conditional sentences

and

causal reasoning

Lecture 2: Causal Bayes Nets

Katrin Schulz
ILLC
University of Amsterdam
GENERAL INTRODUCTION

Lecture 1: The logic of conditionals: The standard view

Tutorial 1: Challenges for the similarity approach

Lecture 2: Bayes Nets and Causal Bayes Nets

Tutorial 2: Counterfactuals as Interventions

SPECIFIC TOPICS

Seminar 1 (Practicum, part 1): 3 challenges for the framework
(I will first introduce the challenges individually, then you can choose and work in groups on one of them for 40 minutes)

Seminar 2: The relation between the similarity approach and the causal approach

Seminar 3: Using Logic Programming to model causal inferences

Seminar 4 (Practicum, part 2): presentations and discussion
(each group will shortly present their ideas, we will discuss them and I will comment on the state of the arts on each of the challenges)
Causal Bayesian Nets

Plan today

• introduce Bayesian Nets …
• … and then Causal Bayesian Nets (CBNs)
• model counterfactual reasoning with them

— BREAK —

• an alternative to CBN: structural equations
• model counterfactual reasoning with them
• discuss examples
Bayesian Nets

representation 1

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>¬A</th>
<th>B</th>
<th>¬B</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>.007125</td>
<td>.00015</td>
<td>.0007425</td>
<td>.296703</td>
</tr>
<tr>
<td>¬C</td>
<td>.002375</td>
<td>.00035</td>
<td>.0002475</td>
<td>.692307</td>
</tr>
</tbody>
</table>

representation 2

|   | P(A) | P(B|A) | P(C|B) |
|---|------|-------|-------|
| P(A) | .01  | .95   | .75   |
| P(B|A) | 0.001 | .3    |

P(a,b,c)=\prod P(c|ab)P(b|a)P(a)
=\prod P(c|b) P(b|a)P(a)
Bayesian Nets

representation 2

\[
P(a,b,c) = \prod P(c|ab) P(b|a) P(a)
= \prod P(c|b) \ P(b|a) P(a)
\]

|   | P(A) | P(B|A) | P(C|B) |
|---|------|--------|--------|
| 0 | .01  | .95    | .75    |
| 1 | 0,001| .3     |        |

P(A) = .01
P(B|A) = .95
P(C|B) = .75
Bayesian Nets

- A graph is a pair \( \langle V, N \rangle \), where \( V \) is a set of vertices and \( N \) is a set of nodes.
- A graph that is directed and doesn’t contain any directed circles is called a directed acyclic graph (DAG).
Bayesian Nets

A graph construction mechanism

Let \( P \) be a probability distribution over the variables \( X_1, X_2, \ldots, X_n \)
- If \( X_2 \) depends on \( X_1 \), then draw an arrow from \( X_1 \) to \( X_2 \), otherwise no arrow is drawn.
- If \( X_3 \) doesn’t depend on \( X_1 \) and \( X_2 \), then no arrow is drawn. Otherwise check whether one of \( X_1 \) and \( X_2 \) screens off \( X_3 \) from the other. If yes draw an arrow from this variable to \( X_3 \). If none screens off \( X_3 \) from the other, draw arrows from both variables to \( X_3 \).
- In general, for \( X_i \) let \( Z \) be the minimal set of predecessors that screens off \( X_i \) from all other predecessors. Then \( Z \) will be the set of parents \( \text{PA}_i \) of \( X_i \).
Bayesian Nets

\[ P(a,b,c) = \prod P(c|b)P(b|a)P(a) \]

\[ P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i|pa_i) \]

parents of \( X_i \): the set of variables pointing to \( X_i \)
Bayesian Nets

\[ P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i | \text{pa}_i) \]

parents of \( X_i \): the set of variables pointing to \( X_i \)

\[ P(a, b, c) = \prod P(c | b) P(b | a) P(a) \]
Bayesian Nets

The Markov condition

**Definition 1.** If a probability distribution $P$ admits the factorisation of (1) relative to a DAG $G$, we say that $P$ is Markov relative to $G$.

$$P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i | pa_i)$$

**Theorem 1.** A necessary and sufficient condition for a probability distribution $P$ to be Markov relative to a DAG $G$ is that every variable be independent of all its non-defendants (in $G$) conditional on its parents.
Bayesian Nets - example

P(C)?
P(D)?
P(A|C)?
P(B|C)?
P(B|C, D)?
P(A|C, D)?
Bayesian Nets - example

P(C) = 0.3061
P(D) = 0.0194
P(A|C) = 0.0003
P(B|C) = 0.0294
P(B|C, D) = 0.7422
P(A|C, D) = 0.0002

A | B | P(CIA,B)
---|---|---
1 | 1 | 0.99
1 | 0 | 0.9
0 | 1 | 0.9
0 | 0 | 0.3

A
---
P(A) = 0.0001

B
---
P(B) = 0.01

B | P(DIB)
---|---
1 | 0.95
0 | 0.01
Causal Bayesian Nets (CBN)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>P(AIA,B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.99</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>.9</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>.9</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

P(A) = .0001

P(B) = .01

<table>
<thead>
<tr>
<th>B</th>
<th>P(DIB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.95</td>
</tr>
<tr>
<td>0</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Causal Bayesian Nets (CBN)

A graph construction mechanism
Let P be a probability distribution over the variables $X_1, X_2, \ldots, X_n$

- If $X_2$ depends on $X_1$, then draw an arrow from to $X_2$, otherwise no arrow is drawn.
- If $X_3$ doesn’t depend on $X_1$ and $X_2$, then no arrow is drawn. Otherwise check whether one of $X_1$ and $X_2$ screens of $X_3$ from the other. If yes draw an arrow from this variable to $X_3$. If none screens of from the other, draw arrows from both variables to $X_3$.
- In general, for $X_i$ let $Z$ be the minimal set of predecessors that screens of $X_i$ from all other predecessors. Then $Z$ will be the set of parents $PA_i$ of $X_i$. 
Causal Bayesian Nets (CBN)

Why causal bayesian nets?

- the “true” graph
“One of the central ideas of the causal modelling framework is that stable probabilistic relations between the observed variables of a system are generated by an underlying causal structure. In other words, the world we see around us with all its uncertainty can be attributed to the operation of a big, complicated network of causal mechanisms. On a smaller scale, particular kinds of causal structure will lead to particular patterns of probability in the form of particular patterns of dependence and independence.”

—Sloman, 2005: 43
Causal Bayesian Nets (CBN)

Why causal bayesian nets?

- the “true” graph
- the causal net is a Bayesian Net with a minimal number of edges
Causal Bayesian Nets (CBN)

stochastic information cannot completely resolve the graph!!!

- 3 variables: A, B, C
- A depends on B, A depends on C, but B screens off A from C
Causal Bayesian Nets (CBN)

Why causal bayesian nets?

- the “true” graph
- the causal net is a Bayesian Net with a minimal number of edges
- allows us to model change in the world
Intervention
Intervention

**observation**
learning about the world

**intervention**
doing; changing the world
Intervention

intervention ¬B formally:

- cut variable off its parents
- fix the value of the variable
- leave everything else unchanged
Intervention

an example:

\[
P(D|\neg B)\
P_{\neg B}(D)\
P_{\neg B,A}(D)\
P_A(D)??
\]
Intervention

an example:

P(D|¬B)=0.055
P¬B(D)=0.0554
P¬B,A(D)=0.0955
PA(D)=0.47
Reasoning with CBN

two modes of reasoning in CBN:

1. **observational inference**: probabilistic update, belief change

2. **interventional inference**: interventionist update, reasoning from change in the world

based on these two Pearl defines a third mode:

3. **counterfactual inference**: what if X had been x, given that we observed that X=y?
counterfactual reasoning: what if B had been the case given that we observed ¬B?

1. update with ¬B (gives P*)
2. intervene with B
3. compute inferences

P*_B(D)=?
counterfactual reasoning: what if B had been the case given that we observed \( \neg B \)?

1. update with \( \neg B \) (gives \( P^* \))
2. intervene with B
3. compute inferences

\[ P^*_{B}(D)=0.4002 \]
Applications: decision taking

Causality and correlations:
An advertisement of an insurance company: “Independent research has shown that people who buy our insurance products live longer than the average American. Thus: ...

(1) If you buy our products, you will (probably) live longer.”

\[ EU(x) = \sum P(y|x)U(y) \]
Applications: decision taking

Causality and correlations:
An advertisement of an insurance company: “Independent research has shown that people who buy our insurance products live longer than the average American. Thus: ...

(1) If you buy our products, you will (probably) live longer.”

\[ EU(x) = \sum P_x(y)U(y) \]
Applications: decision taking

Causality and correlations:

An advertisement of an insurance company: “Independent research has shown that people who buy our insurance products live longer than the average American. Thus: ... (1) If you buy our products, you will (probably) live longer.”

$$EU(x) = \sum P_x(y)U(y)$$

- as far as testable the predictions turn out to be correct
References

good introduction (though for AI):


primary texts/overviews:


application to cognition:


application to philosophy of causation:


critical analyses: