NCTIONAL CAUSAL MODELS

① Consider the physics formula $a = \frac{F}{m}$, i.e. the acceleration of an object is directly proportional to the applied force F and inversely proportional to its mass m.

To answer a question like "how does the acceleration change as we vary the applied force?", we can compute the partial derivative $\frac{\partial a}{\partial F}$, which is the usual derivative of a with respect to F, given that we treat the other quantities (i.e. mass m) as constants.

$$\frac{\partial a}{\partial F} = \frac{1}{m}$$

② This is fine, but what about a similar formula describing the internal air pressure of a balloon as a function of altitude and temperature:

$$p = \frac{t}{a^2}$$

To answer a question like "how does the pressure change as we change the altitude?", we could compute $\frac{\partial p}{\partial a}$ ...

FUNCTIONAL CAUSAL MODELS

(3) ... but what if the temperature also changes when we change the altitude? e.g. say that $t = -a$.

Then there are two partial derivatives that might make sense:

$\left(\frac{\partial p}{\partial a}\right)_t$ denotes the derivative of p wrt a if we keep t fixed

$\frac{\partial p}{\partial a}$ denotes the derivative of p wrt a if we allow t to change as a function of a

---

(4) This can all get rather confusing. It helps to first specify a causal network over the variables, e.g.

$$\begin{tikzcd} F \arrow[dr] & & m \arrow[dl] \\ & a & \end{tikzcd} \qquad \text{or} \qquad \begin{tikzcd} a \arrow[r] \arrow[dr] & t \arrow[d] \\ & p \end{tikzcd}$$

Here, the edges have a causal interpretation, specifically a node is a function of its parents. Also, we assume that if I change a variable, then the values of its descendants are updated, but the values of its nondescendants remain the same.

For instance, if I change the altitude (in the network above), then the temperature changes. But if I change the temperature, the altitude stays the same.

FUNCTIONAL CAUSAL MODELS

5) This interpretation makes it easier to talk about partial derivatives. In the network:

$$\begin{tikzcd} a \arrow[r] \arrow[dr] & t \arrow[d] \\ & p \end{tikzcd}$$

changing a causes p to change in two different ways, each corresponding to a path from a to p:

$$\begin{tikzcd} a \arrow[r] \arrow[dr, thick] & t \arrow[d] \\ & p \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} a \arrow[r, thick] \arrow[dr] & t \arrow[d, thick] \\ & p \end{tikzcd}$$

Since we want to know about the total effect of changing a on the value of p, we define $\frac{\partial p}{\partial a}$ as the derivative of p wrt a, keeping all nondescendants (here, nothing) fixed.

(6) To hammer home the value of the network, consider the following equations:

$$x = u$$
$$y = u + v$$
$$z = x + y$$

Now let's compute $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial u}$.

$$\frac{\partial z}{\partial x} = \frac{\partial (x+y)}{\partial x} = 1 + 0 = 1$$

$$\frac{\partial z}{\partial u} = \frac{\partial (x+y)}{\partial u} = \frac{\partial x}{\partial u} + \frac{\partial y}{\partial u} = \frac{\partial u}{\partial u} + \frac{\partial (u+v)}{\partial u} = 1 + 1 + 0 = 2$$

It's possibly confusing that $\frac{\partial z}{\partial x} \neq \frac{\partial z}{\partial u}$, given $x = u$.

It only makes sense under a causal (assymmetric) interpretation of $x = u$ as $x := u$:

$$\begin{tikzcd}
u \arrow[r] \arrow[dr] & x \arrow[dr] & \\
 & & z \\
v \arrow[r] & y \arrow[ur] &
\end{tikzcd}$$

Under this interpretation, u affects z through two different causal paths, while x affects z through only one, explaining the difference.

Functional Causal Models

⑦ Sometimes it is useful to categorize the variables of the causal graph according to two criteria:

- is the variable <u>observable</u>?
- is the variable a <u>background variable</u>?

To show these criteria in action, consider the following simple game show. The host is holding an envelope which contains a slip of paper with a number written on it. You roll two dice. If the total of the dice is greater than the number in the envelope (which you never get to see), then you win a prize.

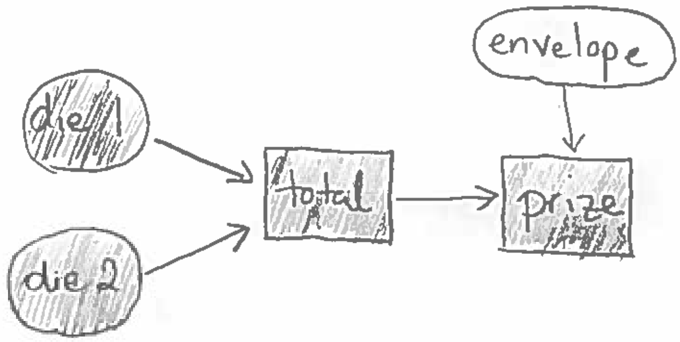
We can model this scenario with the following graph:

$$\begin{tikzcd}
\text{die 1} \arrow[dr] & & \text{envelope} \arrow[d] \\
& \text{total} \arrow[r] & \text{prize} \\
\text{die 2} \arrow[ur] & &
\end{tikzcd}$$

where die 1, die 2 are the numbers you rolled
total is their sum
envelope is the number inside the envelope
prize is whether you win a prize

FUNCTIONAL CAUSAL MODELS

(8) In this course, I'll sometimes use the following visual notation to show if variables are observable or background:



- shaded variables are <u>observable</u>: we can directly observe our dice roll, our total, our prize, but we never directly observe the envelope contents
- circled variables are <u>background</u>: we cannot (or do not wish to) express these variables as a function of other variables.

from the Greek endo- (within) and -genous (producing)

Non-background variables are called <u>endogenous</u>. We assume these can be expressed as a function of their parents:

$$\text{total} = \text{die1} + \text{die2}$$

$$\text{prize} = \begin{cases} 1 & \text{if total} > \text{envelope} \\ 0 & \text{o.w.} \end{cases}$$

FUNCTIONAL CAUSAL MODELS

(9) Formally, a <u>functional causal model</u> is defined as a triple $M = \langle U, V, F \rangle$ where:

definition from J. Pearl, <u>Causality</u>, Ch. 7

- U is a set of <u>background</u> variables that are determined by factors outside the model
- V is a set $\{V_1, V_2, \ldots, V_n\}$ of variables, called <u>endogenous</u>, that are determined by variables in the model — that is, variables in $U \cup V$.
- F is a set $\{f_1, f_2, \ldots, f_n\}$ of functions $f_i : U \cup (V \setminus V_i) \to V_i$ s.t. the entire set F forms a mapping from U to V.

---

(10) We'll be exclusively concerned with <u>acyclic functional causal models</u>, which have the following additional property:

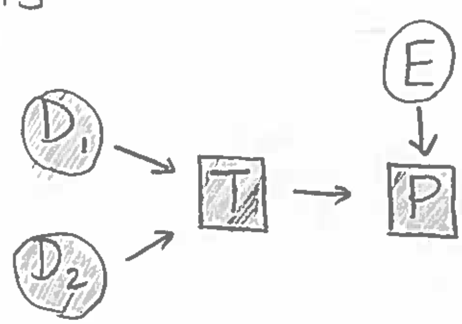
- each function $f_i \in F$ can be expressed as a mapping $f_i' : PA_i \to V_i$ for $PA_i \subseteq U \cup (V \setminus V_i)$ s.t. $f_i'(pa_i) = f_i(pa_i, o_i)$ for all instantiations pa_i of PA_i and o_i of $(U \cup (V \setminus V_i)) \setminus PA_i$.
- the graph produced by drawing edges from P to V_i iff $P \in PA_i$ is <u>acyclic</u>.

Functional Causal Model

(11) So for our game show example, $M = \langle U, V, F \rangle$ where:

- $U = \{D_1, D_2, E\}$
- $V = \{T, P\}$
- $F = \left\{ \begin{array}{l} f_T : f_T(d_1, d_2) = d_1 + d_2, \\ f_P : f_P(t, e) = \begin{cases} 1 & \text{if } t > e \\ 0 & \text{o.w} \end{cases} \end{array} \right\}$

And the <u>causal diagram</u> $G(M)$ is defined as the directed graph in which each node corresponds to a variable and the directed edges point from members of PA_i toward V_i. So the causal diagram for the game show example is:



(12) Basically, making a variable endogenous is a <u>choice</u>. Do we want to model how this variable is generated, or do we want to assume it as given?

FUNCTIONAL CAUSAL MODEL

(13) A probabilistic causal model adds a probability distribution over the background variables, i.e. it is a pair $\langle M, P(u) \rangle$ where M is a causal model and $P(u)$ is a probability distribution over U.

e.g. $P(d_1) = \frac{1}{6}$ for $d_1 \in \{1, \dots, 6\}$

$P(d_2) = \frac{1}{6}$ for $d_2 \in \{1, \dots, 6\}$

$P(e) = \frac{1}{11}$ for $e \in \{2, \dots, 12\}$

$P(d_1, d_2, e) = P(d_1) P(d_2) P(e)$

FUNCTIONAL CAUSAL MODELS

(14) Let's exercise our causal modeling skills. Suppose we collected data about people with a particular (often fatal) disease. Some of them took a particular drug ($D=1$) and some of them recovered ($R=1$):

	$R=0$	$R=1$
$D=0$	24	16
$D=1$	20	20

From the data, 50% (20 out of 40) of the people who took the drug recovered. Only 40% (16 out of 40) of the people who did not take the drug recovered.

(i) Should we recommend the drug?

(ii) Is it possible for the following to be also true:

- if you were male, the likelihood of recovery went down if you took the drug
- if you were female, the likelihood of recovery went down if you took the drug
- if you were gender nonbinary, the likelihood of recovery went down if you took the drug

15. Surprisingly, (ii) is possible.

Males	R=0	R=1
D=0	3	7
D=1	12	18

70% of males who didn't take drug RECOVERED
60% of males who took drug RECOVERED

Females	R=0	R=1
D=0	18	8
D=1	7	2

31% of females who didn't take drug RECOVERED
22% of females who took drug RECOVERED

Nonbinary	R=0	R=1
D=0	3	1
D=1	1	0

25% of nonbinary who didn't take drug RECOVERED
0% of nonbinary who took drug RECOVERED

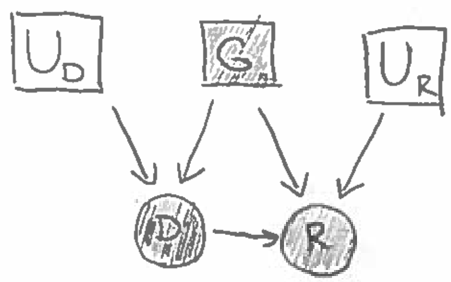
Total	R=0	R=1
D=0	24	16
D=1	20	20

40% of people who didn't take drug RECOVERED
50% of people who took drug RECOVERED

This is a well-known phenomenon called Simpson's Paradox.

Functional Causal Models

(16) One explanation: males are both more likely to recover and more inclined to take this particular drug (maybe it's Rogaine or Propecia). Here is that explanation as a causal model:



Suppose U_D and U_R are both real numbers between 0 and 1 generated uniformly at random. Gender G is also treated as a background variable. The two endogenous variables D and R can be expressed as functions of their parents in the causal diagram:

an indicator function $1_{pred}(x) = \begin{cases} 1 & \text{if } pred(x) \\ 0 & \text{o.w.} \end{cases}$

$$D = \begin{cases} 1_{U_D \leq 0.6}(U_D) & \text{if } G = \text{"male"} \\ 1_{U_D \leq 0.25}(U_D) & \text{o.w.} \end{cases}$$

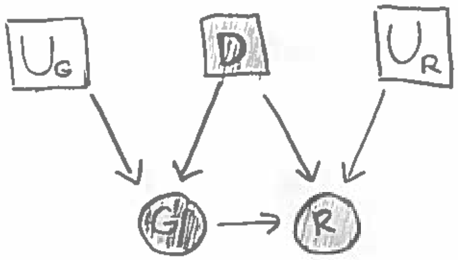
$$R = 1_{U_R \leq [0.3 + 0.4 \cdot 1_{G=male}(g) - 0.1 \cdot 1_{D=1}(d)]}(U_R)$$

FUNCTIONAL CAUSAL MODELS

17) According to this explanation, we should not recommend the drug, since it decreases the probability of recovery for everyone:

$$R = \mathbb{1}_{U_R \leq [0.3 + 0.4 \cdot \mathbb{1}_{G=male}(g) \underbrace{- 0.1 \cdot \mathbb{1}_{D=1}(d)}_{\text{10\% absolute decrease in recovery probability if } D=1}]}(U_R)$$

(18) But there are other explanations. Here is another explanation (as a causal diagram):



Here, the drug's direct effect still slightly hurts recovery, but it also has a possible side effect of causing a gender transition to male, which greatly increases recovery probability. For instance, here's one set of structural equations that fit the data:

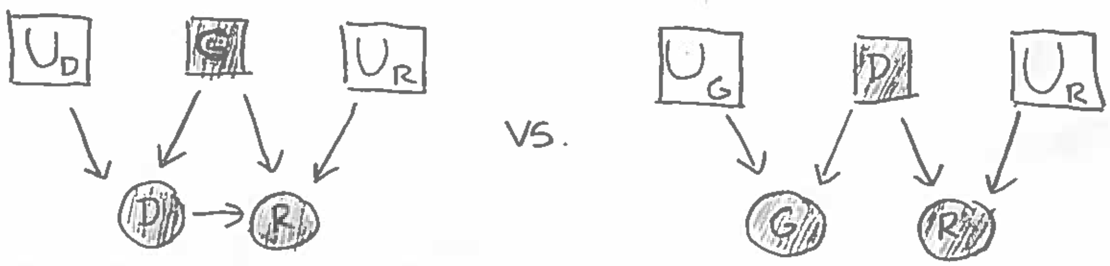
$$G = \begin{cases} \text{"nonbinary" if } U_G \leq 0.07 \\ \text{"male" if } U_G \leq 0.53 + 0.25 \cdot 1_{D=1}(d) \\ \text{"female" o.w.} \end{cases}$$

$$R = 1_{U_R \leq [0.3 + 0.4 \cdot 1_{G=\text{male}}(g) - 0.1 \cdot 1_{D=1}(d)]}(U_R)$$

So should we recommend the drug? Well... it does increase the probability of recovery, though it's problematic to potentially cause gender transition as an accidental side effect, so probably not. But if we substitute gender for, say, cholesterol level, then yes.

FUNCTIONAL CAUSAL MODELS

(19) The point is that the data <u>by itself</u> doesn't tell us whether administering the drug will (overall) help or hurt recovery. We need to decide which causal diagram is more plausible in order to interpret the data:



They allow for a formal way of representing assumptions / intuition about the causal mechanisms producing the data. You can look at one, and easily say that "yeah, that seems right" or "no, I doubt the drug caused gender transitions — they probably would have mentioned that when they gave me the data."