LECTURE 12: CAUSAL MODELING

So Far: probabilistic modeling

- model joint distribution $p(x_1, \ldots, x_d)$ from observations
- encodes probabilistic questions about observed data
  - → conditional independencies (graphical models)
  - → inference ($p(x_1)$, $p(x_5 | x_1)$ ...)
- Ex: - is the rate of car accidents higher for 16-y.o. than 18-y.o. ?

$$p(\text{accident} \mid \text{age} = 16) \overset{?}{\gtrless} p(\text{accident} \mid \text{age} = 18)$$

- is the rate of admissions higher for men vs women ?

$$p(\text{admitted} \mid \text{gender} = M) \overset{?}{\gtrless} p(\text{admitted} \mid \text{gender} = F)$$

- is the rate of recovery higher for medical treatment A or B ?

$$p(R \mid T = A) \overset{?}{\gtrless} p(R \mid T = B)$$

1. From probabilistic to causal models

→ We often care about "alternative worlds" that are not well modeled using a fixed joint distribution

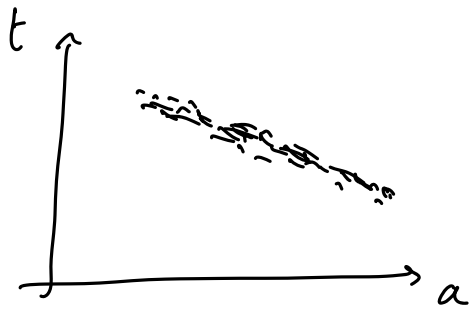
→ causal questions / interventions

e.g. "what is the effect of a new treatment / policy ?"

Three examples

- cause-effect models

data: altitude vs temperature (Peters et al.) for different cities



Q: (A) → (T) or (T) → (A) ?

→ probabilistic modeling: both work!

$$p(a,t) = p(a)\,p(t|a)$$
$$= p(t)\,p(a|t)$$

→ causal modeling: interventions

"if we change A, does T change?" ✓

" ——— T — A — ?" ✗

(e.g. using a heater)

→ $p(a)$ and $p(t|a)$ can be viewed as independent, "physical" mechanisms

→ If we change $p(a)$ to $\tilde{p}(a)$, we expect a modified observed distribution

$$\tilde{p}(a,t) = \tilde{p}(a)\, p(t|a)$$

($p(t|a)$: unchanged)

→ in contrast, changing $p(t)$ to $\tilde{p}(t)$ would require a different $\tilde{p}(a|t)$ in order for $\tilde{p}(a,t) = \tilde{p}(t)\,\tilde{p}(a|t)$ to make sense physically

⇒ Causal models satisfy this independence of mechanisms

- Confounder on treatment

→ Kidney stones example (Bottou et al. 2013)

Success Rates Y	Overall	Patients with small stones	Patients with large stones
Treatment *a*: Open surgery	78% (273/350)	**93%** (81/87)	**73%** (192/263)
Treatment *b*: Percutaneous nephrolithotomy	**83%** (289/350)	87% (234/270)	69% (55/80)

T

Overall, b seems better

Yet, a is better on individual groups

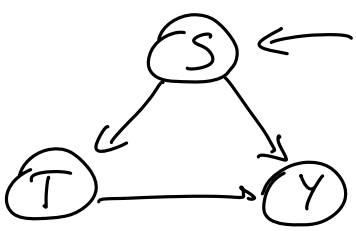
(Simpson's paradox / reversal)

$$P(Y \mid T=a) < P(Y \mid T=b)$$

But: $P(Y \mid T=a, S=s) > P(Y \mid T=b, S=s) \quad \forall s.$

Why? e.g. T=a was given to more severe cases, patients with large stones.

Possible causal model:



confounder

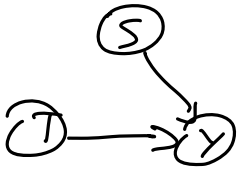
→ $P(Y|T)$ is "confounded" by S

→ "Adjustment": use $P(Y|T,S)$ instead

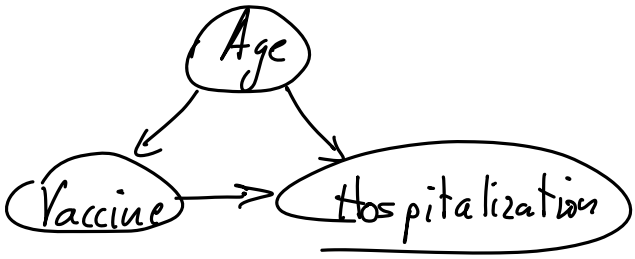
→ Causal effect / treatment effect:

$$P(Y \mid do(T=a)) - P(Y \mid do(T=b))$$

"intervention"



→ other example: Covid-19



Mediation

Berkeley admissions example (Bickel et al. 1975) (Hardt & Recht)

UC Berkeley admissions data from 1973.

	Men		Women	
Department	Applied	Admitted (%)	Applied	Admitted (%)
A	825	62	108	82
B	520	60	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7

Overall: higher admission rate for men

But: _____ for women in multiple departments

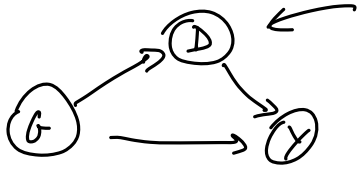
(another Simpson's reversal)

Why? e.g. women apply to more crowded, competitive departments

Bickel: ≠ discrimination does not come from admission process itself

→ Possible causal model:

D: dept.
G: "gender"
Y: admission



"mediator"

Can study direct/indirect causal paths $G \overset{(D)}{\rightsquigarrow} Y$
("mediation analysis")

Note: The question of discrimination is more nuanced (see fairmlbook.org, chap. 4)

---

2. Structural Causal Models

(also Structural Equation Models in econometrics)

Def. A **Structural Causal Model** (SCM) M on variables $X_1, \dots, X_d$ is given by a set of **assignments** of the form:

$$X_i := f_i(X_{p(i)}, U_i) \qquad i = 1, \dots, d$$

- $p(i) \subseteq \{1, \dots, d\}$ are parents of i
- $U_1, \dots, U_d$ are **noise variables**, jointly independent
- We associate the corresponding graph G, assumed acyclic.

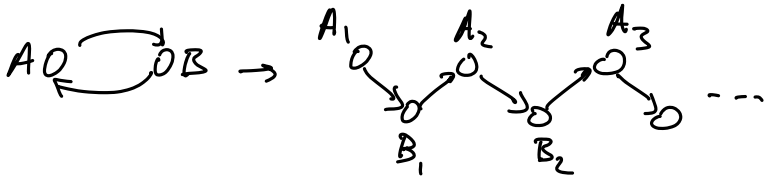
Notes:

- An SCM $\mathcal{M}$ defines an **observational distribution**

$$P^{\mathcal{M}}(x) = \prod_i P^{\mathcal{M}}(x_i \mid x_{p(i)})$$

where the conditionals $P^{\mathcal{M}}(x_i \mid x_{p(i)})$ correspond to the assignments $X_i := f_i(X_{p(i)}, U_i)$

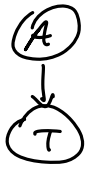
- "assignments" vs "equations": these are not mathematical equations, but rather "code instructions" for a sampling procedure ("ancestral sampling")
- cyclic models can often be unrolled to DAG over time



Ex:

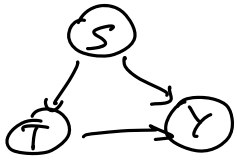
- altitude / temperature

$$A := U_A$$
$$T := f(A, U_T)$$



- Kidney stones

$$S := U_S$$
$$T := f_T(S, U_T)$$
$$Y := f_Y(T, S, U_Y)$$



Interventions and causal effects

We can model interventions by substituting assignments.

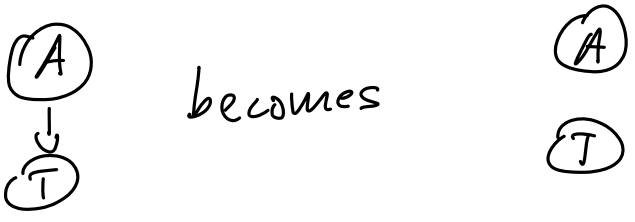
e.g. $X_i := f_i(X_{p(i)}, U_i)$ is replaced by $X_i := \bar{x}$

The resulting model is denoted $M; do(X_i := x)$

$p^{M; do(X_i := x)}$ is the interventional distribution

Ex: • changing temperature in A-T model:

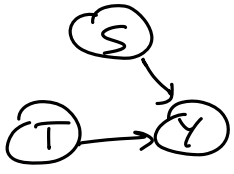
$$A := U_A$$
$$T := t$$



(vs $A := a$, $T := f(A, U_T)$ for intervention on A)

• Intervention on medical treatment

$$S := U_S$$
$$T := t$$
$$Y := f_Y(S, T, U_Y)$$



→ We often denote

$$P(Y \mid do(T:=t)) = P^{M;do(T:=t)}(Y)$$

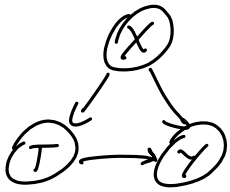
$$\left(\neq P(Y \mid T=t) \text{ in general !!} \right)$$

Def: For binary treatments $T \in \{0,1\}$, we define

- Average Treatment Effect (ATE)

$$\mathbb{E}[Y \mid do(T:=1)] - \mathbb{E}[Y \mid do(T:=0)]$$

- Conditional/Heterogeneous ATE (CATE/HTE) when conditioning on observed confounders

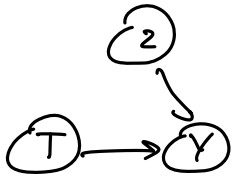


$$\mathbb{E}[Y \mid X=x, do(T:=1)] - \mathbb{E}[Y \mid X=x; do(T:=0)]$$

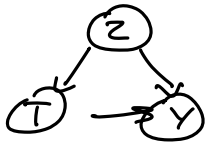
- Adjustment for confounders

Q: How can we compute $P(Y \mid do(T:=t))$ using observational data?

→ Try to rewrite in terms of conditional distr. in the observational model:



$P(X)$ and $P(Y \mid Z, T)$ are the same as in observed model:



$$P(Y \mid do(T=t)) = \sum_z P(Y \mid T=t, Z=z) P(Z=z)$$

("adjustment formula")

More generally: backdoor criterion / do-calculus

Alternative: Randomized Control Trial

(experiment s.t. T is randomly assigned)