Reasoning with Bayesian Networks
Lecture 1: Probability Calculus, Bayesian Networks

Jinbo Huang

NICTA and ANU
Overview of the Course

- Probability calculus, Bayesian networks
- Inference by variable elimination, factor elimination, conditioning
- Models for graph decomposition
- Most likely instantiations
- Complexity of probabilistic inference
- Compiling Bayesian networks
- Inference with local structure
- Applications
### Probabilities of Worlds

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>Earthquake</th>
<th>Burglary</th>
<th>Alarm</th>
<th>$\Pr(.)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>.0190</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>.0010</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>.0560</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>.0240</td>
</tr>
<tr>
<td>$\omega_5$</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>.1620</td>
</tr>
<tr>
<td>$\omega_6$</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>.0180</td>
</tr>
<tr>
<td>$\omega_7$</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>.0072</td>
</tr>
<tr>
<td>$\omega_8$</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>.7128</td>
</tr>
</tbody>
</table>

**Worlds modeled with propositional variables**

**A world is complete instantiation of variables**

**Probabilities of all worlds add up to 1**

**An event is a set of worlds**
Probabilities of Events

<table>
<thead>
<tr>
<th>world</th>
<th>Earthquake</th>
<th>Burglary</th>
<th>Alarm</th>
<th>Pr(.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ω₁</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>.0190</td>
</tr>
<tr>
<td>ω₂</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>.0010</td>
</tr>
<tr>
<td>ω₃</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>.0560</td>
</tr>
<tr>
<td>ω₄</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>.0240</td>
</tr>
<tr>
<td>ω₅</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>.1620</td>
</tr>
<tr>
<td>ω₆</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>.0180</td>
</tr>
<tr>
<td>ω₇</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>.0072</td>
</tr>
<tr>
<td>ω₈</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>.7128</td>
</tr>
</tbody>
</table>

\[ \Pr(\alpha) \overset{\text{def}}{=} \sum_{\omega \in \alpha} \Pr(\omega) \]

\[ \Pr(\text{Earthquake}) = \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \]
Probabilities of Events

<table>
<thead>
<tr>
<th>world</th>
<th>Earthquake</th>
<th>Burglary</th>
<th>Alarm</th>
<th>Pr(.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>.0190</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>.0010</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>.0560</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>.0240</td>
</tr>
<tr>
<td>$\omega_5$</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>.1620</td>
</tr>
<tr>
<td>$\omega_6$</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>.0180</td>
</tr>
<tr>
<td>$\omega_7$</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>.0072</td>
</tr>
<tr>
<td>$\omega_8$</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>.7128</td>
</tr>
</tbody>
</table>

\[
\Pr(\alpha) \overset{\text{def}}{=} \sum_{\omega \in \alpha} \Pr(\omega)
\]

- $\Pr(\text{Earthquake}) = \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1$
- $\Pr(\neg \text{Burglary}) = .8$
- $\Pr(\text{Alarm}) = .2442$
### Propositional Sentences as Events

<table>
<thead>
<tr>
<th>world</th>
<th>Earthquake</th>
<th>Burglary</th>
<th>Alarm</th>
<th>Pr(.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>.0190</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>.0010</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>.0560</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>.0240</td>
</tr>
<tr>
<td>$\omega_5$</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>.1620</td>
</tr>
<tr>
<td>$\omega_6$</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>.0180</td>
</tr>
<tr>
<td>$\omega_7$</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>.0072</td>
</tr>
<tr>
<td>$\omega_8$</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>.7128</td>
</tr>
</tbody>
</table>

Pr($((\text{Earthquake} \lor \text{Burglary}) \land \neg \text{Alarm}) = ?$)
Propositional Sentences as Events

<table>
<thead>
<tr>
<th>world</th>
<th>Earthquake</th>
<th>Burglary</th>
<th>Alarm</th>
<th>Pr(.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>.0190</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>.0010</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>.0560</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>.0240</td>
</tr>
<tr>
<td>$\omega_5$</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>.1620</td>
</tr>
<tr>
<td>$\omega_6$</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>.0180</td>
</tr>
<tr>
<td>$\omega_7$</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>.0072</td>
</tr>
<tr>
<td>$\omega_8$</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>.7128</td>
</tr>
</tbody>
</table>

$\Pr((\text{Earthquake} \lor \text{Burglary}) \land \neg \text{Alarm}) =$?

Interpret world $\omega$ as model (satisfying assignment) for propositional sentence $\alpha$

$$\Pr(\alpha) \overset{\text{def}}{=} \sum_{\omega \models \alpha} \Pr(\omega)$$

Jinbo Huang  Reasoning with Bayesian Networks
Suppose we’re given evidence $\beta$ that $\text{Alarm} = \text{true}$.

Need to update $\Pr(.)$ to $\Pr(.|\beta)$.
Updating Beliefs

\[ \Pr(\beta | \beta) = 1, \Pr(\neg \beta | \beta) = 0 \]

Partition worlds into those \( \models \beta \) and those \( \models \neg \beta \)

\[ \Pr(\omega | \beta) = 0, \text{ for all } \omega \models \neg \beta \]
Pr(β|β) = 1, Pr(¬β|β) = 0

Partition worlds into those |= β and those |= ¬β

- Pr(ω|β) = 0, for all ω |= ¬β
- Relative beliefs stay put, for ω |= β
Updating Beliefs

\[ \Pr(\beta|\beta) = 1, \Pr(\neg\beta|\beta) = 0 \]

Partition worlds into those \( \models \beta \) and those \( \models \neg\beta \)

- \( \Pr(\omega|\beta) = 0, \) for all \( \omega \models \neg\beta \)
- Relative beliefs stay put, for \( \omega \models \beta \)
- \( \sum_{\omega \models \beta} \Pr(\omega) = \Pr(\beta) \)
Updating Beliefs

Pr(\(\beta|\beta\)) = 1, Pr(\(\neg\beta|\beta\)) = 0

Partition worlds into those \(\models \beta\) and those \(\models \neg\beta\)

- \(\Pr(\omega|\beta) = 0\), for all \(\omega \models \neg\beta\)
- Relative beliefs stay put, for \(\omega \models \beta\)
- \(\sum_{\omega \models \beta} \Pr(\omega) = \Pr(\beta)\)
- \(\sum_{\omega \models \beta} \Pr(\omega|\beta) = 1\)
Updating Beliefs

\[ \Pr(\beta|\beta) = 1, \Pr(\neg\beta|\beta) = 0 \]

Partition worlds into those \( \models \beta \) and those \( \models \neg\beta \)

- \( \Pr(\omega|\beta) = 0 \), for all \( \omega \models \neg\beta \)
- Relative beliefs stay put, for \( \omega \models \beta \)
- \( \sum_{\omega \models \beta} \Pr(\omega) = \Pr(\beta) \)
- \( \sum_{\omega \models \beta} \Pr(\omega|\beta) = 1 \)
- Above imply we normalize by constant \( \Pr(\beta) \)
Updating Beliefs

\[ \Pr(\beta | \beta) = 1, \Pr(\neg \beta | \beta) = 0 \]

Partition worlds into those \( \models \beta \) and those \( \models \neg \beta \)

- \( \Pr(\omega | \beta) = 0 \), for all \( \omega \models \neg \beta \)
- Relative beliefs stay put, for \( \omega \models \beta \)
- \( \sum_{\omega \models \beta} \Pr(\omega) = \Pr(\beta) \)
- \( \sum_{\omega \models \beta} \Pr(\omega | \beta) = 1 \)
- Above imply we normalize by constant \( \Pr(\beta) \)

\[ \Pr(\omega | \beta) \overset{\text{def}}{=} \begin{cases} 0, & \text{if } \omega \models \neg \beta \\ \Pr(\omega) / \Pr(\beta), & \text{if } \omega \models \beta \end{cases} \]
## Updating Beliefs

|  world  | Earthquake | Burglary | Alarm   | Pr(.)  | Pr(.|Alarm)  |
|---------|------------|----------|---------|--------|-------------|
| $\omega_1$ | true       | true     | true    | .0190  | .0190/.2442 |
| $\omega_2$ | true       | true     | false   | .0010  | 0           |
| $\omega_3$ | true       | false    | true    | .0560  | .0560/.2442 |
| $\omega_4$ | true       | false    | false   | .0240  | 0           |
| $\omega_5$ | false      | true     | true    | .1620  | .1620/.2442 |
| $\omega_6$ | false      | true     | false   | .0180  | 0           |
| $\omega_7$ | false      | false    | true    | .0072  | .0072/.2442 |
| $\omega_8$ | false      | false    | false   | .7128  | 0           |
Bayes Conditioning

\[
\Pr(\alpha | \beta) = \sum_{\omega \models \alpha} \Pr(\omega | \beta)
\]

\[
= \sum_{\omega \models \alpha, \omega \models \beta} \Pr(\omega | \beta) + \sum_{\omega \models \alpha, \omega \models \neg \beta} \Pr(\omega | \beta)
\]

\[
= \sum_{\omega \models \alpha, \omega \models \beta} \Pr(\omega | \beta) = \sum_{\omega \models \alpha \land \beta} \Pr(\omega | \beta)
\]

\[
= \sum_{\omega \models \alpha \land \beta} \Pr(\omega) / \Pr(\beta) = \frac{1}{\Pr(\beta)} \sum_{\omega \models \alpha \land \beta} \Pr(\omega)
\]

\[
= \frac{\Pr(\alpha \land \beta)}{\Pr(\beta)}
\]
Bayes Conditioning

\[ \Pr(Burglary) = 0.2 \]
Bayes Conditioning

\[
\begin{align*}
\Pr(Burglary) &= .2 \\
\Pr(Burglary|Alarm) &\approx .741 \uparrow
\end{align*}
\]
Bayes Conditioning

\[
\begin{align*}
\Pr(\text{Burglary}) &= .2 \\
\Pr(\text{Burglary} | \text{Alarm}) &\approx .741 \uparrow \\
\Pr(\text{Burglary} | \text{Alarm} \land \text{Earthquake}) &\approx .253 \downarrow
\end{align*}
\]
Bayes Conditioning

\[
\begin{align*}
\Pr(Burglary) &= .2 \\
\Pr(Burglary | Alarm) &\approx .741 \uparrow \\
\Pr(Burglary | Alarm \land Earthquake) &\approx .253 \downarrow \\
\Pr(Burglary | Alarm \land \neg Earthquake) &\approx .957 \uparrow
\end{align*}
\]
Independence

\[
\begin{align*}
\Pr(\text{Earthquake}) & = .1 \\
\Pr(\text{Earthquake} \mid \text{Burglary}) & = .1
\end{align*}
\]
Independence

\[
\begin{align*}
\Pr(Earthquake) &= .1 \\
\Pr(Earthquake|Burglary) &= .1
\end{align*}
\]

Event \( \alpha \) independent of event \( \beta \) if

\[
\Pr(\alpha|\beta) = \Pr(\alpha) \text{ or } \Pr(\beta) = 0
\]
Independence

\[
\begin{align*}
\Pr(\text{Earthquake}) & = .1 \\
\Pr(\text{Earthquake} | \text{Burglary}) & = .1
\end{align*}
\]

Event \( \alpha \) independent of event \( \beta \) if

\[
\Pr(\alpha | \beta) = \Pr(\alpha) \text{ or } \Pr(\beta) = 0
\]

Or equivalently

\[
\Pr(\alpha \land \beta) = \Pr(\alpha) \Pr(\beta)
\]
Independence

\[ \Pr(\text{Earthquake}) = .1 \]
\[ \Pr(\text{Earthquake}|\text{Burglary}) = .1 \]

Event \( \alpha \) independent of event \( \beta \) if

\[ \Pr(\alpha|\beta) = \Pr(\alpha) \text{ or } \Pr(\beta) = 0 \]

Or equivalently

\[ \Pr(\alpha \land \beta) = \Pr(\alpha) \Pr(\beta) \]

Independence is symmetric
Conditional Independence

Independence is a dynamic notion

Burglary independent of Earthquake, but not after accepting evidence Alarm

\[
\Pr(Burglary | Alarm) \approx 0.741
\]
\[
\Pr(Burglary | Alarm \land Earthquake) \approx 0.253
\]
Conditional Independence

Independence is a dynamic notion

Burglary independent of Earthquake, but not after accepting evidence Alarm

\[
\begin{align*}
\Pr(Burglary|Alarm) & \approx .741 \\
\Pr(Burglary|Alarm \land Earthquake) & \approx .253
\end{align*}
\]

\(\alpha\) independent of \(\beta\) given \(\gamma\) if

\[
\Pr(\alpha \land \beta|\gamma) = \Pr(\alpha|\gamma) \Pr(\beta|\gamma) \text{ or } \Pr(\gamma) = 0
\]
Variable Independence

For disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$

$I_{Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) : \mathbf{X}$ independent of $\mathbf{Y}$ given $\mathbf{Z}$ in $Pr$

Means $x$ independent of $y$ given $z$ for all instantiations of $x$, $y$, $z$
Properties of Beliefs: Chain Rule

\[
\Pr(\alpha_1 \land \alpha_2 \land \ldots \land \alpha_n)
\]

\[
= \Pr(\alpha_1 | \alpha_2 \land \ldots \land \alpha_n) \Pr(\alpha_2 | \alpha_3 \land \ldots \land \alpha_n) \ldots \Pr(\alpha_n)
\]
Properties of Beliefs: Case Analysis

\[ \Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha \land \beta_i) \]

Or equivalently

\[ \Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha|\beta_i) \Pr(\beta_i) \]

Where events \( \beta_i \) are mutually exclusive and collectively exhaustive.
Properties of Beliefs: Case Analysis

Special case of $n = 2$

$$\Pr(\alpha) = \Pr(\alpha \land \beta) + \Pr(\alpha \land \lnot \beta)$$

Or equivalently

$$\Pr(\alpha) = \Pr(\alpha|\beta) \Pr(\beta) + \Pr(\alpha|\lnot \beta) \Pr(\lnot \beta)$$
Properties of Beliefs: Bayes Rule

\[
Pr(\alpha|\beta) = \frac{Pr(\beta|\alpha) \cdot Pr(\alpha)}{Pr(\beta)}
\]
Example: patient tested positive for disease

- $D$: patient has disease
- $T$: test comes out positive
- Would like to know $\Pr(D|T)$
Example: patient tested positive for disease

- $D$: patient has disease
- $T$: test comes out positive
- Would like to know $\Pr(D|T)$

We know

- $\Pr(D) = 1/1000$
- $\Pr(T|\neg D) = 2/100$ (false positive)
- $\Pr(\neg T|D) = 5/100$ (false negative)
Properties of Beliefs: Bayes Rule

Example: patient tested positive for disease

- $D$: patient has disease
- $T$: test comes out positive
- Would like to know $\Pr(D | T)$

We know

- $\Pr(D) = 1/1000$
- $\Pr(T | \neg D) = 2/100$ (false positive)
- $\Pr(\neg T | D) = 5/100$ (false negative)

Using Bayes rule: $\Pr(D | T) = \frac{\Pr(T | D) \Pr(D)}{\Pr(T)}$
Computing $\Pr(T)$: case analysis

$$\Pr(T) = \Pr(T|D) \Pr(D) + \Pr(T|\neg D) \Pr(\neg D)$$

$$= \frac{95}{100} \times \frac{1}{1000} + \frac{2}{200} \times \frac{999}{1000} = \frac{2093}{100000}$$
Computing $\Pr(T)$: case analysis

$$\Pr(T) = \Pr(T|D) \Pr(D) + \Pr(T|\neg D) \Pr(\neg D)$$

$$= \frac{95}{100} \times \frac{1}{1000} + \frac{2}{200} \times \frac{999}{1000} = \frac{2093}{100000}$$

$$\Pr(D|T) = \frac{95}{2093} \approx 4.5\%$$
So far we’ve considered *hard evidence*, that some event has occurred

*Soft evidence*: neighbor with hearing problem calls to tell us they heard alarm

May not confirm Alarm, but can still increase our belief in Alarm

Two main methods to specify strength of soft evidence
"After receiving my neighbor’s call, my belief in Alarm is now .85”

Soft evidence as constraint $\Pr'(\beta) = q$

Not a statement about strength of evidence *per se*, but result of its integration with our initial beliefs
Soft Evidence: All Things Considered

“After receiving my neighbor’s call, my belief in Alarm is now .85”

Soft evidence as constraint $\Pr'(\beta) = q$

Not a statement about strength of evidence *per se*, but result of its integration with our initial beliefs

Scale beliefs in worlds $\models \beta$ so they add up to $q$
Scale beliefs in worlds $\models \neg \beta$ so they add up to $1 - q$
Soft Evidence: All Things Considered

For worlds

$$
\Pr'(\omega) \overset{\text{def}}{=} \begin{cases} 
\frac{q}{\Pr(\beta)} \Pr(\omega), & \text{if } \omega \models \beta \\
\frac{1-q}{\Pr(\neg \beta)} \Pr(\omega), & \text{if } \omega \models \neg \beta 
\end{cases}
$$
Soft Evidence: All Things Considered

For worlds

\[ \Pr'(\omega) \overset{\text{def}}{=} \begin{cases} q \frac{\Pr(\omega)}{\Pr(\beta)}, & \text{if } \omega \models \beta \\ \frac{1-q}{\Pr(\neg\beta)} \Pr(\omega), & \text{if } \omega \models \neg\beta \end{cases} \]

For events (Jeffery’s rule)

\[ \Pr'(\alpha) = q \Pr(\alpha|\beta) + (1 - q) \Pr(\alpha|\neg\beta) \]
Soft Evidence: All Things Considered

For worlds

\[
Pr'(\omega) \overset{\text{def}}{=} \begin{cases} 
q \frac{Pr(\omega)}{Pr(\beta)}, & \text{if } \omega \models \beta \\
1 - q \frac{Pr(\omega)}{Pr(\neg \beta)}, & \text{if } \omega \models \neg \beta
\end{cases}
\]

For events (Jeffery’s rule)

\[
Pr'(\alpha) = q Pr(\alpha|\beta) + (1 - q) Pr(\alpha|\neg \beta)
\]

Bayes conditioning is special case of \( q = 1 \)
Soft Evidence: Nothing Else Considered

Would like to declare strength of evidence independently of current beliefs
Would like to declare strength of evidence independently of current beliefs

Need notion of *odds* of event $\beta$: $O(\beta) \overset{\text{def}}{=} \frac{\Pr(\beta)}{\Pr(\neg\beta)}$
Soft Evidence: Nothing Else Considered

Would like to declare strength of evidence independently of current beliefs

Need notion of odds of event $\beta$: $O(\beta) \overset{\text{def}}{=} \frac{\Pr(\beta)}{\Pr(\neg\beta)}$

Declare evidence $\beta$ by Bayes factor $k = \frac{O'(\beta)}{O(\beta)}$
Soft Evidence: Nothing Else Considered

Would like to declare strength of evidence independently of current beliefs.

Need notion of *odds* of event $\beta$: $O(\beta) \overset{\text{def}}{=} \frac{\Pr(\beta)}{\Pr(\neg\beta)}$

Declare evidence $\beta$ by *Bayes factor* $k = \frac{O'(\beta)}{O(\beta)}$

“My neighbor’s call increases odds of Alarm by factor of 4”
Soft Evidence: Nothing Else Considered

Would like to declare strength of evidence independently of current beliefs.

Need notion of *odds* of event \( \beta \): \( O(\beta) \overset{\text{def}}{=} \frac{\Pr(\beta)}{\Pr(\neg \beta)} \)

Declare evidence \( \beta \) by *Bayes factor* \( k = \frac{O'(\beta)}{O(\beta)} \)

“My neighbor’s call increases odds of Alarm by factor of 4”

Compute \( \Pr'(\beta) \) and fall back on Jeffrey’s rule to update beliefs.
Probabilistic Reasoning

Basic task: belief updating in face of hard and soft evidence

Can be done in principle using definitions we’ve seen

Inefficient, requires joint probability tables (exponential)

Also a modeling difficulty

- Specifying joint probabilities is tedious
- Beliefs (e.g., independencies) held by human experts not easily enforced
Capturing Independence Graphically

Evidence $R$ could increase $\Pr(A)$, $\Pr(C)$, but not if we know $\neg A$. $C$ independent of $R$ given $\neg A$. 

Jinbo Huang
Reasoning with Bayesian Networks
Evidence $R$ could increase $\Pr(A)$, $\Pr(C)$
Evidence $R$ could increase $\Pr(A)$, $\Pr(C)$
But not if we know $\neg A$
Evidence $R$ could increase $\Pr(A)$, $\Pr(C)$

But not if we know $\neg A$

$C$ independent of $R$ given $\neg A$
Capturing Independence Graphically

Parents(V): variables with edge to V

Descendants(V): variables to which \( \exists \) directed path from V

Nondescendants(V): all except V, Parents(V), and Descendants(V)
Markovian Assumptions

Every variable is conditionally independent of its nondescendants given its parents

\[ I(V, \text{Parents}(V), \text{Nondescendants}(V)) \]
Markovian Assumptions

Every variable is conditionally independent of its nondescendants given its parents

\[ I(V, \text{Parents}(V), \text{Nondescendants}(V)) \]

Given direct causes of variable, our beliefs in that variable are no longer influenced by any other variable except possibly by its effects
Capturing Independence Graphically

\[ I(C, A, \{B, E, R\}) \]
\[ I(R, E, \{A, B, C\}) \]
\[ I(A, \{B, E\}, R) \]
\[ I(B, \emptyset, \{E, R\}) \]
\[ I(E, \emptyset, B) \]
DAG $G$ together with $\text{Markov}(G)$ declares a set of independencies, but does not define $\text{Pr}$.

$\text{Pr}$ is uniquely defined if a conditional probability table ($CPT$) is provided for each variable $X$.

$CPT$: $\text{Pr}(x|u)$ for every value $x$ of variable $X$ and every instantiation $u$ of parents $U$. 


Conditional Probability Tables

- $\Pr(c|a)$
- $\Pr(r|e)$
- $\Pr(a|b, e)$
- $\Pr(e)$
- $\Pr(b)$

Diagram:

- Earthquake? ($E$)
- Burglary? ($B$)
- Radio? ($R$)
- Alarm? ($A$)
- Call? ($C$)
Conditional Probability Tables

- Earthquake? (E)
- Burglary? (B)
- Radio? (R)
- Alarm? (A)
- Call? (C)

Conditional Probability Table:

| A   | C   | Pr(c|a) |
|-----|-----|--------|
| true| true| .80    |
| true| false| .20  |
| false| true| .001  |
| false| false| .999  |
Conditional Probability Tables

Half of probabilities are redundant

Only need 10 probabilities for whole DAG

| $A$     | $C$    | $\Pr(c|a)$ |
|---------|--------|------------|
| true    | true   | .80        |
| true    | false  | .20        |
| false   | true   | .001       |
| false   | false  | .999       |
A pair \((G, \Theta)\) where

- \(G\) is DAG over variables \(Z\), called network *structure*
- \(\Theta\) is set of CPTs, one for each variable in \(Z\), called network *parametrization*
Bayesian Network

\[ \Theta_{X|U}: \text{CPT for } X \text{ and parents } U \]

\[ XU: \text{network family} \]

\[ \theta_{x|u} = \Pr(x|u): \text{network parameter} \]

\[ \sum_{x} \theta_{x|u} = 1 \text{ for every parent instantiation } u \]
Network instantiation: e.g., $a, b, \overline{c}, d, \overline{e}$

$\theta_{x|u} \sim z$: instantiations $xu$ & $z$ are compatible

$\theta_{a}, \theta_{b|a}, \theta_{\overline{c}|a}, \theta_{d|b, \overline{c}}, \theta_{\overline{e}|\overline{c}}$ all $\sim z$ above
Bayesian Network

Independence constraints imposed by network structure (DAG) and numeric constraints imposed by network parametrization (CPTs) are satisfied by \textit{unique} \( \Pr \)

\[
\Pr(z) \overset{\text{def}}{=} \prod_{\theta_{x|u} \sim z} \theta_{x|u} \quad (\text{chain rule})
\]

Bayesian network is implicit representation of this probability distribution
Chain Rule Example

\[
\Pr(z) \overset{\text{def}}{=} \prod_{\theta_{x|u} \sim z} \theta_{x|u} \quad \text{(chain rule)}
\]

\[
\Pr(a, b, \overline{c}, d, \overline{e}) = \theta_a \theta_{b|a} \theta_{\overline{c}|a} \theta_{d|b, \overline{c}} \theta_{\overline{e}|\overline{c}}
\]

\[
= (.6)(.2)(.2)(.9)(.1)
\]

\[
= .0216
\]
Compactness of Bayesian Networks

Let $d$ be max variable domain size, $k$ max number of parents

Size of any CPT = $O(d^{k+1})$

Total number of network parameters = $O(n \cdot d^{k+1})$ for $n$-variable network

Compact as long as $k$ is relatively small, compared with joint probability table (up to $d^n$ entries)
Properties of Probabilistic Independence

Distribution \( \Pr \) represented by Bayesian network \((G, \Theta)\) is guaranteed to satisfy every independence assumption in \( \text{Markov}(G) \)

\[ I_{\Pr}(X, \text{Parents}(X), \text{Nondescendants}(X)) \]

It may satisfy more

Derive independence by \textit{graphoid axioms}
Graphoid Axioms: Symmetry

\[ I_{Pr}(X, Z, Y) \leftrightarrow I_{Pr}(Y, Z, X) \]

If learning \( y \) does not influence our belief in \( x \), then learning \( x \) does not influence our belief in \( y \) either.
Graphoid Axioms: Decomposition

\[ I_{Pr}(X, Z, Y \cup W) \rightarrow I_{Pr}(X, Z, Y) \land I_{Pr}(X, Z, W) \]

If learning \(yw\) does not influence our belief in \(x\), then learning \(\text{y}\) alone, or \(w\) alone, does not influence our belief in \(x\) either.

If some information is irrelevant, then any part of it is also irrelevant.

Reverse does not hold in general: Two pieces of information may each be irrelevant on their own, yet their combination may be relevant.
Graphoid Axioms: Decomposition

\[ I_{Pr}(X, Z, Y \cup W) \rightarrow I_{Pr}(X, Z, Y) \land I_{Pr}(X, Z, W) \]

In particular, can strengthen \textit{Markov}(G)

\[ I_{Pr}(X, \text{Parents}(X), W) \text{ for } W \subseteq \text{Nondescendants}(X) \]

Useful for proving chain rule for Bayesian networks from chain rule of probability calculus.
Graphoid Axioms: Weak Union

\[ I_{Pr}(X, Z, Y \cup W) \rightarrow I_{Pr}(X, Z \cup Y, W) \]

If information \(yw\) is not relevant to our belief in \(x\), then partial information \(y\) will not make rest of information, \(w\), relevant.

In particular, can strengthen Markov(\(G\))

\[ I_{Pr}(X, Parents(X) \cup W, Nondescendants(X)\setminus W) \]

for \(W \subseteq Nondescendants(X)\)
Graphoid Axioms: Contraction

\[ I_{Pr}(X, Z, Y) \land I_{Pr}(X, Z \cup Y, W) \rightarrow I_{Pr}(X, Z, Y \cup W) \]

If, after learning irrelevant information \( y \), information \( w \) is found to be irrelevant to our belief in \( x \), then combined information \( yw \) must have been irrelevant to start with.
Intersection Axiom

\[ I_{Pr}(X, Z \cup W, Y) \land I_{Pr}(X, Z \cup Y, W) \rightarrow I_{Pr}(X, Z, Y \cup W) \]

for strictly positive distribution \( Pr \)

If information \( w \) is irrelevant given \( y \), and information \( y \) is irrelevant given \( w \), then their combination is irrelevant to start with.

Not true in general, when there’re zero probabilities (determinism)
Graphical Test of Independence

Graphoid axioms produce independencies beyond Markov($G$)

Derivation of independencies is nontrivial; need efficient way

Linear-time graphical test: d-separation
d-separation

\[ dsep_G(X, Z, Y) : \text{every path between } X \text{ & } Y \text{ is blocked by } Z \]

\[ dsep_G(X, Z, Y) \rightarrow I_{Pr}(X, Z, Y) \text{ for every Pr induced by } G \] (soundness)

Also enjoys \textit{weak completeness} (discussed later)
d-separation: Paths and Valves

- Sequential
- Divergent
- Convergent
d-separation: Paths and Valves

- **Sequential Valve**
  - Earthquake? (E)
  - Burglary? (B)
  - Radio? (R)
  - Alarm? (A)
  - Call? (C)

- **Divergent Valve**
  - Earthquake? (E)
  - Burglary? (B)
  - Radio? (R)
  - Alarm? (A)
  - Call? (C)

- **Convergent Valve**
  - Earthquake? (E)
  - Burglary? (B)
  - Radio? (R)
  - Alarm? (A)
  - Call? (C)
d-separation: Path Blocking

Path is *blocked* if any valve is *closed*

Sequential valve $\rightarrow W \rightarrow$ closed iff $W \in Z$

Divergent valve $\leftarrow W \rightarrow$ closed iff $W \in Z$

Convergent value $\rightarrow W \leftarrow$ closed iff $W \not\in Z$ and none of $\text{Descendants}(W)$ appears in $Z$
d-separation: Examples

- Earthquake? (E)
  - Radio? (R)
  - Call? (C)
- Burglary? (B)
  - Alarm? (A)

- Earthquake? (E)
  - Radio? (R)
  - Call? (C)
- Burglary? (B)
  - Alarm? (A)
d-separation: Examples

- Visit to Asia? ($A$)
- Smoker? ($S$)
- Tuberculosis? ($T$)
- Lung Cancer? ($C$)
- Bronchitis? ($B$)
- Tuberculosis or Cancer? ($P$)
- Positive X-Ray? ($X$)
- Dyspnoea? ($D$)

Diagram showing relationships and d-separations among variables.
Need not enumerate paths between $\mathbf{X}$ & $\mathbf{Y}$

\[ dsep_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \text{ iff } \mathbf{X} \text{ and } \mathbf{Y} \text{ are disconnected in new DAG } G' \]

obtained by pruning DAG $G$ as follows

▶ Delete all leaves $W \notin \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$
▶ Delete all edges outgoing from $\mathbf{Z}$

Linear in size of $G$
d-separation: Soundness

\[ dsep_G(X, Z, Y) \rightarrow I_{Pr}(X, Z, Y) \text{ for } Pr \text{ induced by } G \]

Prove by showing every independence claimed by d-separation can be derived from graphoid axioms
d-separation: Completeness?

\[ dsep_G(X, Z, Y) \leftarrow I_{Pr}(X, Z, Y)? \]

Does not hold

- Consider \( X \rightarrow Y \rightarrow Z \)
- \( X \) not d-separated from \( Z \) (given \( \emptyset \))
- However, if \( \theta_y|_x = \theta_y|_{\overline{x}} \) then \( X \) & \( Y \) are independent, so are \( X \) & \( Z \)

Not surprising as d-separation is based on structure alone
For every DAG $G$, $\exists$ parametrization $\Theta$ such that

$$dsep_G(X, Z, Y) \leftrightarrow I_{\Pr}(X, Z, Y)$$

where $\Pr$ is induced by Bayesian network $(G, \Theta)$

Implies that one cannot improve on d-separation
Summary

Probabilities of worlds & events, Bayes conditioning, independence, chain rule, case analysis, Bayes rule, two methods for soft evidence

Capturing independence graphically, Markovian assumptions, conditional probabilities tables, Bayesian networks, chain rule, graphoid axioms, d-separation