Formalizing causal models

To make formal causal reasoning possible, we will use structural causal models. Such a model will describe:

- A number of variables (usually, but not always, these will be *random* variables)
- The causal relations between them: which variable is a cause of which other variable?

Structural causal models: an example

Example from the book

Variables:

$$X, Y, Z, U_X, U_Y, U_Z$$

Structural equations:

$$f_X : X = U_X$$
$$f_Y : Y = \frac{X}{3} + U_Y$$
$$f_Z : Z = \frac{Y}{16} + U_Z$$

(Possible interpretation: X is a high school's funding; Y is its average final exam score; Z is its college acceptance rate)

Structural causal models: an example

Example

Variables:

$$X, Y, Z, U_X, U_Y, U_Z$$

Structural equations:

$$f_X : X = U_X$$

$$f_Y : Y = \frac{X}{3} + U_Y$$

$$f_Z : Z = \frac{Y}{16} + U_Z$$

X, Y, Z: endogenous variables
(part of the system being modelled)

U_X, U_Y, U_Z: exogenous variables
(outside influences on the system)

(Possible interpretation: X is a high school's funding; Y is its average final exam score; Z is its college acceptance rate)

Structural causal models: an example

Example

Variables:

$$X, Y, Z, U_X, U_Y, U_Z$$

Structural equations:

$$f_X : X = U_X$$
$$f_Y : Y = \frac{X}{3} + U_Y$$
$$f_Z : Z = \frac{Y}{16} + U_Z$$

X, Y, Z: endogenous variables (part of the system being modelled)

U_X, U_Y, U_Z: exogenous variables (outside influences on the system)

f_X, f_Y, f_Z: describe the process by which 'nature' assigns values to X, Y, Z

(Possible interpretation: X is a high school's funding; Y is its average final exam score; Z is its college acceptance rate)

Structural causal models: definition (slightly more formal)

A *structural causal model (SCM)* consists of

- a set V of endogenous variables,
- a set U of exogenous variables, and,
- for each endogenous variable $X \in V$, a structural equation $f_X : X = \ldots$ that expresses X in terms of other (endo/exogenous) variables

Note that in f_X, the X is always to the left of the $=$. Rewriting it to a different form would change its meaning as a structural equation! (compare assignment in a programming language)

In the SCMs we will see today, each endogenous variable X will have a corresponding exogenous variable U_X; U_X will appear in f_X but not in any other f

Graphs

To visualize the qualitative aspects of an SCM, we can draw a graph as follows:

- for each variable (in $V \cup U$), draw a node
- for each $X \in V \cup U$ that appears in f_Y (for $Y \in V$), draw a directed edge from X to Y

Example SCM: corresponding graph

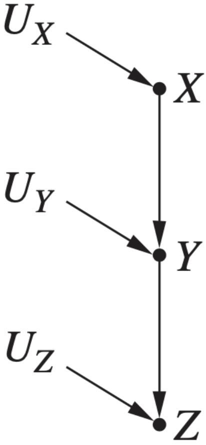
Example

Variables:

$$X, Y, Z, U_X, U_Y, U_Z$$

Structural equations:

$$f_X : X = U_X$$
$$f_Y : Y = \frac{X}{3} + U_Y$$
$$f_Z : Z = \frac{Y}{16} + U_Z$$



Graph terminology

- Two nodes are *adjacent* if there is an edge between them
- A *path* is a sequence of nodes, each adjacent to its predecessor
  - A path is *directed* if all edges along it point in the same direction
  - A path is *simple* if it has no repeated nodes
- If $X \to Y$, X is called a *parent* of Y, and Y a *child* of X
- If there is a directed path from X to Y (consisting of at least two nodes), X is called an *ancestor* of Y, and Y a *descendant* of X
- If any node is a descendant of itself, a graph is called *cyclic*; otherwise, it's called *acyclic*

We will only consider acyclic graphs in this course (though cyclic SCMs do exist!)

SCM-specific graph terminology

Consider an SCM, and draw its corresponding graph —

- If X is a parent of Y in the graph, X is called a direct cause of Y
- If X is an ancestor of Y in the graph, X is called a cause of Y

Advantages of the graph

Advantages of graph over the fully specified SCM:

- Quick visual representation
- Often, we will know where the causal relations are, but not be able to express them in formulas. Then the graph is exactly the part we do know!

Probabilities

Usually, we will treat the SCMs variables as random variables

- If we know the distribution of the exogenous variables, the joint distribution of $U \cup V$ is also known

We will always require the exogenous variables to be independent of each other

- This is to ensure that all dependences are visible in the graph
- To 'repair' an SCM that doesn't meet this requirement, we can add a new exogenous variable for each set S of dependent exogenous variables, making it a common parent of all the children of some node in S

Product decomposition rule

Product decomposition rule

The joint distribution of $U \cup V$ is given by

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_i P(X_i = x_i \,|\, \mathrm{Pa}(X_i) = \mathrm{pa}(x_i))$$

where $\mathrm{Pa}(X_i)$ denotes the set of parents of X_i, and $\mathrm{pa}(x_i)$ the corresponding values $(x_j, \dots, x_k)$

For $X_i \in U$: X_i has no parents, so the factor is simply $P(X_i = x_i)$

For $X_i \in V$: $X_i = f_{X_i}(\mathrm{pa}_i)$ with probability 1

Product decomposition rule: combining factors

If an exogenous variable has only one child, we can take their factors together

Let X_i be the child, and U_i the exogenous variable. Then

$$
\begin{aligned}
& P(x_i \mid \mathrm{pa}(x_i)) P(u_i \mid \mathrm{pa}(u_i)) \\
& = P(x_i \mid \mathrm{pa}(x_i)) P(u_i) \\
& = P(x_i \mid (\mathrm{pa}(x_i) \setminus u_i), u_i) P(u_i) \\
& = P(x_i \mid (\mathrm{pa}(x_i) \setminus u_i))
\end{aligned}
$$

Product decomposition rule: combining factors

If an exogenous variable has only one child, we can take their factors together

Let X_i be the child, and U_i the exogenous variable. Then

$$\begin{aligned} & P(x_i \mid \mathrm{pa}(x_i))P(u_i \mid \mathrm{pa}(u_i)) \\ & = P(x_i \mid \mathrm{pa}(x_i))P(u_i) \\ & = P(x_i \mid (\mathrm{pa}(x_i) \setminus u_i), u_i)P(u_i) \\ & = P(x_i \mid (\mathrm{pa}(x_i) \setminus u_i)) \end{aligned}$$

If all exogenous and endogenous variables are paired like this, we get (for $V = \{x_1, \ldots, x_n\}$)

$$P(x_1, \ldots, x_n) = \prod_i P(x_i \mid (\mathrm{pa}(x_i) \setminus u_i))$$

where all exogenous variables have disappeared

Importance of product decomposition

The product decomposition is a very useful tool, even if we are not interested in causality but only in probabilities:

- Fewer numbers / equations are required to specify the joint distribution (e.g. for binary variables, we need a table of 2^n probabilities, but often much fewer to specify all factors in the decomposition)
- As a consequence, it becomes much more feasible to learn these numbers from data

Importance of product decomposition

The product decomposition is a very useful tool, even if we are not interested in causality but only in probabilities:

- Fewer numbers / equations are required to specify the joint distribution (e.g. for binary variables, we need a table of 2^n probabilities, but often much fewer to specify all factors in the decomposition)
- As a consequence, it becomes much more feasible to learn these numbers from data

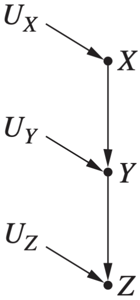
This is possible because the graph of the SCM puts certain *constraints* on the joint distributions that can be represented by that SCM

Graphs with one path

We will now look at different graphs, all having just one simple path between X and Y, and look at the properties of the corresponding joint distributions

Chains

Chain: a directed path of three nodes



Independences in chains where the decomposition rule holds and where the endogenous and exogenous variables are paired

- X and Y are likely dependent
- Y and Z are likely dependent
- X and Z are likely dependent

(and the same is true for pairs of nodes along longer directed paths)

Pathological cases (1/3)

Why 'likely' independent? Consider this SCM:

$$f_X : X = U_X$$
$$f_Y : Y = X - X + U_Y$$

Then X 'occurs' in f_Y, but doesn't actually have any effect on it

Pathological cases (2/3)

Less obvious example:

$$f_X : X = U_X$$
$$f_Y : Y = U_Y$$
$$f_Z : \begin{cases} X + 1 & \text{if } Y = 1083.4 \\ X & \text{otherwise} \end{cases}$$

If Y never equals 1083.4, we will never see the 'dependence' of Z on Y

Pathological cases (2/3)

Less obvious example:

$$f_X : X = U_X$$
$$f_Y : Y = U_Y$$
$$f_Z : \begin{cases} X + 1 & \text{if } Y = 1083.4 \\ X & \text{otherwise} \end{cases}$$

If Y never equals 1083.4, we will never see the 'dependence' of Z on Y

Both these examples are pathological: they are the exception rather than the rule

Pathological cases (3/3)

In a chain, even if the relation between X and Y as well as the relation between Y and Z are not pathological, the relation between X and Z may be pathological: this is called an intransitive case

$$V = \{X, Y, Z\}, U = \{U_X, U_Y, U_Z\}, F = \{f_X, f_Y, f_Z\}$$

$$f_X : X = U_X$$

$$f_Y : Y = \begin{cases} a & \text{IF } X = 1 \text{ AND } U_Y = 1 \\ b & \text{IF } X = 2 \text{ AND } U_Y = 1 \\ c & \text{IF } U_Y = 2 \end{cases}$$

otherwise

$$f_Y : Z = \begin{cases} i & \text{IF } Y = c \text{ OR } U_Z = 1 \\ j & \text{IF } U_Z = 2 \end{cases}$$

Conditional independence in chains

Definition

Two random variables X and Y are conditionally independent given a third (set of) random variable(s) Z if: for all values of Z, X and Y are independent

Conditional independence in chains

Definition

Two random variables X and Y are conditionally independent given a third (set of) random variable(s) Z if: for all values of Z, X and Y are independent

In a chain $X \rightarrow Y \rightarrow Z$,

- X and Z are independent given Y

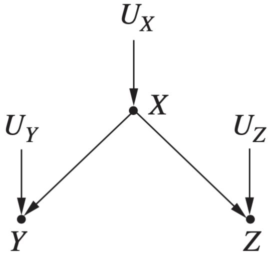
This is not 'likely', but always true!

Chains: rule for conditional independence

Rule 1: conditional independence in chains

If the only path between two variables X and Y is directed, and $\mathbf{Z}$ is a set of variables at least one of which is on that path, then X and Y are conditionally independent given $\mathbf{Z}$

Forks



Independences in forks

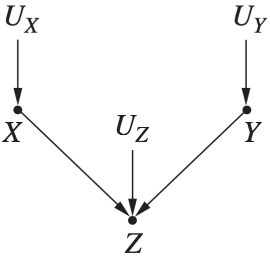
- X and Y are likely dependent
- X and Z are likely dependent
- Y and Z are likely dependent
- Y and Z are independent given X

Forks: rule for conditional independence

Rule 2: conditional independence in forks

If the only path between two variables Y and Z passes through a common cause X, then Y and Z are conditionally independent given X

Colliders



Independences in colliders

- X and Z are likely dependent
- Y and Z are likely dependent
- X and Y are independent
- X and Y are likely dependent given Z

Same is true if a descendant of the collider Z is conditioned on

Colliders: rule for conditional independence

Rule 3: conditional independence in colliders

If the only path between two variables X and Y passes through a collider Z, then X and Y are marginally (i.e. unconditionally) independent, but likely dependent given Z, or given any descendant of Z

Colliders: counterintuitive

The conditional dependence here is harder to reason about intuitively / surprising to many people.

Possible reason: no easy explanation in terms of causation (one variable causing the other, or a third variable causing both)