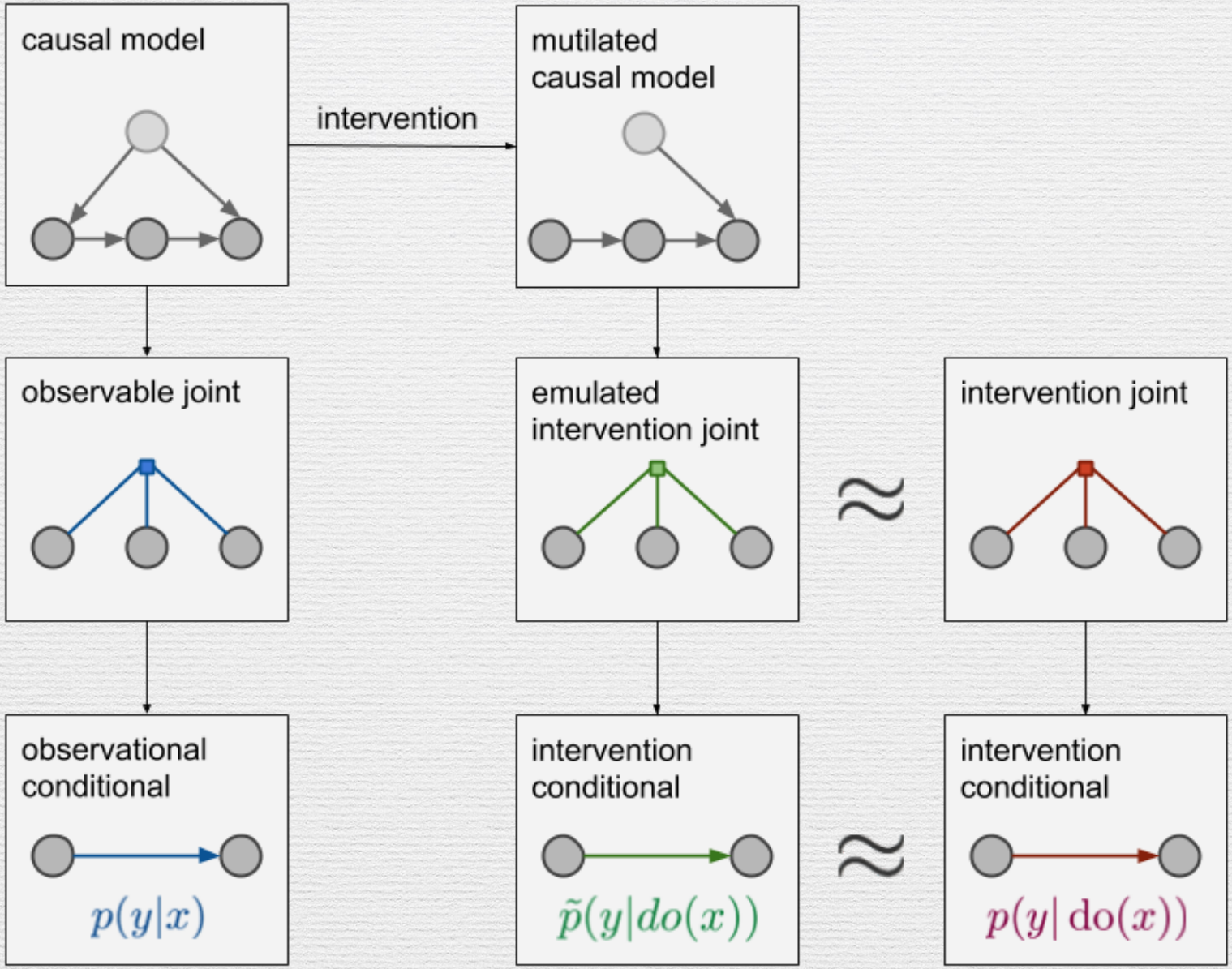
How to get causal graph

- Bayesian Reasoning: the most appealing answer is that you have to accept that your analysis is conditional on the graph you choose, and your conclusions are valid under the assumptions encoded there. In a way, causal inference from observational data is *subjective*. When you publish a result, you should caveat it with "under these assumptions, this is true". Readers can then dispute and question your assumptions if they disagree.
- Methods to be discussed

- y and x are correlated or statistically dependent and therefore seeing x allows me to predict the value of y, but y is not caused by x so setting the value of x won't effect the distribution of y.
- p(y|do(x)) is from a joint distribution of data which we would observe if we actually carried out the intervention in question.
  - is the conditional distribution we would learn from data collected in randomized controlled trials or A/B tests where the experimenter controls x.
  - Note that actually carrying out the intervention or randomized trials may be impossible or at least impractical or unethical in many situations.
- If I cannot measure p(y|do(x)) directly in a randomized controlled trial, can I estimate it based on data I observed outside of a controlled experiment?


observable joint
training data
supervised
learning
observational
conditional
trained ML model
$q(y|x;\theta)$
$p(y|x)$


observable joint
training data
intervention joint
???
observational conditional
$p(y|x)$
intervention model
$q(y|x;\theta)$
intervention conditional
$p(y|\,\mathrm{do}(x))$


causal model
intervention
mutilated
causal model
observable joint
emulated
intervention joint
≈
intervention joint
observational
conditional
$p(y|x)$
intervention
conditional
$\tilde{p}(y|do(x))$
≈
intervention
conditional
$p(y|\,\mathrm{do}(x))$