Causality

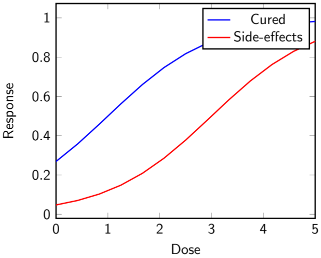
Actions, Confounders and Interventions

Christos Dimitrakakis

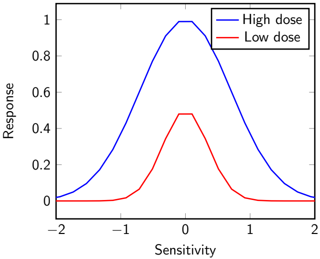
October 25, 2018

Headaches and aspirins

Example 1 (Population effects)



(a) Dose-response curve.



(b) Response distribution

Figure: Investigation the response of the population to various doses of the drug.

- Is aspirin an effective cure for headaches?
- Does having a headache lead to aspirin-taking?

Example 2 (Individual effects)

- Effects of Causes: Will my headache pass if I take an aspirin?
- Causes of Effects: Would my headache have passed if I had not taken an aspirin?

Overview

Inferring causal models

We can distinguish different models from observational or experimental data.

Inferring individual effects

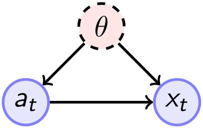
The effect of possible intervention on an individual is not generally determinable. We usually require strong assumptions.

Decision-theoretic view

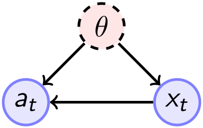
There are many competing approaches to causality. We will remain within the decision-theoretic framework, which allows us to crisply define both our knowledge and assumptions.

What causes what?

Example 3



(a) Independence of a_t.



(b) Independence of x_t.

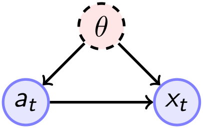
Suppose we have data x_t, a_t where

- x_t: lung cancer
- a_t: smoking

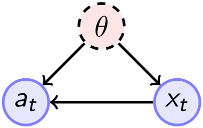
Does smoking cause lung cancer or does lung cancer make people smoke? Can we compare the two models above to determine it?

What causes what?

Example 3



(a) Independence of a_t.



(b) Independence of x_t.

Suppose we have data x_t, a_t where

- x_t: lung cancer
- a_t: smoking

Does smoking cause lung cancer or does lung cancer make people smoke? Can we compare the two models above to determine it?

$$P_\theta(D) = \prod_t P_\theta(x_t, a_t) = \prod_t P_{\theta'}(x_t \mid a_t) P_{\theta'}(a_t) = \prod_t P_{\theta''}(a_t \mid x_t) P_{\theta''}(x_t).$$

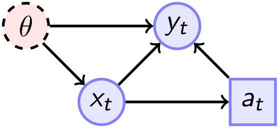


Figure: A typical decision diagram where x_t: individual information, y_t: individual result, a_t: action, π: policy

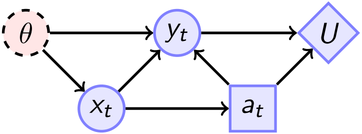


Figure: A typical decision diagram where x_t: individual information, y_t: individual result, a_t: action, π: policy

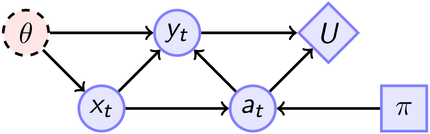


Figure: A typical decision diagram where x_t: individual information, y_t: individual result, a_t: action, π: policy

Example 4 (Taking an aspirin)

- Individual t
- Individual information x_t
- $a_t = 1$ if t takes an aspirin, and 0 otherwise.
- $y_t = 1$ if the headache is cured in 30 minutes, 0 otherwise.
- π: intervention policy.

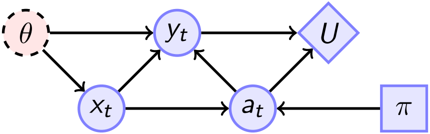


Figure: A typical decision diagram where x_t: individual information, y_t: individual result, a_t: action, π: policy

Example 4 (A recommendation system)

- x_t: User information (random variable)
- a_t: System action (random variable)
- y_t: Click (random varaible)
- π: recommendation policy (decision variable).

Conditional distributions and decision variables.

$$P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)}.$$

The conditional distribution of decisions

$$\pi(a) \equiv \mathbb{P}^{\pi}(a) \equiv \mathbb{P}(a \mid \pi).$$

$$\mathbb{P}^{\pi}_{\theta}(a) \equiv \mathbb{P}(a \mid \theta, \pi).$$

Basic causal structures

Non-cause

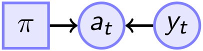


Figure: π does not cause y

No confounding

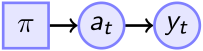


Figure: No confounding: π causes y_t

Basic causal structures

Non-cause

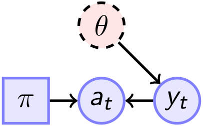


Figure: π does not cause y

No confounding

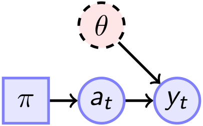


Figure: No confounding: π causes y_t

Covariates

Sufficient covariate

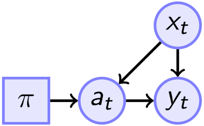


Figure: Sufficient covariate x_t

Instrumental variables and confounders

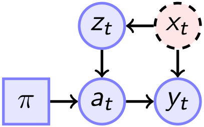


Figure: Instrumental variable z_t

Covariates

Sufficient covariate

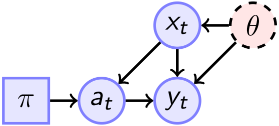


Figure: Sufficient covariate x_t

Instrumental variables and confounders

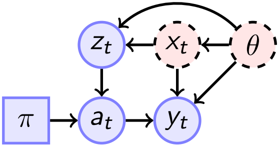


Figure: Instrumental variable z_t

Modelling interventions

- Observational data D.
- Policy space Π.

Default policy

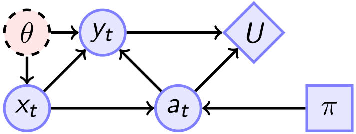
The space of policies Π includes a default policy π_0, under which the data was collected.

Intervention policies

Except π_0, policies $\pi \in \Pi$ represent different interventions specifying a distribution $\pi(a_t \mid x_t)$.

- Direct interventions.
- Indirect interventions and non-compliance.

Example 5 (Weight loss)



Example 5 (Weight loss)

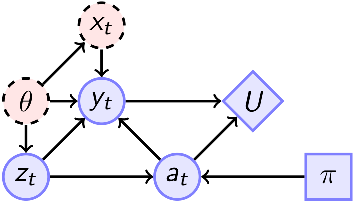


Figure: Model of non-compliance as a confounder.

The value of an observed policy

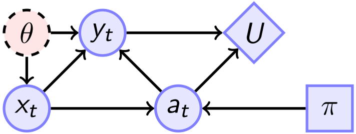


Figure: Basic decision diagram

$$\hat{a}^*_D \in \arg\max_a \hat{\mathbb{E}}_D(U \mid a),$$

The value of an observed policy

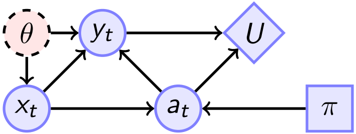


Figure: Basic decision diagram

$$\hat{\mathbb{E}}_D(U \mid a) \triangleq \frac{1}{|\{t \mid a_t = a\}|} \sum_{t : a_t = a} U(a_t, y_t) \tag{3.1}$$

$$\approx \mathbb{E}^{\pi_0}_\theta(U \mid a) \qquad (a_t, y_t) \sim \mathbb{P}^{\pi_0}_\theta . \tag{3.2}$$

$$\hat{a}^*_D \in \arg\max_a \hat{\mathbb{E}}_D(U \mid a),$$

$$x_t \mid \theta \sim P_\theta(x)$$
$$y_t \mid \theta, x_t, a_t \sim P_\theta(y \mid x_t, a_t)$$
$$a_t \mid x_t, \pi \sim \pi(a \mid x_t).$$

The value of a policy

$$\mathbb{E}^\pi_\theta(U) = \int_\mathcal{X} \mathrm{d}P_\theta(x) \sum_{y \in \mathcal{Y}} P_\theta(y \mid x, a) U(a, y) \sum_{a \in \mathcal{A}} \pi(a \mid x).$$

The optimal policy under a known parameter θ is given simply by

$$\max_{\pi \in \Pi} \mathbb{E}^\pi_\theta(U),$$

where Π is the set of allowed policies.

Monte-Carlo estimation

Importance sampling[1]

We can obtain an unbiased estimate of the utility in a model-free manner through importance sampling:

$$\mathbb{E}^{\pi}_{\theta}(U) = \int_{\mathcal{X}} \mathrm{d}P_{\theta}(x) \sum_{a} \mathbb{E}_{\theta}(U \mid a, x)\pi(a \mid x)$$
$$\approx \frac{1}{T} \sum_{t=1}^{T} U_t \frac{\pi(a_t \mid x_t)}{\pi_0(a_t \mid x_t)}.$$

[1] Also known as Propensity Scoring

Bayesian estimation

If we π_0 is given, we can calculate the utility of any policy to whatever degree of accuracy we wish.

$$\xi(\theta \mid D, \pi_0) \propto \prod_t \mathbb{P}_\theta^{\pi_0}(x_t, y_t, a_t)$$

$$\mathbb{E}_\xi^\pi(U \mid D) = \int_\Theta \mathbb{E}_\theta^\pi(U) \, \mathrm{d}\xi(\theta \mid D)$$

$$= \int_\Theta \int_\mathcal{X} \mathrm{d}P_\theta(x) \sum_{t=1}^T \sum_a \mathbb{E}_\theta(U \mid a, x) \pi(a \mid x) \, \mathrm{d}\xi(\theta \mid D).$$

Causal inference and policy optimisation

Example 6

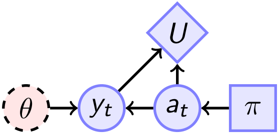


Figure: Simple decision problem.

Let $a_t, y_t \in \{0, 1\}$, $\theta \in [0, 1]^2$ and

$$y_t \mid a_t = a \sim \mathcal{B}ernoulli(\theta_a)$$

Then, by estimating θ, we can predict the effect of any action.

Causal inference and policy optimisation

Example 6

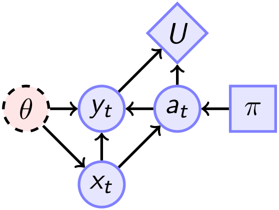


Figure: Decision problem with covariates.

Let $a_t, x_t = \{0, 1\}$, $y_t \in \mathbb{R}$, $\theta \in \mathbb{R}^4$ and

$$y_t \mid a_t = a, x_t = x \sim \mathcal{B}ernoulli(\theta_{a,x})$$

Then, by estimating θ, we can predict the effect of any action.

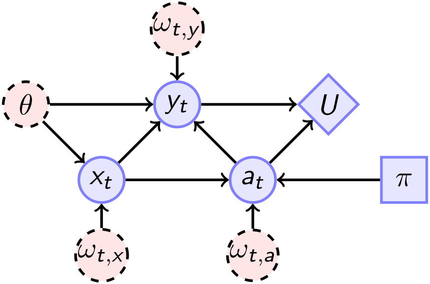


Figure: Decision diagram with exogenous disturbances ω.

Example 7 (Structural equation model for Figure 12)

$$\theta \sim \mathcal{N}(\mathbf{0}_4, \boldsymbol{I}_4),$$

$$x_t = \theta_0 \omega_{t,x}, \qquad \omega_{t,x} \sim \mathit{Bernoulli}(0.5)$$

$$y_t = \theta_1 y_t + \theta_2 x_t + \theta_3 a_t + \omega_{t,y}, \qquad \omega_{t,y} \sim \mathcal{N}(0,1)$$

$$a_t = \pi(x_t) + \omega_{t,a} \mod |\mathcal{A}| \qquad \omega_{t,a} \sim 0.1\,\mathcal{D}(0) + 0.9\,\mathcal{Unif}(\mathcal{A}),$$

Treatment-unit additivity

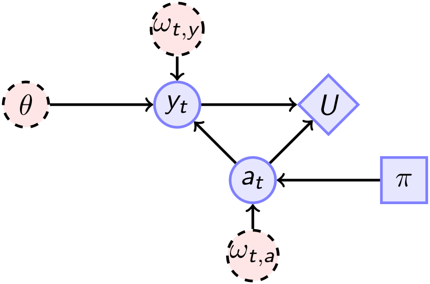


Figure: Decision diagram for treatment-unit additivity

Assumption 1 (TUA)

For any given treatment $a \in \mathcal{A}$, the response variable satisfies

$$y_t = g(a_t) + \omega_{t,y}$$

Example 8 (Pricing model)

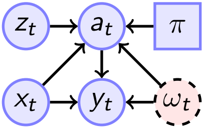


Figure: Graph of structural equation model for airport pricing policy π: a_t is the actual price, z_t are fuel costs, x_t is the customer type, y_t is the amount of sales, ω_t is whether there is a conference. The dependency on θ is omitted for clarity.

Assumption 2 (Relevance)

a_t depends on z_t.

Assumption 3 (Exclusion)

$z_t \perp\!\!\!\perp y_t \mid x_t, a_t, \omega_t$.

Assumption 4 (Unconfounded instrument)

$z_t \perp\!\!\!\perp \omega_t \mid x_t$.

Prediction tasks

$$y_t = g_\theta(a_t, x_t) + \omega_t, \qquad \mathbb{E}_\theta\, \omega_t = 0, \qquad \forall \theta \in \Theta \tag{4.1}$$

Standard prediction

$$\mathbb{P}^\pi_\theta(y_t \mid x_t, a_t), \qquad \mathbb{E}^\pi_\theta(y_t \mid x_t, a_t) = g_\theta(x_t, a_t) + \mathbb{E}^\pi_\theta(\omega_t \mid x_t, a_t).$$

Counterfactual prediction

$$\mathbb{E}^\pi_\theta(y_t \mid x_t, z_t) = \int_{\mathcal{A}} \underbrace{[g(a_t \mid x_t, z_t) + \mathbb{E}_\theta(\omega \mid x_t)]}_{h(a_t, x_t)} \mathrm{d}\pi(a_t \mid x_t)$$

Further reading

- Pearl, *Causality*.
- **?**