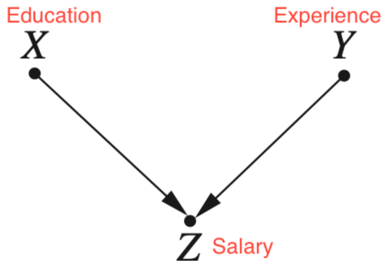
1. Introduction

2. Structural Causal Models

Structural causal models (SCMs) provide a comprehensive theory of causality.

2.1. Causal Graph

A causal graph $G = (V, E)$ is a directed graph that describes the causal effects between variables, where V is the node set and E the edge set. In a causal graph, each node represents a random variable including the treatment, the outcome, other observed and unobserved variables. A directed edge $x \rightarrow y$ denotes a causal effect of x on y.



2.2. 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

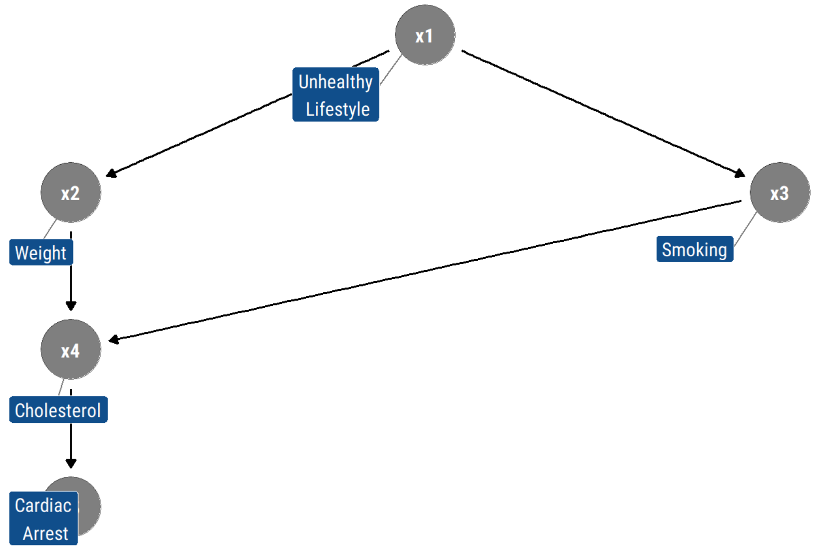
$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i|pa_i)$$

where pa_i stands for the values of the parents of variable X_i.

For example, in a simple chain graph $X \rightarrow Y \rightarrow Z$, we can write directly:

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

This knowledge allows us to save an enormous amount of space when laying out a joint distribution. To estimate the joint distribution from a data set generated by the above model, we need not count the frequency of every triple (x, y, z).; we can instead count the frequencies of each x, $(y|x)$, and $(z|y)$ and multiply.



$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)$$

2.3. $\boldsymbol{d}$-Separation

Given a SCM, the conditional independence embedded in its causal graph provides sufficient information to confirm whether it satisfies the criteria such that we can apply certain causal inference methods. To understand the **conditional independence**, we need the concept of **dependency-separation** (d-separation) based on the definition of **blocked** path.

Two events A and B are conditionally independent given a third event C if $P(A|B, C) = P(A|C)$ and $P(B|A, C) = P(B|C)$.

The probabilistic implications of the d-separation criterion:

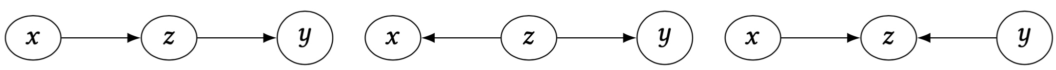
- If X and Y are d-separated by Z in causal graph G, then X is independent of Y conditional on Z in every distribution compatible with G
- Conversely, if X and Y are not d-separated by Z in causal graph G, then X and Y are dependent conditional on Z

There are three typical DAGs for conditional independence:

a) **Chain**: x causally affects y through its influence on z

b) **Fork**: z is the common cause of both x and y, that is, x is associated with y but there is no causation between them

c) **Collider**: both x and y cause z but there is no causal effect or association between x and y



(a) Chain (b) Fork (c) Collider

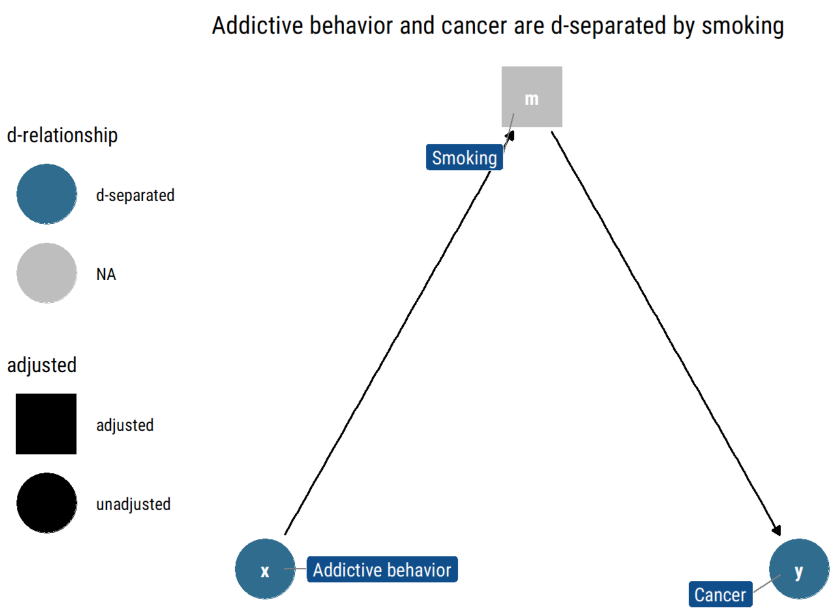
In the chain and fork, the path between x and y is blocked, that is, the variables become independent, if we condition on z. Contrarily, in a collider, conditioning on z introduces an association between x and y. Generally, we say conditioning on a set of nodes $\mathcal{Z}$ blocks a path p iff there exists at least one node $z \in \mathcal{Z}$ in the path p.

Definition 1 (d-separation) A path p is blocked (or d-separated) by a set of nodes $\mathcal{Z}$ if and only if

1. p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in $\mathcal{Z}$ (i.e., B is conditioned on), or
2. p contains a collider $A \leftarrow B \rightarrow C$ such that the collision node B is not in $\mathcal{Z}$, and no descendant of B is in $\mathcal{Z}$.

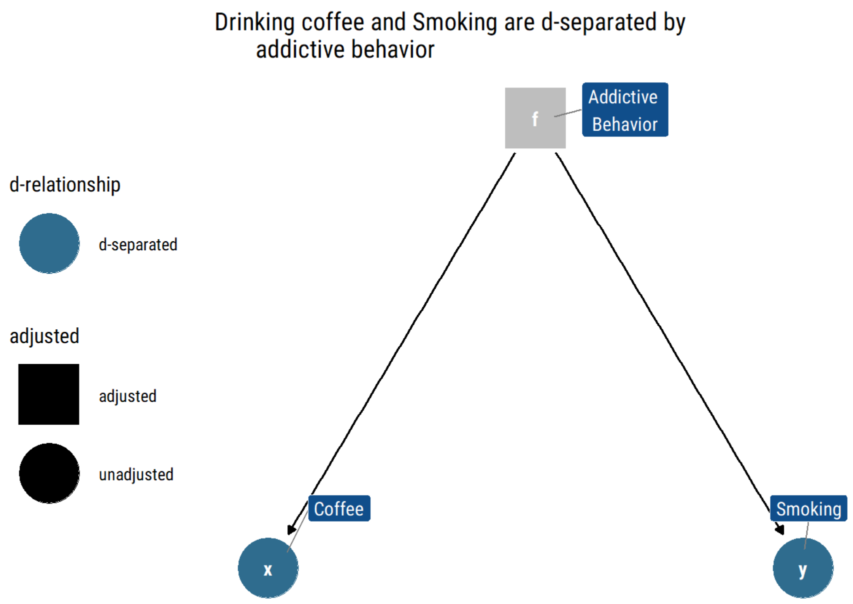
Chain

Rule 1 (Conditional Independence in Chains) Two variables, X and Y, are conditionally independent given Z, if there is only one unidirectional path between X and Y, and Z is any set of variables that intercepts that path.



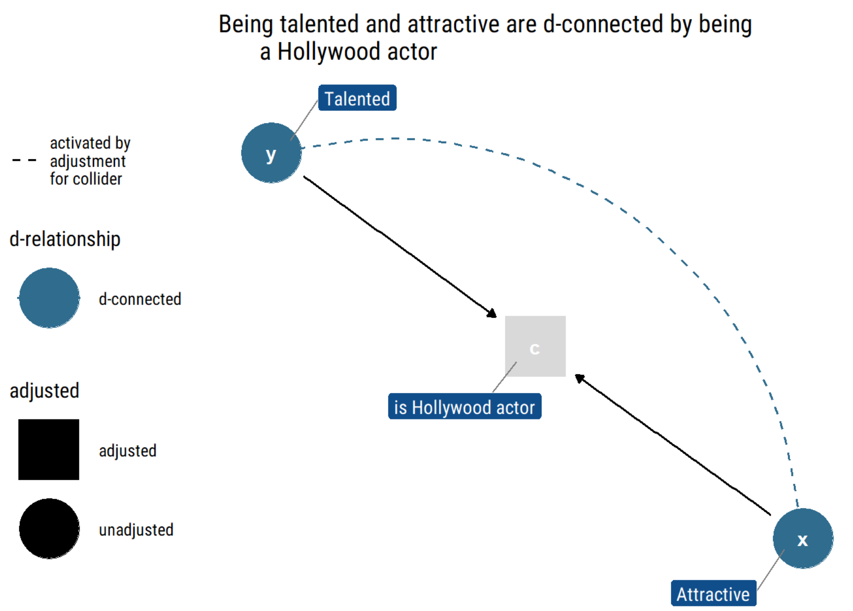
Fork

Rule 2 (Conditional Independence in Forks) If a variable X is a common cause of variables Y and Z, and there is only one path between Y and Z, then Y and Z are independent conditional on X.

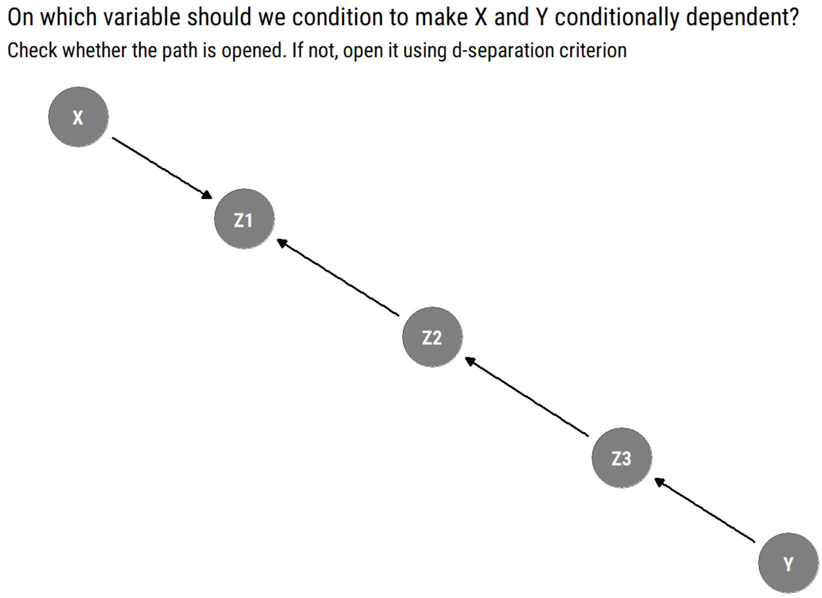


Collider

Rule 3 (Conditional Independence in Colliders) If a variable Z is the collision node between two variables X and Y, and there is only one path between X and Y, then X and Y are unconditionally independent but are dependent conditional on Z and any descendants of Z.



Armed with the tool of d-separation, we can now look at some more complex graphical models and determine which variables in them are independent and dependent, both marginally and conditional on other variables.

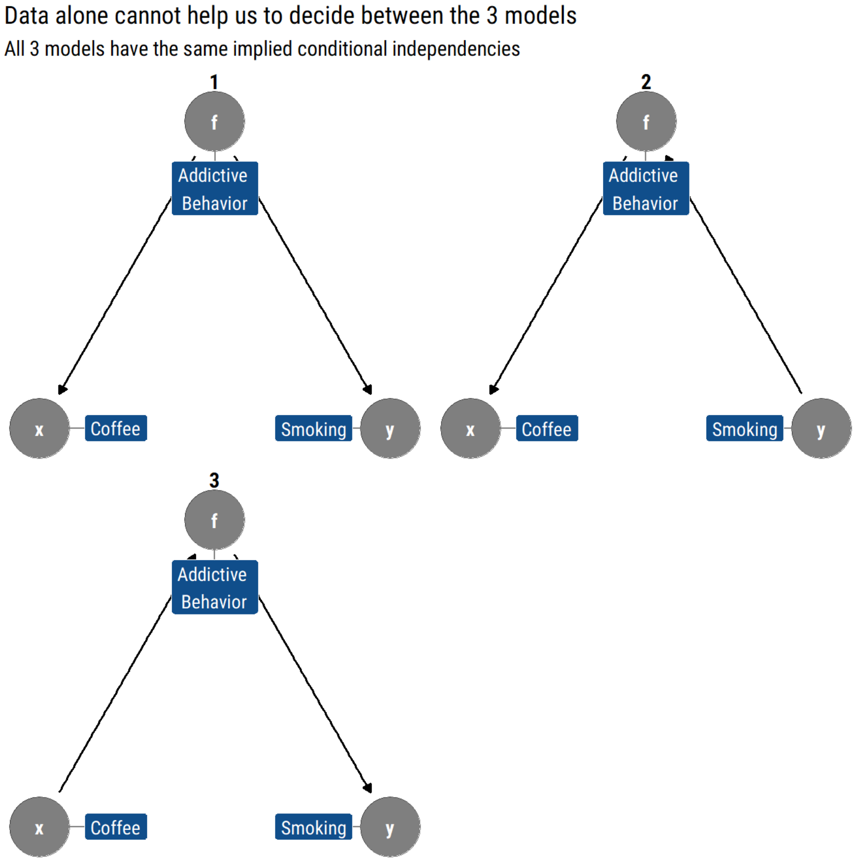


Can we distinguish models from data alone?

The preceding sections demonstrate that causal models have **testable implications** in the data sets they generate. For instance, if we have a graph G that we believe might have generated a data set S, d-separation will tell us which variables in G must be independent conditional on which other variables. Conditional independence is something we can test for using a data set. Suppose we list the d-separation conditions in G, and note that variables A and B must be independent conditional on C. Then, suppose we estimate the probabilities based on S, and discover that the data suggests that A and B are not independent conditional on C. We can then reject G as a possible causal model for S.

But there are other models that imply the same conditional independencies. Two graphical models G_1 and G_2 are observationally equivalent when they imply the same conditional independencies. The set of all the models with indistinguishable implications is called an **equivalence class**.

If two models are observationally equivalent, we **cannot use data alone to distinguish from them**. We must use our structural knowledge about the problem at hand to decide which model is the right one.



As Pearl says, **data are fundamentally dumb**: if we rely only in data to inform our models, we are extremely limited on what we can learn from them. We cannot predict the consequences of intervening (i.e., causal effects) in one of the variables with only observational data. Therefore, we must extend our theory beyond conditional probabilities gain causal understanding with only observational data.