Belief networks

Artificial intelligence (CK0031)

Francesco Corona

Belief networks

We make a first connection between probability and graph theory

- **Belief networks** introduce structure into a probabilistic model by using graphs to represent **independence assumptions** among vars
- Probability operations (marginalisation and conditioning) correspond to simple operations on the graph
- Details about the model can be 'read' from the graph
- There is a benefit in terms of computational efficiency

Belief networks (cont.)

Belief networks cannot capture all possible relations among variables

- They are natural for representing 'causal' relations, and they belong to the family of (**probabilistic**) **graphical models** we study further

Benefits of structure

Belief networks

Benefits of structure

The many possible ways variables can interact is extremely large

- Without assumptions we are unlikely to make a useful model
- Independently specifying all entries of a table $p(x_1, \dots, x_N)$ over binary variables x_i takes $\mathcal{O}(2^N)$ space, and might be impractical

This grow is infeasible in many application areas where we need to deal with distributions on potentially hundreds if not millions of variables

Structure is important for tractability of inferring quantities

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Benefits of structure (cont.)

Remark

Given a distribution on N binary variables, $p(x_1, \dots, x_N)$, computing a marginal $p(x_i)$ requires summing over the 2^{N-1} states of the other vars

- Even on the most optimistically fast supercomputer this would take too long, even for a $N = 100$ variable system

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Benefits of structure (cont.)

To render specification/inference in such systems tractable, the only way with such distributions is to constrain the nature of variable interactions

- The idea is to specify which variables are independent of others, to get a **structured factorisation** of the **joint probability distribution**
- For a distribution on a chain, $p(x_1, \dots, x_{100}) = \sum_{i=1}^{99} \phi(x_i, x_{i+1})$, computing a marginal $p(x_1)$ is fast

Belief networks are a valid framework for representing independence assumptions and they play a (quasi) natural role as '*causal*' models

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Benefits of structure (cont.)

Belief networks (BN, or Bayes' networks or Bayesian belief networks) are a way to depict the independence assumptions in a distribution

- Their application domain is widespread, ranging from expert reasoning under uncertainty to machine learning

Modelling independences

Benefits of structure

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies

Example

One morning Tracey leaves her house and realises that her grass is wet

- Is it due to overnight rain or did she forget to turn off the sprinkler?

Next she notices that the grass of her neighbour, Jack, is also wet

- This explains away to some extent the possibility that her sprinkler was left on, and she concludes that it has probably been raining

We can model the situation by defining the variables we wish to include

$$\begin{aligned}
R &\in \{0,1\} : R = 1 \text{ It has been raining} && (R = 0, \text{ otherwise})\\
S &\in \{0,1\} : S = 1 \text{ Tracey' sprinkler was on} && (S = 0, \text{ otherwise})\\
J &\in \{0,1\} : J = 1 \text{ Jack's grass is wet} && (J = 0, \text{ otherwise})\\
T &\in \{0,1\} : T = 1 \text{ Tracey's grass is wet} && (T = 0, \text{ otherwise})
\end{aligned}$$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies (cont.)

A model of Tracey's world corresponds to $p(T, J, R, S)$, a distribution on the joint set of variables of interest (the order of which is irrelevant)

Since each of the variables can take one of two states, it would appear that we have to specify the values for each of the $2^4 = 16$ states

- $p(T = 1, J = 0, R = 0, S = 1) = 0.057$
- ...

This is not truly true, there are normalisation conditions for probabilities

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies (cont.)

How many states need to be specified? Consider the decomposition ...

Without loss of generality and repeatedly using the definition of conditional probability, we may write

$$\begin{aligned}
p(T, J, R, S) &= p(T|J, R, S)p(J, R, S)\\
&= P(T|J, R, S)p(J|R, S)p(R, S) \qquad (1)\\
&= P(T|J, R, S)p(J|R, S)p(R|S)p(S)
\end{aligned}$$

The joint distro as a product of conditional distros

Modelling independencies (cont.)

$$p(T, J, R, S) = P(T|J, R, S)p(J|R, S)p(R|S)p(S)$$

The first term $p(T|J, R, S)$ requires us to specify $2^3 = 8$ values

- $p(T = 1|J, R, S)$ for the 8 joint states of (J, R, S)
- $p(T = 0|J, R, S) = 1 - p(T = 1|J, R, S)$, by normalisation

- $p(J = 1|R, S)$ for the 4 joint states of (R, S)
- $p(J = 0|R, S) = 1 - p(J = 1|R, S)$, by normalisation

- ...

Similarly, $2 + 1$ values for the other factors, a total of 15

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies (cont.)

Remark

In general, for a distribution on n binary variables,
we need to specify $2^n - 1$ values in the range $[0, 1]$

The important point: The number of values that need to be specified, in general, scales exponentially with the number of variables in the model

- This is impractical, in general, and motivates simplifications

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies - Conditional independence

The modeller often knows constraints on the system

Example

We may assume that ...

... Tracey's grass (T) is wet only depends directly on whether or not it has been raining (R) and whether or not her sprinkler (S) was on

- That is, we make a conditional independence assumption

$$p(T|J, R, S) = p(T|\cancel{J}, R, S) \tag{2}$$

- Similarly, assume that Jack's grass (J) is wet is influenced only directly by whether or not it has been raining (R),

$$p(J|R, S) = p(J|R, \cancel{S}) \tag{3}$$

- Moreover, assume the rain (R) is not directly influenced by the sprinkler (S)

$$p(R|S) = p(R|\cancel{S}) \tag{4}$$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies - Conditional independence (cont.)

$$p(T, J, R, S) = p(T|R, S)p(J|R)p(R)p(S) \tag{5}$$

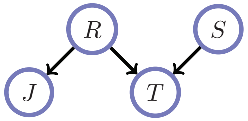
This reduces to $4 + 2 + 1 + 1 = 8$ the number of values to be specified

- A saving over the 15 values in the case where no conditional independencies had been assumed

Modelling independencies - Conditional independence (cont.)

We can represent these conditional independencies graphically

$$p(T, J, R, S) = p(T|R, S)p(J|R)p(R)p(S)$$



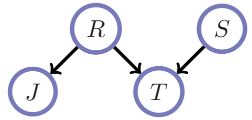
Each node in the graph represents a variable in the joint distribution

Variables which feed in (parents) to another variable (children) represent which variables are to the right of the conditioning bar

To complete the model, we need to specify the 8 values of each CPT

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

$$p(T, J, R, S) = p(T|R, S)p(J|R)p(R)p(S)$$



Let prior probabilities for R and S be

- $p(R = 1) = 0.2$
- $p(S = 1) = 0.1$

We set the remaining probabilities to

- $p(J = 1|R = 1) = 1.0$
- $p(J = 1|R = 0) = 0.2$ ⊗

- $p(T = 1|R = 1, S = 0) = 1.0$
- $p(T = 1|R = 1, S = 1) = 1.0$
- $p(T = 1|R = 0, S = 1) = 0.9$ ⊙
- $p(T = 1|R = 0, S = 0) = 0.0$

⊗ Jack's grass is wet due to unknown effects, other than rain

⊙ There is a small chance that even though the sprinkler was left on, it did not wet the grass noticeably

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies - Inference

$$p(T, J, R, S) = p(T|R, S)p(J|R)p(R)p(S)$$

We made a model of the environment, let us calculate the probability that the sprinkler was on overnight, given that Tracey's grass is wet

$$p(S = 1|T = 1)$$

$$\begin{aligned} p(S = 1|T = 1) &= \frac{p(S = 1, T = 1)}{p(T = 1)} = \frac{\sum_{J,R} p(T = 1, J, R, S = 1)}{\sum_{J,R,S} p(T = 1, J, R, S)} \\ &= \frac{\sum_{J,R} p(J|R)p(T = 1|R, S = 1)p(R)p(S = 1)}{\sum_{J,R,S} p(J|R)p(T = 1|R, S)p(R)p(S)} \\ &= \frac{\sum_{R} p(T = 1|R, S = 1)p(R)p(S = 1)}{\sum_{R,S} p(T = 1|R, S)p(R)p(S)} \end{aligned} \tag{6}$$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies - Inference (cont.)

$$\begin{aligned} p(S = 1|T = 1) &= \frac{(0.9 \cdot 0.8 \cdot 0.1) + (1 \cdot 0.2 \cdot 0.1)}{0.9 \cdot 0.8 \cdot 0.1 + 1 \cdot 0.2 \cdot 0.1 + 0 \cdot 0.8 \cdot 0.9 + 1 \cdot 0.2 \cdot 0.9} \\ &= 0.3382 \end{aligned}$$

The (posterior) belief that the sprinkler is on increases above the prior probability $p(S = 1) = 0.1$, due to the evidence that the grass is wet

Modelling independencies - Inference (cont.)

Remark

Note that the summation over J in the numerator is unity since, for any function $f(R)$, a summation of the form $\sum_J p(J|R)f(R)$ equals $f(R)$

- This follows from the definition that a distribution $p(J|R)$ must sum to one, and the fact that $f(R)$ does not depend on J
- A similar effect occurs for the summation over J in the denominator

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies - Inference (cont.)

$$p(T, J, R, S) = p(T|R, S)p(J|R)p(R)p(S)$$

Let us calculate the probability that Tracey's sprinkler was on overnight, given that her and Jack's grass are wet

$$p(S = 1|T = 1, J = 1)$$

We use conditional probability again:

$$\begin{aligned} p(S = 1|T = 1, J = 1) &= \frac{p(S = 1, T = 1, J = 1)}{p(T = 1, J = 1)} \\ &= \frac{\sum_R p(T = 1, J = 1, R, S = 1)}{\sum_{R,S} p(T = 1, J = 1, R, S)} \\ &= \frac{\sum_R p(J = 1|R)p(T = 1|R, S = 1)p(R)p(S)}{\sum_{R,S} p(J = 1|R)p(T = 1)p(R)p(S)} \end{aligned} \quad (7)$$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies - Inference (cont.)

$$p(S = 1|T = 1, J = 1) = \frac{0.0344}{0.2144} = 0.1604$$

Probability that the sprinkler is on, given extra evidence (Jack's wet grass), is lower than it is given only that Tracey's grass is wet (0.34)

- This occurs since the fact that Jack's grass is also wet increases the chance that the rain has played a role in making Tracey's grass wet

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies (cont.)

Example

Sally comes home to find that the burglar alarm is sounding ($A = 1$)

- Has she been burgled ($B = 1$), or was it an earthquake ($E = 1$)?

Soon, she finds that the radio broadcasts an earthquake alert ($R = 1$)

Using Bayes' rule, we can write

$$p(B, E, A, R) = p(A|B, E, R)p(R|B, E)p(E|B)p(B) \quad (8)$$

However, the alarm is surely not directly influenced by radio reports

$$P(A|B, E, R) = p(A|B, E, \cancel{R})$$

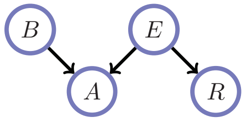
and we can other conditional independence assumptions such that

$$p(B, E, A, R) = p(A|B, E)p(R|\cancel{B}, E)P(E|\cancel{B})p(B) \quad (9)$$

Modelling independencies (cont.)

$$p(B, E, A, R) = p(A|B, E)p(R|E)P(E)p(B)$$

Graphical representation of the factorised joint and CPT specification



$A = 1$ (Alarm is on)	B (Burglar)	E (Earthquake)
0.9999	1	1
0.99	1	0
0.99	0	1
0.0001	0	0

$p(B = 1) = 0.01$

$p(E = 1) = 0.000001$

$R = 1$ (Earthquake alert)	E (Earthquake)
1	1
0	0

The tables and graphical structure fully specify the distribution

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies (cont.)

What happens when we observe evidence?

- Initial evidence: The alarm is sounding

$$\begin{aligned} p(B = 1|A = 1) &= \frac{\sum_{E,R} p(B = 1, E, A = 1, R)}{\sum_{B,E,R} p(B, E, A = 1, R)} \\ &= \frac{\sum_{E,R} p(A = 1|B = 1, E)p(B = 1)p(E)p(R|E)}{\sum_{B,E,R} p(A = 1|B, E)p(B)p(E)p(R|E)} \\ &\simeq 0.99 \end{aligned} \tag{10}$$

- Additional evidence: The earthquake alarm is broadcasted and a similar calculation gives $p(B = 1|A = 1, R = 1) \simeq 0.01$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Modelling independencies (cont.)

Remark

Causal intuitions: BNs, as defined, express independence statements

- In expressing these independencies it can be useful, though potentially misleading, to think of '*what causes what*'

The ordering of variables is used to reflect our intuition on root causes

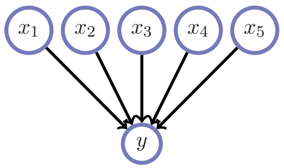
Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Reducing specifications

Benefits of structure

Reducing specifications

Consider a discrete variable y with discrete parental variables $x_1, \ldots, x_n$



Formally, the structure of the graph implies nothing about the form of the parameterisation of the table

$$p(y|x_1, \ldots, x_n)$$

If all variables are binary, $2^5 = 32$ states to specify $p(y|x_1, \ldots, x_n)$

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Reducing specifications (cont.)

Remark

If each parent x_i has $\dim(x_i)$ states and there is no constraint on the table, then $p(y|x_1, \ldots, x_n)$ contains $(\dim(y) - 1) \prod_i \dim(x_i)$ entries

- If stored explicitly for each state, a potentially huge storage

An alternative is to constrain the table to a simpler parametric form

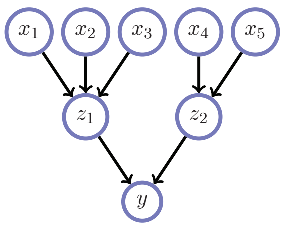
Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Reducing specifications (cont.)

Divorcing parents: One might write a decomposition in which only a limited number of parental interactions are required



Assuming all variables are binary, $2^3 + 2^2 + 2^2 = 16$ states require specification, $2^5 = 32$ states in the unconstrained case

$$p(y|x_1, \ldots, x_n) = \sum_{z_1, z_2} p(y|z_1, z_2) p(z_1|x_1, x_2, x_3) p(z_2|x_4, x_5) \quad (11)$$

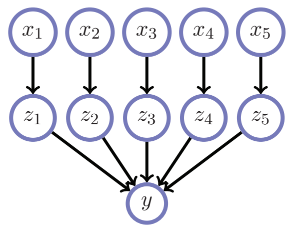
Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Reducing specifications (cont.)

Logical gates: Another technique to constrain tables uses simple classes of conditional tables



Use a logical OR gate on binary z_i

$$p(y|z_1, \ldots, z_5) = \begin{cases} 1 & \text{if at least one } z_i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

We can make table $p(y|x_1, \ldots, x_5)$ by including terms $p(z_i = 1|x_i)$

- When each x_i is binary there are $2 + 2 + 2 + 2 + 2 = 10$ quantities that are required for specifying $p(y|x)$

Reducing specifications (cont.)

Remark

The graph can be used to represent any **noisy logical state**, such as the **noisy OR** or **noisy AND**, where the number of parameters required to specify the noisy gate is linear in the number of parents

The noisy-OR is particularly common in disease-symptom networks in which many diseases x can give rise to the same symptom y

Provided that at least one of the diseases is present, the probability that the symptom will be present is high

Uncertain and unreliable evidence

Belief networks

Uncertain and unreliable evidence

We now make a distinction between two types of evidence

- **Evidence** that is **uncertain**
- **Evidence** that is **unreliable**

Uncertain evidence

Uncertain and unreliable evidence

Uncertain evidence

In **soft** or **uncertain evidence**, evidence is in more than one state, with the strength of our belief about each state being given by probabilities

- If x has states $\text{dom}(x) = \{\texttt{red}, \texttt{blue}, \texttt{green}\}$, then vector $(0.6, 0.1, 0.3)$ represents the belief in the states

In **hard evidence**, we are certain that a variable is a particular state

- All the probability mass is in one vector component $(0, 0, 1)$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Uncertain evidence (cont.)

Performing inference with soft-evidence is straightforward (Bayes' rule)

- For model $p(x, y)$, consider some soft evidence $\tilde{y}$ about variable y, we wish to know the effect this has on variable x, $p(x|\tilde{y})$
- Compute $p(x|\tilde{y})$, under the assumption that $p(x|y, \tilde{y}) = p(x|y)$

$$\begin{aligned} p(x|\tilde{y}) &= \sum_y p(x, y|\tilde{y}) = \sum_y p(x|y, \tilde{y}) p(y|\tilde{y}) \\ &= \sum_y p(x|y) p(y|\tilde{y}) \end{aligned} \tag{13}$$

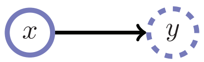
$p(y = \texttt{i}|\tilde{y})$ is the probability of y in state $\texttt{i}$ under soft-evidence

- This is a generalisation of hard-evidence in which vector $p(y|\tilde{y})$ has all zeros except for single component

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Uncertain evidence (cont.)

The procedure in which we define the model conditioned on evidence, and then average over the distribution of the evidence is **Jeffrey's rule**



In the BN we use a dashed circle to represent that a variable is in a soft-evidence state

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Uncertain evidence (cont.)

Example

Soft-evidence: We can revisit the burglar scenario by imagining that we are only 70% sure we heard the burglar alarm sounding

For this binary variable case, we represent soft-evidence for states $(1, 0)$

$$\tilde{A} = (0.7, 0.3)$$

What is the probability of a burglar under the soft-evidence?

$$\begin{aligned} p(B = 1|\tilde{A}) &= \sum_A p(B = 1|A) p(A|\tilde{A}) \\ &= p(B = 1|A = 1) \times 0.7 + p(B = 1|A = 0) \times 0.3 \\ &\simeq 0.6930 \end{aligned} \tag{14}$$

$p(B = 1|A = 1) \simeq 0.99$ and $p(B = 1|A = 0) \simeq 0.0001$ from Bayes' rule

- This is lower than 0.99, the probability of having been burgled when we are sure we heard the alarm

Uncertain evidence - Holmes, Watson and Mrs Gibbon

An entertaining example with an environment containing four variables

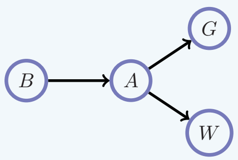
Example

$B \in \{\texttt{tr}, \texttt{fa}\}$: $B = \texttt{tr}$ Holmes' house has been Burgled

$A \in \{\texttt{tr}, \texttt{fa}\}$: $A = \texttt{tr}$ Holmes' house Alarm went off

$W \in \{\texttt{tr}, \texttt{fa}\}$: $W = \texttt{tr}$ Watson heard the alarm

$G \in \{\texttt{tr}, \texttt{fa}\}$: $G = \texttt{tr}$ Mrs Gibbon heard the alarm



The BN and a factorisation of the joint probability for this scenario

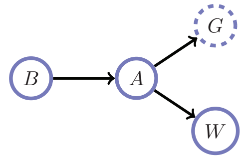
$$p(B, A, G, W) = p(A|B)p(B)p(W|A)p(G|A) \qquad (15)$$

Watson states (100% sure) that he heard the alarm sounding, whereas Mrs Gibbon is a little deaf and cannot be sure she heard it (80% sure)

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

From Jeffrey's rule, we can use the original model equation to compute the model conditioned on the evidence

$$\begin{aligned} p(B = \texttt{tr}|W = \texttt{tr}, G) &= \frac{p(B = \texttt{tr}, W = \texttt{tr}, G)}{p(W = \texttt{tr}, G)} \\ &= \frac{\sum_A p(G|A)p(W = \texttt{tr}|A)p(A|B = \texttt{tr})p(B = \texttt{tr})}{\sum_{B,A} p(G|A)p(W = \texttt{tr})p(A|B)p(B)} \end{aligned} \qquad (16)$$



and then we can use soft evidence

$$P(G|\tilde{G}) = \begin{cases} 0.8 & G = \texttt{tr} \\ 0.2 & G = \texttt{fa} \end{cases} \qquad (17)$$

$$\begin{aligned} p(B = \texttt{tr}|W = \texttt{tr}, \tilde{G}) = & p(B = \texttt{br}|W = \texttt{tr}, G = \texttt{tr})p(G = \texttt{tr}|\tilde{G}) + \\ & p(B = \texttt{tr}|W = \texttt{br}, G = \texttt{fa})p(G = \texttt{fa}|\tilde{G}) \end{aligned} \qquad (18)$$

A full calculation requires us to numerically specify all terms in Eq. 15

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Unreliable evidence

Uncertain and unreliable evidence

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Unreliable evidence

Example

Holmes telephones Mrs Gibbon and realises that he does not trust her evidence and his interpretation is that if the alarm sounded it was 80% probable to have resulted in Mrs Gibbon stating that she heard it

- If the alarm did not sound, there is a 20% chance that Mrs Gibbon would have stated she heard it

This is not like Mrs Gibbon being 80% sure herself that she heard the alarm (soft-evidence, whose effect is calculations containing $p(G|A)$)

Unreliable evidence (cont.)

Holmes will discard all of this and replace it with his own interpretation

- He can do that by replacing $p(G|A)$ by **virtual evidence**

$$p(G|A) \rightarrow p(\mathtt{H}|A), \quad \text{where } p(\mathtt{H}|A) = \begin{cases} 0.8 & A = \mathtt{tr} \\ 0.2 & A = \mathtt{fa} \end{cases} \tag{19}$$

Here the state $\mathtt{H}$ is arbitrary and fixed and it is used to modify the joint

$$p(B, A, \mathtt{H}, W) = p(A|B)p(B)p(W|A)p(\mathtt{H}|A) \tag{20}$$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Unreliable evidence (cont.)

The effect of Holmes' judgement when computing $p(B = \mathtt{tr}|W = \mathtt{tr}, \mathtt{H})$ counts 4 times more in favour of the alarm sounding than not

- The value of the table entries are irrelevant up to normalisation
- Any constants can be absorbed into the proportionality constant

$p(\mathtt{H}|A)$ is not a distribution over A, and so no normalisation is required

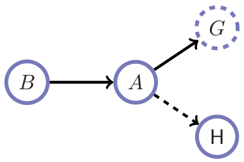
Remark

This form of evidence is called **likelihood evidence**

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Uncertain and unreliable evidence

To demonstrate how to combine such effects as unreliable and uncertain evidence, consider the situation in which Mrs Gibbon is uncertain in her evidence and Holmes feels that Watson's evidence is unreliable



Holmes wishes to use its own interpretation

We first deal with unreliable evidence:

$$p(A, B, W, G) \rightarrow p(B, A, \mathtt{H}, G) = p(B)p(A|B)p(G|A)p(\mathtt{H}|A) \tag{21}$$

We use Jeffrey's rule to compute a model conditioned on evidence

$$p(B, A|\mathtt{H}, G) = \frac{p(B)p(A|B)p(G|A)p(\mathtt{H}|A)}{\sum_{A,B} p(B)p(A|B)p(G|A)p(\mathtt{H}|A)} \tag{22}$$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Uncertain and unreliable evidence (cont.)

We include uncertain evidence $\tilde{G}$ to form the final model

$$p(B, A|\mathtt{H}, \tilde{G}) = \sum_G p(B, A|\mathtt{H}, G)p(G|\tilde{G}) \tag{23}$$

from which we may then compute the marginal $p(B|\mathtt{H}, \tilde{G})$

$$p(B|\mathtt{H}, \tilde{G}) = \sum_A p(B, A|\mathtt{H}, \tilde{G}) \tag{24}$$

Belief networks

Belief networks

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Belief networks

Definition

Belief networks: A belief network is a distribution of form

$$p(x_1, \dots, x_D) = \prod_{i=1}^{D} p(x_i | \text{pa}(x_i)) \quad (25)$$

where $\text{pa}(x_i)$ represent the **parental variables** of variable x_i

As directed graph, with an arrow pointing from parent to child, a Bayesian BN corresponds to a Directed Acyclic Graph (DAG)

- The i-th node in the graph corresponds to factor $p(x_i|\text{pa}(x_i))$

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Belief networks (cont.)

Remark

Graphs and distributions: A subtle point is whether a BN corresponds to an instance of a distro requiring specification of the CPTs or not

- Or, whether or not it refers to any distribution which is consistent with the graph structure

In this one, one can distinguish

- A BN distribution (with numerical specification)
- A BN graph (without numerical specification)

Important to clarify the scope of independence/dependence statements

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Belief networks (cont.)

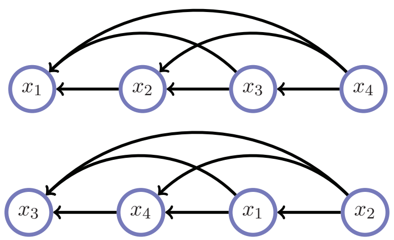
Remark

In the grass and burglar cases, WE chose how to recursively use Bayes

Belief networks (cont.)

$$\begin{aligned} p(x_1, x_2, x_3, x_4) &= p(x_1|x_2, x_3, x_4)p(x_2|x_3, x_4)p(x_3|x_4)p(x_4) \\ &= p(x_3|x_4, x_1, x_2)p(x_4|x_1, x_2)p(x_1|x_2)p(x_2) \end{aligned} \quad (26)$$

The two choices are equivalently valid and the two associated graphs, though different, represent the same independence assumptions



Both graphs represent the same distribution

$$p(x_1, \dots, x_4)$$

Both graphs, the same lack of independence assumptions

To make independence assumptions, the choice of factorisation is crucial

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Belief networks (cont.)

The observation that any distribution may be written in the **cascade form** suggests an algorithm for constructing a BN on vars $x_1, \dots, x_n$

1. write the n-node cascade graph, label the nodes with the variables in any order
2. each successive independence statement corresponds to deleting one of the edges

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Belief networks (cont.)

Definition

More formally, this corresponds to an ordering of the variables which, without loss of generality, we may write as $x_1, \dots, x_n$, from Bayes' rule

$$\begin{aligned} p(x_1, \dots, x_n) &= p(x_1|x_2, \dots, x_n)p(x_2, \dots, x_n) \\ &= p(x_1|x_2, \dots, x_n)p(x_2|x_3, \dots, x_n)p(x_3, \dots, x_n) \\ &= \cdots \\ &= p(x_n)\sum_{i=1}^{n-1} p(x_i|x_i, \dots, x_n) \end{aligned} \quad (27)$$

The representation of any BN is thus a **Direct Acyclic Graph** (**DAG**)

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Belief networks (cont.)

Remark

Every probability distribution can be written as a Bayesian BN, though it may correspond to a fully connected 'cascade' DAG

The role of a BN is that the structure of the DAG corresponds to a set of conditional independence assumptions of variables on their ancestors

- Namely, which ancestral parental variables are sufficient to specify each conditional probability table
- This does not mean that non-parental variables have no influence

Belief networks (cont.)

Example

Consider the distribution $p(x_1|x_2)p(x_2|x_3)p(x_3)$ with DAG $x_1 \leftarrow x_2 \leftarrow x_3$

- This does not imply $p(x_2|x_1, x_3) = p(x_2|x_3)$

The DAG specifies conditional independence statements of variables on their ancestors (which ancestors are direct 'causes' for the variable)

- The effects, given by the descendants of the variable, will generally be dependent on the variable

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

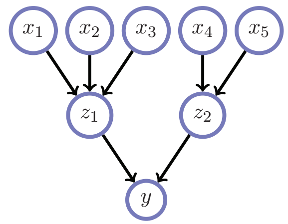
Belief networks (cont.)

Remark

Dependencies and **Markov blanket**: Consider a distribution on a set of variables $\mathcal{X}$ and for a variable $x_i \in \mathcal{X}$ and corresponding BN represented by a DAG G, let $\text{MB}(x_i)$ be the variables in the Markov blanket of x_i

- For any other variable y that is also not in the Markov blanket of x_i $(y \in \mathcal{X}) \backslash \{x_i \cup \text{MB}(x_i)\}$ then $x_i \perp\!\!\!\perp y | \text{MB}(x_i)$

The Markov blanket of x_i carries all information about x_i



$\text{MB}(z_1) = \{x_1, x_2, x_3, y, z_2\}$

$z_1 \perp\!\!\!\perp x_4 | \text{MB}(z_1)$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Belief networks (cont.)

The DAG corresponds to a statement of conditional independencies

- To complete the specification of the BN, we need to define all the elements of the conditional probability tables $p(x_i|\text{pa}(x_i))$

Once the structure is defined, the entries of the CPTs are expressed

For every possible state of the parental variables $\text{pa}(x_i)$, a value for each of the states of x_i needs to be specified (except one, by normalisation)

- For a large number of parents, the specification is intractable
- Tables are usually parameterised in a low-dimensional manner

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Conditional independence
Belief networks

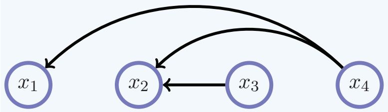
Conditional independence (cont.)

A BN corresponds to sets of conditional independence assumptions

- It is not always immediately clear from the DAG whether a set of variables is conditionally independent of a set of other variables

Example

$$p(x_1, \dots, x_4) = p(x_1|x_4)p(x_2|x_3, x_4)p(x_3)p(x_4)$$



Are x_1 and x_2 independent, given the state of x_4?

Belief networks
UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Conditional independence (cont.)

$$\begin{aligned} p(x_1, x_2|x_4) &= \frac{1}{p(x_4)} \sum_{x_3} p(x_1, x_2, x_3, x_4) \\ &= \frac{1}{p(x_4)} \sum_{x_3} p(x_1|x_4)p(x_2|x_3, x_4)p(x_3)p(x_4) \\ &= p(x_1|x_4) \sum_{x_3} p(x_2|x_3, x_4)p(x_3) \end{aligned} \tag{28}$$

$$\begin{aligned} p(x_2|x_4) &= \frac{1}{p(x_4)} \sum_{x_1, x_3} p(x_1, x_2, x_3, x_4) \\ &= \frac{1}{p(x_4)} \sum_{x_1, x_3} p(x_1|x_4)p(x_2|x_3, x_4)p(x_3)p(x_4) \\ &= \sum_{x_3} p(x_2|x_3, x_4)p(x_3) \end{aligned} \tag{29}$$

Belief networks
UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Conditional independence (cont.)

Combining the two results, we have $P(x_1, x_2|x_4) = p(x_1|x_4)p(x_2|x_4)$

- Hence, variable x_1 and x_2 are independent conditioned on x_4

Belief networks
UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Conditional independence (cont.)

We would like to have a general algorithm that allows to avoid doing such tedious manipulations by reading the results directly from a graph

Example

To help develop intuitions towards building such algorithm

- We consider the three variable distribution $p(x_1, x_2, x_3)$

We can write the distribution in a total of six ways of type

$$p(x_1, x_2, x_3) = p(x_{i_1}|x_{i_2}, x_{i_3})p(x_{i_2}|x_{i_3})p(x_{i_3}) \tag{30}$$

where (i_1, i_2, i_3) is any of the six permutations of $(1, 2, 3)$

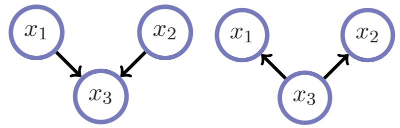
Each factorisation produces a different DAG, all representing the same distribution and none making independence statement

- If DAGs are cascades, no independence assumptions were made

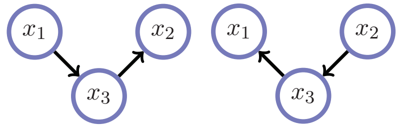
Conditional independence (cont.)

Minimal independence assumptions correspond to dropping any link

- Say, we drop the link between x_1 and x_2



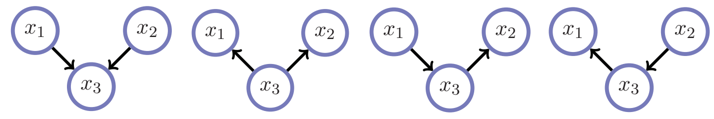
(a) $x_1 \to x_2 / x_2 \to x_1$ (b) $x_1 \to x_2 / x_2 \to x_1$



(c) $x_1 \to x_2$ (d) $x_2 \to x_1$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Conditional independence (cont.)



(a) (b) (c) (d)

Are theses graphs equivalent in representing the same distribution?

$$\underbrace{p(x_2|x_3)p(x_3|x_1)p(x_1)}_{\text{graph (c)}} = \frac{p(x_2,x_3)p(x_3,x_1)}{p(x_3)} = p(x_1|x_3)p(x_2,x_3)$$
$$= \underbrace{p(x_1|x_3)p(x_3|x_2)p(x_2)}_{\text{graph (d)}} = \underbrace{p(x_1|x_3)p(x_2|x_3)p(x_3)}_{\text{graph (b)}} \quad (31)$$

DAGs (b), (c) and (d) represent the same conditional independence assumptions (given x_3, x_1 and x_2 are independent $x_1 \perp\!\!\!\perp x_2 | x_3$)

- DAG (a) is fundamentally different ($p(x_1,x_2) = p(x_1)p(x_2)$) and there is no way to transform $p(x_3|x_1,x_2)p(x_1)p(x_2)$ into the others

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Conditional independence (cont.)

Remark

Graphical dependence: BNs (graphs) are good for encoding conditional independence but they are not appropriate for encoding dependence

Graph $a \to b$ may seem to encode a relation that a and b are dependent

- However, a specific numerical instance of a BN distribution could be such that $p(b|a) = p(b)$ for which we have $a \perp\!\!\!\perp b$

When a graph appears to show 'graphical' dependence there can be instances of the distributions for which dependence does not follow

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

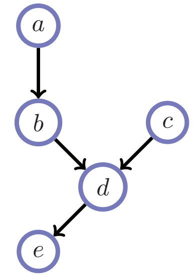
Impact of collisions

Belief networks

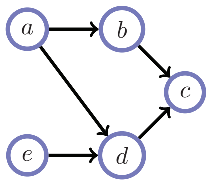
Impact of collisions

Definition

Given a path $\mathcal{P}$, a **collider** is a node c on $\mathcal{P}$ with neighbours a and b on $\mathcal{P}$ such that $a \rightarrow c \leftarrow b$



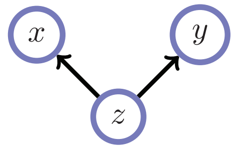
Variable d is a collider along path $a - b - d - c$, but not along path $a - b - d - e$



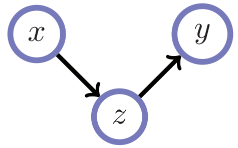
Variable d is a collider along path $a - d - e$, but not along path $a - b - c - d$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

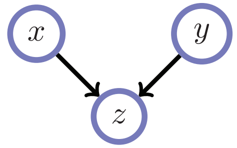
Impact of collisions (cont.)



(a) z is a not collider



(b) z is a not collider

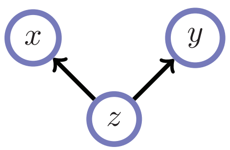


(c) z is a collider

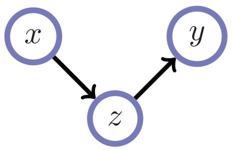
Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Impact of collisions (cont.)

In a general BN, how can we check if $x \perp\!\!\!\perp y|z$?



(a) $x \perp\!\!\!\perp y|z$



(b) $x \perp\!\!\!\perp y|z$

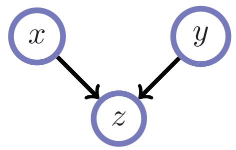
In these DAGs, x and y are independent given z

(a) Since $p(x, y|z) = p(x|z)p(y|z)$

(b) Since $p(x, y|z) \propto p(z|x)p(x)p(y|z)$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

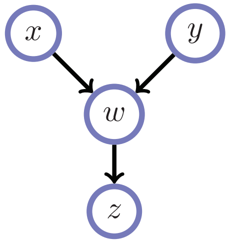
Impact of collisions (cont.)



(a) $x \top\!\!\!\top y|z$

In this DAG, x and y are graphically dependent given z

(c) Since $p(x, y|z) \propto p(z|x, y)p(x)p(y)$



(a) $x \top\!\!\!\top y|z$

When we condition on z, x and y will be graphically dependent

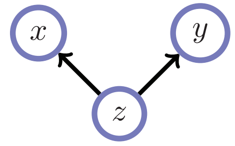
$$p(x, y, |z) = \frac{p(x, y, z)}{p(z)} = \frac{1}{p(z)} \sum_w p(z|w)p(w|x, y)p(x)p(y)$$
$$\neq p(x|z)p(y|z)$$

The inequality holds due to the term $p(w|x, y)$ and only in special cases such as $p(w|x, y) = const$ would x and y be independent

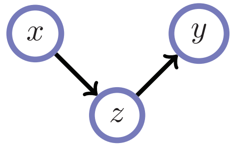
- w becomes dependent on the value of z, and since x and y are conditionally dependent on w, are conditionally dependent on z

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

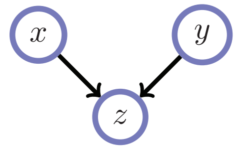
Impact of collisions (cont.)



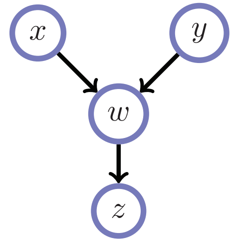
(a) $x \perp\!\!\!\perp y|z$



(b) $x \perp\!\!\!\perp y|z$



(c) $x \top\!\!\!\top y|z$

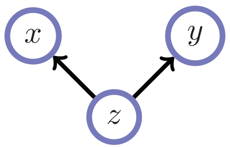


(d) $x \top\!\!\!\top y|z$

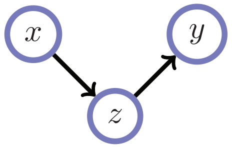
- If there is a non-collider z, conditioned along the path between x and y, then this path cannot induce dependence between x and y

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

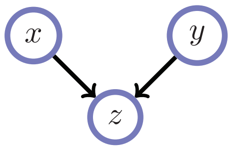
Impact of collisions (cont.)



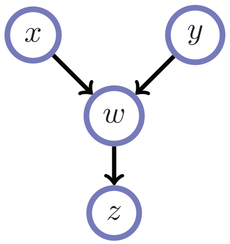
(a) $x \perp\!\!\!\perp y|z$



(b) $x \perp\!\!\!\perp y|z$



(c) $x \top\!\!\!\top y|z$

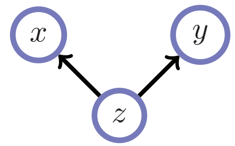


(d) $x \top\!\!\!\top y|z$

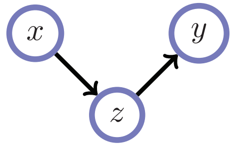
- If there is a path between x and y which contains a collider, and this collider is not in the conditioned set and neither are any of its descendants, then this path does not make x and y dependent

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

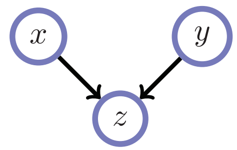
Impact of collisions (cont.)



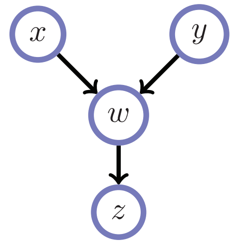
(a) $x \perp\!\!\!\perp y|z$



(b) $x \perp\!\!\!\perp y|z$



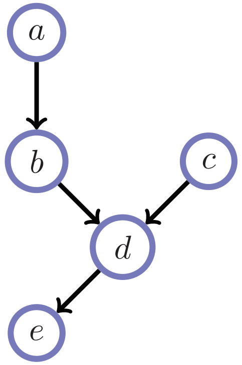
(c) $x \top\!\!\!\top y|z$



(d) $x \top\!\!\!\top y|z$

- If there is path between x and y which contains no colliders and no conditioning variables, then this path '**d-connects**' x and y

Impact of collisions (cont.)



Variable d is a collider along the path $a - b - d - c$ but not along the path $a - b - d - c$

- Is $a \perp\!\!\!\perp e | b$?

a and e are not d-connected (no colliders on the path between them) and since there is a non-collider b which is in the conditioning set

- Hence, a and e are d-separated by b

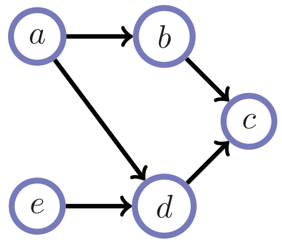
$$a \perp\!\!\!\perp e | b$$

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Impact of collisions (cont.)



Variable d is a collider along the path $a - d - e$ but not along the path $a - b - c - d - e$

- Is $a \perp\!\!\!\perp e | c$?

There are two paths between a and c ($a - b - c - d - e$ and $a - d - e$)

Path $a - d - e$ is not blocked since, although d is a collider on this path, and d is not in the conditioning set we have a descendant of the collider d in the conditioning set, c

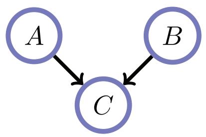
Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Impact of collisions (cont.)

Consider $A \rightarrow B \leftarrow C$ with A and C are (unconditionally) independent



$$p(A, B, C) = p(C|A, B)p(A)p(B)$$

- Conditioning of B makes them 'graphically' dependent

From a 'causal' perspective, this models the 'causes' A and B as a priori independent, both determining effect C

Intuitively, whilst we believe the root causes are independent given the value of the observation, this tells us something about the state of both the causes coupling them and making them (generally) dependent

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Impact of collisions (cont.)

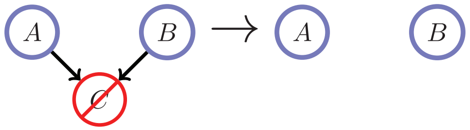
On the effect that conditioning/marginalisation has on the graph of the remaining variables

Definition

Some properties of belief networks: It is useful to understand what effect conditioning or marginalising a variable has on a belief network

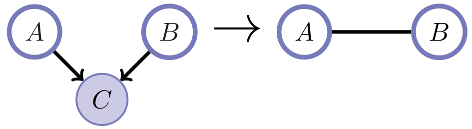
- We state how these operations effect the remaining variables in the graph and use this intuition to develop a more complete description

Impact of collisions (cont.)



- Marginalising over C makes A and B independent
- A and B are conditionally independent $p(A, B) = p(A)p(B)$
- In the absence of any info about effect C, we retain this belief

Impact of collisions (cont.)

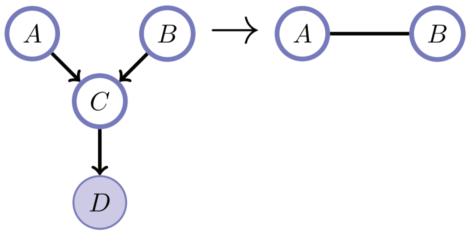


Conditioning on C makes A and B (graphically) dependent

In general, $p(A, B|C) \neq p(A|C)p(B|C)$

- Although the causes are a priori independent, knowing the effect C, in general, tells us something about how the causes colluded to bring about the effect observed

Impact of collisions (cont.)

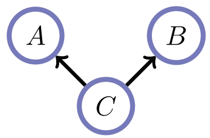


Conditioning on D, a descendent of collider C, makes A and B (graphically) dependent

In general, $p(A, B|D) \neq p(A|D)p(B|D)$

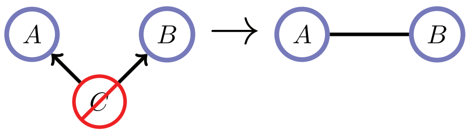
Impact of collisions (cont.)

$$P(A|, B, C) = p(A|C)p(B|C)p(C)$$



Here there is a 'cause' C and independent 'effects' A and B

Impact of collisions (cont.)

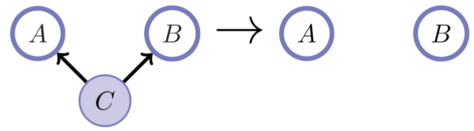


Marginalising over C makes A and B (graphically) dependent

In general, $p(A, B) \neq p(A)P(B)$

Though we do not know the 'cause', the 'effects' will be dependent

Impact of collisions (cont.)



Conditioning on C makes A and B independent

$$p(A, B|C) = p(A|C)p(B|C)$$

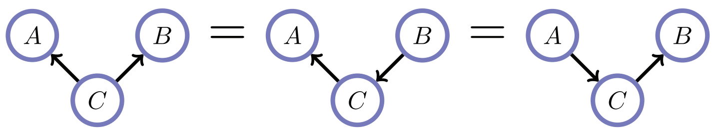
If you know 'cause' C, you know everything about how each effect occurs, independent of the other effect

This is also true from reversing the arrow from A to C

- A would 'cause' C and then C would 'cause' B

Conditioning on C blocks the ability of A to influence B

Impact of collisions (cont.)



These graphs express the same conditional independence assumptions

Path manipulations

Belief networks

Graphical path manipulations for independence

We understand when x is independent of y, conditioned on z ($x \perp\!\!\!\perp y|z$)

- We need to look at each path between x and y

Colouring x as red, y as green and the conditioning node z as yellow

- We examine each path between x and y and adjust the edges, following some intuitive results

After the manipulations, if there is no undirected path between x and y, then x and y are independent, conditioned on z

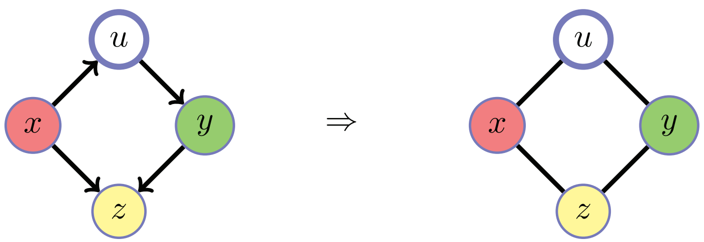
Graphical path manipulations for independence (cont.)

Remark

The graphical rules we define here differ from those provided earlier

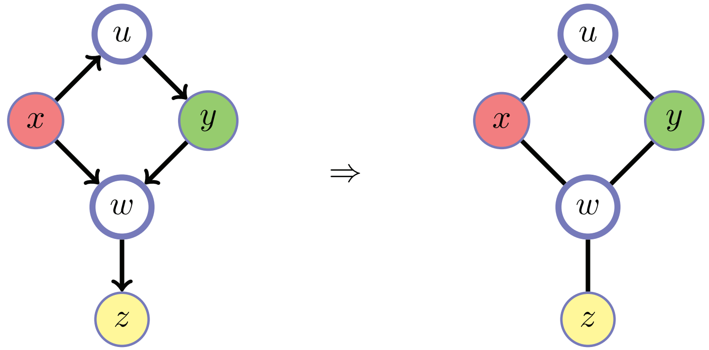
- Earlier we considered the effect on the graph having eliminated a variable (via conditioning or marginalisation)
- Now we consider rules for determining independence based on the graphical representation in which the variables remain in the graph

Graphical path manipulations for independence (cont.)



If z is a collider (bottom path), then we keep undirected links between the neighbours of the collider

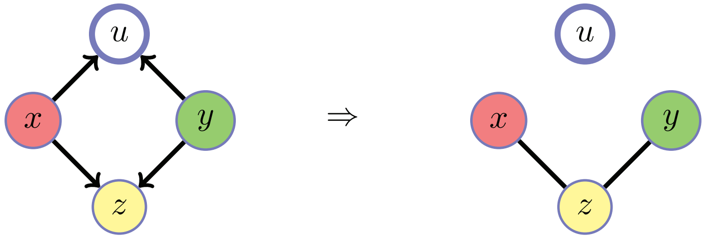
Graphical path manipulations for independence (cont.)



If z is a descendant of a collider, this could induce dependence

- We retain the links, making them undirected

Graphical path manipulations for independence (cont.)

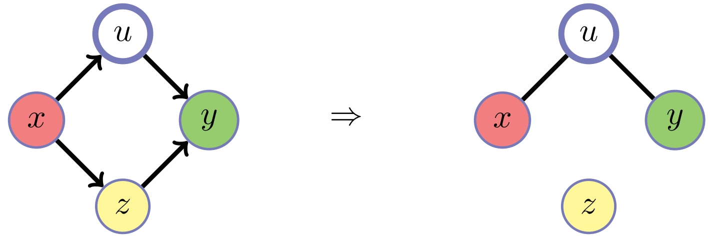


If there is a collider not in the conditioning set (upper path), then we cut the links to the collider variables

- Here, the upper path between x and y is blocked

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical path manipulations for independence (cont.)

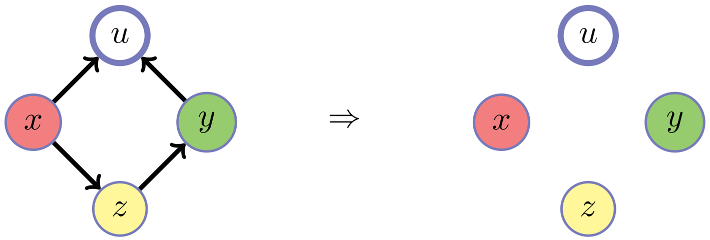


If there is a non-collider which is in the conditioning set (bottom path), then we cut the link between the neighbours of this non-collider which cannot induce dependence between x and y

- The bottom path is blocked

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical path manipulations for independence (cont.)

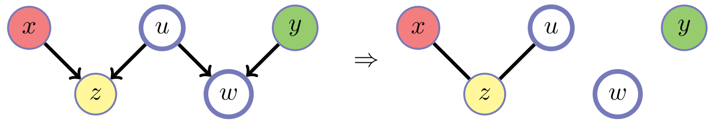


Neither path contributes to dependence, hence $x \perp\!\!\!\perp y | z$

- Both paths are blocked

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical path manipulations for independence (cont.)



Whilst z is a collider in the conditioning set,
w is a collider that is not in the conditioning set

This means that there is no path between x and y,
and hence x and y are independent given z

d-Separation

Belief networks

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

d-separation

The graphical description is intuitive and a formal treatment that is amenable to implementation is straightforward to get from intuitions

- First, we define the DAG concepts of the **d-separation** and **d-connection** that are central to determining conditional independence in any BN with structure given by the DAG

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

d-separation (cont.)

Definition

d-connection and **d-separation**

If G is a directed graph in which $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are disjoint sets of vertices, then $\mathcal{X}$ and $\mathcal{Y}$ are d-connected by $\mathcal{Z}$ in G iff there exists an undirected path U between some vertex in $\mathcal{X}$ and some vertex in $\mathcal{Y}$, and no collider on U is in $\mathcal{Z}$

$\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$ in G iff not d-connected by $\mathcal{Z}$ in G

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

d-separation (cont.)

One may also phrase this as '*For every variable $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, check every path U between x and y and a path U is said to be blocked if there is a node w and U such that either*:

- *w is a collider and neither w nor any of its descendants is in $\mathcal{Z}$*
- *w is not a collider on U and w is in $\mathcal{Z}$*

If all such paths are blocked then $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$, and if the variables sets $\mathcal{X}$ and $\mathcal{Y}$ they are independent conditional on $\mathcal{Z}$ in all probability distributions such a graph can represent

d-separation (cont.)

Remark

Bayes ball

The algorithm provides a linear time complexity algo which given a set of nodes $\mathcal{X}$ and $\mathcal{Z}$ determines the set of nodes $\mathcal{Y}$ such that $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}|\mathcal{Z}$

- $\mathcal{Y}$ is called the set of irrelevant nodes for $\mathcal{X}$ given $\mathcal{Z}$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical and distributional in/dependence

Belief networks

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical and distributional in/dependence

We have shown that $\mathcal{X}$ and $\mathcal{Y}$ d-separated by $\mathcal{Z}$ leads to $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}|\mathcal{Z}$ in all distributions consistent with the belief network structure

If we take any instance of a distribution P which factorises according to the BN structure and then write down a list $\mathcal{L}_p$ of all conditional independence statements that can be obtained from P

1. if $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$, then list $\mathcal{L}_p$ must contain the statement $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}|\mathcal{Z}$
2. List $\mathcal{L}_p$ could contain more statements than those obtained from the graph

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical and distributional in/dependence (cont.)

For the network graph $p(a,b,c) = p(c|a,b)p(a)p(b)$ which is representable by the DAG $a \rightarrow c \leftarrow b$, then $a \perp\!\!\!\perp b$ is the only graphical independence statement we can make

Consider a distribution consistent with $p(a,b,c) = p(c|a,b)p(a)p(b)$

For example, on binary variables $\text{dom}(a) = \text{dom}(b) = \text{dom}(c) = \{0,1\}$

$$p_{[1]}(c = 1|a,b) = (a-b)^2$$
$$p_{[1]}(a = 1) = 0.3$$
$$p_{[1]}(b = 1) = 0.4$$

then numerically we must have $a \perp\!\!\!\perp b$ for this distribution $p_{[1]}$

- $\mathcal{L}_{[1]}$ contains only the statement $a \perp\!\!\!\perp b$

Graphical and distributional in/dependence (cont.)

We can also consider the distribution

$$p_{[2]}(c = 1 | a, b) = 0.5$$
$$p_{[2]}(a = 1) = 0.3$$
$$p_{[2]}(b = 1) = 0.4$$

Here, $\mathcal{L}_{[2]} = \{a \perp\!\!\!\perp b, a \perp\!\!\!\perp c, b \perp\!\!\!\perp c\}$

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical and distributional in/dependence (cont.)

A question is whether or not d-connection similarly implies dependence

- That is, do all distributions P, consistent with the belief network, possess the dependencies implied by the graph?

Consider the BN equation $p(a, b, c) = p(c|a, b)p(a)p(b)$, a and b are d-connected by c, so a and b are dependent, conditioned on c, graphlly

- For instance $p_{[1]}$, numerically $a \top\!\!\!\top b | c$, so the list of dependence statements for $p_{[1]}$ contains the graphical dependence statement
- For instance $p_{[2]}$, list of dependence statements for $p_{[2]}$ is empty

Graphical dependence statements are not necessarily found in all distributions consistent with the belief network

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical and distributional in/dependence (cont.)

$\mathcal{X}$ and $\mathcal{Y}$ d-connected by $\mathcal{Z}$ does NOT lead to $\mathcal{X} \top\!\!\!\top \mathcal{Y} | \mathcal{Z}$ in all distributions consistent with the belief network

Belief networks

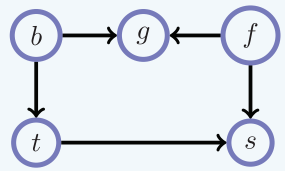
UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Graphical and distributional in/dependence (cont.)

Example

Variables t and f are d-connected by variable g



- Are the variables t and f unconditionally independent ($t \perp\!\!\!\perp f | \emptyset$)?

There are two colliders, namely g and s, however, these are not in the conditioning set (which is empty), hence t and f are d-separated and therefore unconditionally independent

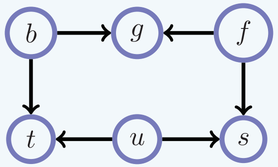
- What about $t \perp\!\!\!\perp f | g$?

There is a path between t and f for which all colliders are in the conditioning set, hence t and f are d-connected by g and thus t and f are graphically dependent conditioned on g

Graphical and distributional in/dependence (cont.)

Example

Variables b and f are d-separated by variable u



- Is $\{b, f\} \perp\!\!\!\perp u | \emptyset$?

Since the conditioning set is empty and every path from either b or f to u contains a collider, b and f are unconditionally independent of u

Belief networks
UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Markov equivalence in BNs

Belief networks

Belief networks
UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Markov equivalence in BNs

We studied how to read conditional independence relations from a DAG

Happily, we can determine whether two DAGs represent the same set of conditional independence statements by using a relatively simple rule

- Even when we don not know what they are!

Definition

Markov equivalence

Two graphs are Markov equivalent if they both represent the same set of conditional independence statements

This definition holds for both directed and undirected graphs

Belief networks
UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Markov equivalence in BNs (cont.)

Example

Consider the belief network with edges $A \rightarrow C \leftarrow B$

- The set of conditional independence statements is $A \perp\!\!\!\perp B | \emptyset$

For the belief network with edges $A \rightarrow C \leftarrow B$ and $A \rightarrow B$

- The set of conditional independence statements is empty

In this case, the two belief networks are not Markov equivalent

Markov equivalence in BNs (cont.)

Pseudocode

Determining Markov equivalence

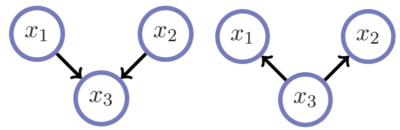
Define an **immorality** in a DAG as a configuration of three nodes A, B and C st C is a child of both A and B, with A and B directly connected

Define the **skeleton** of a graph by removing the directions of the arrow

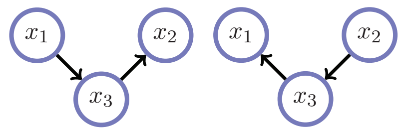
Two DAGS represent the same set of independence assumption (Markov equivalence) iff they share the same skeleton and the same immoralities

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Markov equivalence in BNs (cont.)



(e) (f)



(g) (h)

(b), (c) and (d) are equivalent as they share the same skeleton with no immoralities, (a) has an immorality and it is not equivalent to the others

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Expressibility of BNs

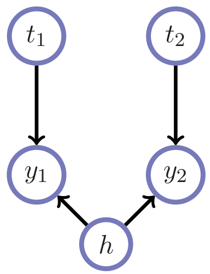
Belief networks

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Expressibility of BNs

BNs fit with our intuitive notion of modelling 'causal' independencies

- Formally they cannot necessarily graphically represent all the independence properties of a given distribution



The DAG can be used to represent two successive experiments where t_1 and t_2 are two treatments and y_1 and y_2 represent two outcomes of interest

- h: Underlying health status of the patient

The first treatment has no effect on the second outcome hence there is no edge from y_1 and y_2

Expressibility of BNs (cont.)

Now consider the implied independencies in the marginal distribution $p(t_1, t_2, y_1, y_2)$, obtained by marginalising the full distribution over h

- There is no DAG containing only the vertices t_1, y_1, t_2, y_2 which represents the independence relations and does not imply some other independence relation that is not implied in the figure

Expressibility of BNs (cont.)

Consequently, any DAG on vertices t_1, y_1, t_2 and y_2 alone will either fail to represent an independence relation of $p(t_1, y_2, t_2, y_2)$, or will impose some additional independence restriction that is not implied by the DAG

In general, $p(t_1, y_1, t_2, y_2) = p(t_1)p(t_2)\sum_h p(y_1|t_1, h)p(y_2|t_2, h)p(h)$ cannot be expressed as a product of functions on a limited set of vars

It is the case, however, that the conditional independence conditions $t_1 \perp\!\!\!\perp (t_2, y_2)$, $t_2 \perp\!\!\!\perp (t_1, y_1)$ hold in $p(t_1, t_2, y_1, y_2)$

- They are there encoded in the form of the CPTs
- We cannot see this independence since it is not present in the structure of the marginalised graph
- Though it can be inferred in a larger graph $p(t_1, t_2, y_1, y_2, h)$

Expressibility of BNs (cont.)

For example, for the BN with link from y_2 to y_1, we have $t_1 \perp\!\!\!\perp t_2 | y_2$

- Not for $p(t_1, y_1, t_2, y_2) = p(t_1)p(t_2)\sum_h p(y_1|t_1, h)p(y_2|t_2, h)p(h)$

Similarly, for the BN with $y_1 \rightarrow y_2$, the implied statement $t_1 \perp\!\!\!\perp t_2 | y_1$ is also not true for that distribution

Expressibility of BNs (cont.)

This example demonstrates that BNs cannot express all the conditional independence statements that could be made on that set of variables

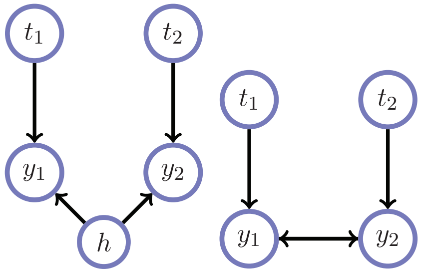
- The set of conditional independence statements can be increased by considering additional variables however

This situation is rather general in the sense that graphical models have limited expressibility in terms of independence statement

Expressibility of BNs (cont.)

It is worth bearing in mind that BNs may not always be the most appropriate framework to express one's independence assumptions

- A natural consideration is to use a bi-directional arrow when a variable is marginalised



One could depict the marginal distribution using a bi-directional edge

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Causality
Belief networks

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Causality

Causality is a contentious topic and some pitfalls can occur

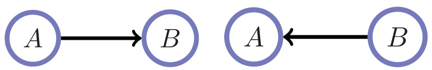
- This may give rise to erroneous inference

The word ''causal' is contentious particularly in cases where the data model contains no explicit temporal implication

- Here, formally only correlations or dependencies can be inferred

A distribution $p(a, b)$ can be written and understood as either

- $p(a|b)p(b)$: We might think that 'b causes a'
- $p(b|a)p(a)$: We might think that 'a causes b'



(a) (b)

Not very meaningful: Both forms represent the same distribution

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Causality (cont.)

Formally BNs only make independence statements not causal ones

- Nevertheless, it can be helpful to think about dependencies in terms of causation, since our intuitive understanding is usually framed in how one variable influences another

Simpson's paradox

Causality

Sympson's paradox

We first discuss a classic conundrum that highlights potential pitfalls

- **Simpson's paradox**: A warning tale in causal reasoning in BNs

Example

Consider a medical trial: Patient treatment and outcome are recovered

Two trials were conducted

- One with 40 females
- One with 40 males

Males	Recovered	Not recovered	Recovery rate
Given drugs	18	12	60%
Not given drugs	7	3	70%

Females	Recovered	Not recovered	Recovery rate
Given drugs	2	8	20%
Not given drugs	9	21	30%

Does the drug cause increased recovery and can be recommended?

Sympson's paradox (cont.)

According to both the male and the female table, the answer is no

- Male: 60% vs 70%
- Female: 20% vs 30%

Ignoring gender information, we find that more people recovered when given the drug than when not and we do not know what to do

Combined	Recovered	Not recovered	Recovery rate
Given drugs	20	20	50%
Not given drugs	16	24	40%

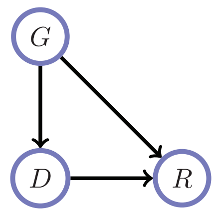
The 'paradox' occurs because we ask a causal (interventional) question

- If we give someone the drug, what happens?

But, we perform an observational calculation and there is a difference between 'given that we see' (observational evidence) and 'given that we do' (interventional evidence)

Sympson's paradox (cont.)

We want to model a causal experiment in which we first intervene, setting the drug state, and observe what effect this has on recovery



A Gender-Drug-Recovery model with no conditional independence assumption is

$$p(G, D, R) = P(R|G, D)p(D|G)p(G) \quad (32)$$

If we intervene and give the drug, term $p(D|G)$ should play no role in the experiment, as we decide to give drug or not independent of gender

- The term $p(D|G)$ therefore needs to be replaced by a term that reflects the set-up of the experiment

Sympson's paradox (cont.)

Atomic intervention: We set a single variable in a particular state

- We set D and we deal with a modified distribution

$$\tilde{p}(G, R|D) = p(R|G, D)p(G) \tag{33}$$

- To denote an intervention, we use the symbol $||$

$$p(R||G, D) \equiv \tilde{p}(R|G, D) = \frac{p(R|G, D)p(G)}{\sum_R p(R|G, D)p(G)} = p(R|G, D) \tag{34}$$

Sympson's paradox (cont.)

Remark

One can also consider here G as being interventional: Irrelevant here

Variable G has no parents, thus for any distribution conditional on G

- the prior factor $p(G)$ will not be present

Using $p(R||G, D) \equiv \tilde{p}(R|G, D) = p(R|G, D)$

For the males given the drug 60% recover, versus 70% recovery when not given the drug

For the females given the drug 20% recover, versus 30% recovery when not given the drug

Belief networks
UFC/DC
AI (CK0031)
2016.2

Sympson's paradox (cont.)

$$p(R||D) \equiv \tilde{p}(R|D) = \frac{\sum_G p(R|G, D)p(G)}{\sum_{R,G} p(R|G, D)p(G)} = \sum_G p(R|G, D)p(G) \tag{35}$$

Using the post intervention distribution above

$$\begin{aligned} p(\text{recovery}|\text{drug}) &= 0.6 \times 0.5 + 0.2 \times 0.5 = 0.4 \\ p(\text{recovery}|\text{no drug}) &= 0.7 \times 0.5 + 0.3 \times 0.5 = 0.5 \end{aligned} \tag{36}$$

We infer that the drug is overall not helpful, as we intuitively expected

- this is consistent with the results from both subpopulations

Sympson's paradox (cont.)

Thus, $p(G, D, R) = p(R|G, D)p(G)p(D)$ means we choose either a Male or Female patient and give drug or not independent of gender

- Hence, the absence of the term $p(D|G)$ from the joint distribution

One way to think about such model is to consider how to draw a sample from the joint distribution of the random variables

- Often this should clarify the role of causality in the experiment

Remark

Observational calculation makes independence assumptions, whereas interventional calculation does not

- This means that term $p(D|G)$ plays a role in the calculation
- It is equivalent to inferring with full distribution in Equation 32

The do-calculus

Causality

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

The do-calculus

In making causal inferences we have seen before that we must adjust the model to reflect any causal experimental conditions

In setting any variable into a state, we need to remove all parental links of that variable

- Pearl calls this the **do operator**, and contrasts observational ('see') inference $p(x|y)$ with causal ('make' or 'do') inference $p(x|\text{do}(y))$

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

The do-calculus (cont.)

Definition

Pearl's DO operator: Let all the variables $\mathcal{X} = \mathcal{X}_C \cup \mathcal{X}_{\overline{C}}$ be written in terms of intervention variables $\mathcal{X}_C$ and non-intervention variables $\mathcal{X}_{\overline{C}}$

For a Bayes BN $p(\mathcal{X}) = \prod_i p(X_i|\text{pa}(X_i))$, inferring the effect of setting variables $X_{c_1}, \dots, X_{c_K}$ (with $c_k \in \mathcal{C}$) in states $\mathtt{x}_{c_1}, \dots, \mathtt{x}_{c_K}$ is equivalent to standard evidential inference in the **post-intervention distribution**

$$p(\mathcal{X}_{\overline{C}}|\text{do}(X_{c_1} = \mathtt{x}_{c_1}), \dots, \text{do}(X_{c_K} = \mathtt{x}_{c_K})) = \prod_{j \in \overline{C}} p(X_j|\text{pa}(X_j)) \quad (37)$$

Any parental var in the intervention set is set to its intervention state

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

The do-calculus (cont.)

For those variables for which we causally intervene and set in a state, the corresponding terms $p(X_{c_i}|\text{pa}(X_{c_i}))$ are removed from the original BN

- The effect is to consider each intervention variable, cut connections to its parents and set intervention variables to its intervention state

For those variables which are evidential but non-causal, the corresponding factors are not removed from the distribution

The do-calculus (cont.)

Interpretation: Post-intervention distributions agree with experiments in which causal variables are first set and non-causal variables are observed

Remark

For a Belief network to have a causal interpretation, it means that the ancestral order of the variables must correspond to the temporal order

- If we start with the variables that have no parents, these must come first in time, with their children coming later
- Ancestral sampling from a causal BN corresponds to the temporal evolution of the physical experiment

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Influence diagrams and do-calculus

Causality

Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Influence diagrams and do-calculus

Influence diagram: A way to modify a BN to represent intervention

- Append a parental decision variable F_X to any variable X on which an intervention can be made

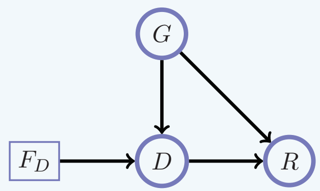
Belief networks
UFC/DC
AI (CK0031)
2016.2
Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Influence diagrams and do-calculus (cont.)

Example

For the Simpson's paradox example, we may use

$$\tilde{p}(D, G, R, F_D) = p(D|F_D, G)p(G)p(R|G, D)p(F_D) \tag{38}$$



$$p(D|F_D = \emptyset, G) \equiv p(D|\text{pa}(D))$$

$$p(D|F_D = \mathtt{d}, G) = \begin{cases} 1 & \text{if } D = \mathtt{d} \\ 0 & \text{otherwise} \end{cases}$$

- If the decision variable F_D is set to the empty state, the variable D is determined by the standard observational term $p(D|\text{pa}(D))$
- If the decision variable F_D is set to a state of D, the variable puts all its probability in that single state of $D = \mathtt{d}$

Influence diagrams and do-calculus (cont.)

This has the effect of replacing the conditional probability term by a unit factor and any instances of D set to the variable in its interventional state (or, distribution of states, in some cases)

- A potential advantage of influence diagrams over do-calculus is that conditional independence statements can be derived using standard techniques for the augmented BN
- Additionally, for learning, standard techniques apply in which decision variables are set to the condition under which each data sample was collected (a causal or non-causal sample)

Belief networks

UFC/DC
AI (CK0031)
2016.2

Benefits of structure
Modelling independencies
Reducing specifications
Uncertain and unreliable evidence
Uncertain evidence
Unreliable evidence
Belief networks
Conditional independence
The impact of collisions
Path manipulations
d-Separation
Graphical and distributional in/dependence
Markov equivalence in BNs
Expressibility of BNs
Causality
Simpson's paradox
The do-calculus
Influence diagrams and do-calculus

Influence diagrams and do-calculus (cont.)

Remark

Learning the edge directions

In the absence of data from causal experiments, one should be justifiably sceptical abut learning 'causal' networks, and might prefer a certain direction of a link based on assumptions of the 'simplicity' of the CPTs

- The preference may come from physical intuition that, whilst root causes may be uncertain, the relation from cause to effect is clear
- A measure of complexity of a CPT is required, such as entropy

Such heuristics can be numerically encoded and the edge directions learned in an otherwise Markov equivalent graph