Bayesian Causal Induction

Pedro A. Ortega

Sensorimotor Learning and Decision-Making Group
MPI for Biological Cybernetics/Intelligent Systems

17th December 2011

Introduction

Causal Induction (AKA Causal Discovery):

- One of the oldest philosophical problems:
  - Aristotle, Kant, Hume, . . .
- The generalization from particular causal instances to abstract causal laws.

Introduction

Causal Induction (AKA Causal Discovery):

- One of the oldest philosophical problems:
  - Aristotle, Kant, Hume, ...
- The generalization from particular causal instances to abstract causal laws.
- Example:
  - 'I had a bad fall on wet floor.'
  - 'Therefore, it is dangerous to ride a bike on ice.'
  - ('Because I learned that a slippery floor can cause a fall')

Introduction

Causal Induction (AKA Causal Discovery):

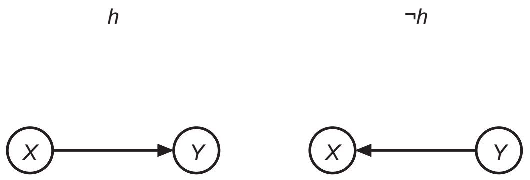
- One of the oldest philosophical problems:
  - Aristotle, Kant, Hume, . . .
- The generalization from particular causal instances to abstract causal laws.
- Example:
  - 'I had a bad fall on wet floor.'
  - 'Therefore, it is dangerous to ride a bike on ice.'
  - ('Because I learned that a slippery floor can cause a fall')
- Two important aspects:
  - Infer causal link from experience.
  - Extrapolate to future experience.

Introduction

Causal Induction (AKA Causal Discovery):

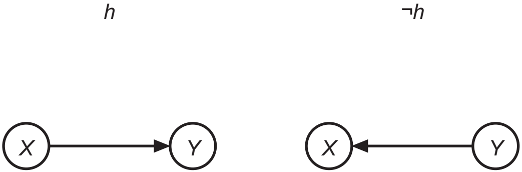
- One of the oldest philosophical problems:
  - Aristotle, Kant, Hume, . . .
- The generalization from particular causal instances to abstract causal laws.
- Example:
  - 'I had a bad fall on wet floor.'
  - 'Therefore, it is dangerous to ride a bike on ice.'
  - ('Because I learned that a slippery floor can cause a fall')
- Two important aspects:
  - Infer causal link from experience.
  - Extrapolate to future experience.
- We all do this in our everyday lives—**but how?**

Causal Graphical Model



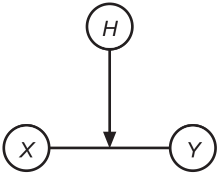
- A pair of (binary) random variables X and Y
- Two candidate **causal hypotheses** $\{h, \neg h\}$ (having identical joint distributions)

Causal Graphical Model



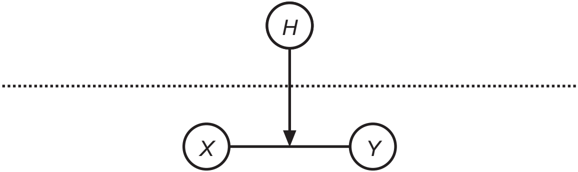
- A pair of (binary) random variables X and Y
- Two candidate **causal hypotheses** $\{h, \neg h\}$ (having identical joint distributions)
- How do we express the problem of causal induction using the language of graphical models **alone**?

Causal Graphical Model



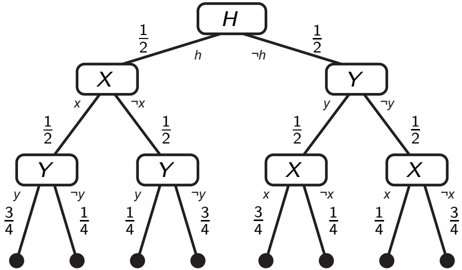
- A pair of (binary) random variables X and Y
- Two candidate **causal hypotheses** $\{h, \neg h\}$ (having identical joint distributions)
- How do we express the problem of causal induction using the language of graphical models **alone**?

Causal Graphical Model



- A pair of (binary) random variables X and Y
- Two candidate **causal hypotheses** $\{h, \neg h\}$ (having identical joint distributions)
- How do we express the problem of causal induction using the language of graphical models **alone**?
- Do we have to introduce a meta-level for H?

Probability Trees



- Node: mechanism, history dependent
  - e.g. $P(y|h, \neg x) = \frac{1}{4}$ and $P(\neg y|h, \neg x) = \frac{3}{4}$
- Path: causal realization of mechanisms
- Tree: causal realizations, possibly **heterogeneous**
- All random variables are first class citizens!

Inferring the Causal Direction

- We observe $X = x$, then we observe $Y = y$.
- What is the probability of $H = h$?
- Calculate posterior probability:

$$P(h|x,y) = \frac{P(y|h,x)P(x|h)P(h)}{P(y|h,x)P(x|h)P(h) + P(x|\neg h,y)P(y|\neg h)P(\neg h)}$$

$$= \frac{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}}{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}}$$

Inferring the Causal Direction

- We observe $X = x$, then we observe $Y = y$.
- What is the probability of $H = h$?
- Calculate posterior probability:

$$P(h|x,y) = \frac{P(y|h,x)P(x|h)P(h)}{P(y|h,x)P(x|h)P(h) + P(x|\neg h,y)P(y|\neg h)P(\neg h)}$$

$$= \frac{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}}{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}} = \frac{1}{2} = P(h)!$$

Inferring the Causal Direction

- We observe $X = x$, then we observe $Y = y$.
- What is the probability of $H = h$?
- Calculate posterior probability:

$$P(h|x,y) = \frac{P(y|h,x)P(x|h)P(h)}{P(y|h,x)P(x|h)P(h) + P(x|\neg h,y)P(y|\neg h)P(\neg h)}$$

$$= \frac{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}}{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}} = \frac{1}{2} = P(h)!$$

- We haven't learned anything!

Inferring the Causal Direction

- We observe $X = x$, then we observe $Y = y$.
- What is the probability of $H = h$?
- Calculate posterior probability:

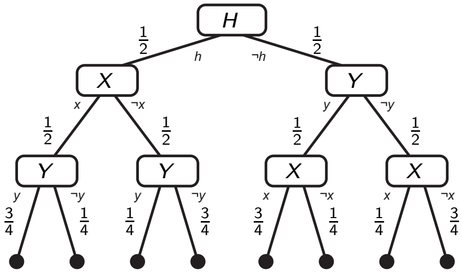
$$P(h|x,y) = \frac{P(y|h,x)P(x|h)P(h)}{P(y|h,x)P(x|h)P(h) + P(x|\neg h,y)P(y|\neg h)P(\neg h)}$$

$$= \frac{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}}{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}} = \frac{1}{2} = P(h)!$$

- We haven't learned anything!
- To extract **new** causal information, we have to supply **old** causal information:
  - "no causes in, no causes out"
  - "to learn what happens if you kick the system, you have to kick the system"

Interventions in a Probability Tree

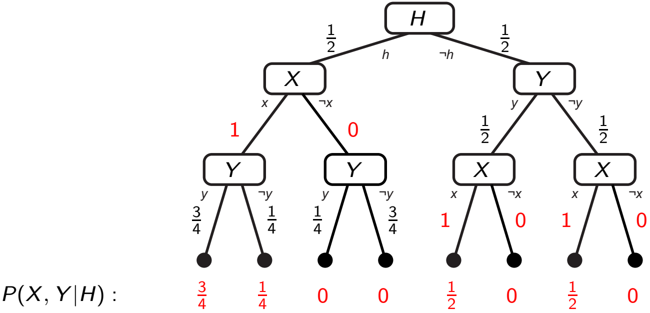
Set $X = x$:



$P(X, Y|H)$: $\frac{3}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{3}{8}$ $\frac{3}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{3}{8}$

Interventions in a Probability Tree

Set $X = x$:



- Replace all mechanisms resolving X with the delta "$X = x$".

Inferring the Causal Direction—2nd Attempt

- We set $X = x$, then we observe $Y = y$.
- What is the probability of $H = h$?

Inferring the Causal Direction—2nd Attempt

- We set $X = x$, then we observe $Y = y$.
- What is the probability of $H = h$?
- Calculate posterior probability:

$$P(h|\hat{x}, y) = \frac{P(y|h,\hat{x})P(\hat{x}|h)P(h)}{P(y|h,\hat{x})P(\hat{x}|h)P(h) + P(\hat{x}|\neg h, y)P(y|\neg h)P(\neg h)}$$

$$= \frac{\frac{3}{4} \cdot 1 \cdot \frac{1}{2}}{\frac{3}{4} \cdot 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \cdot \frac{1}{2}}$$

Inferring the Causal Direction—2nd Attempt

- We set $X = x$, then we observe $Y = y$.
- What is the probability of $H = h$?
- Calculate posterior probability:

$$P(h|\hat{x}, y) = \frac{P(y|h,\hat{x})P(\hat{x}|h)P(h)}{P(y|h,\hat{x})P(\hat{x}|h)P(h) + P(\hat{x}|\neg h, y)P(y|\neg h)P(\neg h)}$$

$$= \frac{\frac{3}{4}\cdot 1 \cdot \frac{1}{2}}{\frac{3}{4}\cdot 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}\cdot\frac{1}{2}} = \frac{3}{5} \neq P(h).$$

- We have have acquired evidence for "$X \rightarrow Y$"!

Conclusions

- Causal induction can be done using purely Bayesian techniques plus a description allowing multiple causal explanations of an experiment.
- Probability trees provide a clean & simple way to encode causal probabilistic information.
- The purpose of an intervention is to introduce statistical asymmetries.
- The causal information that we can acquire is limited by the interventions we can apply to the system.
- In this approach, the causal dependencies are not "in the data", but they rather arise from the data **and** the hypotheses that the reasoner "imprints" on them.