Today

- Last time:
  - Bayes nets
  - Conditional independence

- Today:
  - More conditional independence
  - Inference to answer queries
Reachability (the Bayes’ Ball)

- **Correct algorithm:**
  - Start at source node
  - Try to reach target by search
- **States:** node, along with previous arc
- **Successor function:**
  - **Unobserved nodes:**
    - To any child
    - To any parent if coming from a child (or start)
  - **Observed nodes:**
    - From parent to parent
- If you can’t reach a node, it’s conditionally independent of the start node

Example

\[
\begin{align*}
L & \perp T'|T \\
L & \perp B \\
L & \perp B|T \\
L & \perp B|T'
\end{align*}
\]

Yes

Yes
Example

- **Variables:**
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I’m sad

- **Questions:**
  - $T \perp D$
  - $T \perp D | R$  
    
    Yes
  - $T \perp D | R, S$

Summary

- Bayes nets compactly encode joint distributions

- Guaranteed independencies of distributions can be deduced from BN graph structure

- The Bayes’ ball algorithm (aka d-separation)

- A Bayes net may have other independencies that are not detectable until you inspect its specific distribution
Inference

- Inference: calculating some statistic from a joint probability distribution
- Examples:
  - Posterior marginal probability:
    \[ P(Q|E_1 = e_1, \ldots, E_k = e_k) \]
  - Most likely explanation:
    \[ \text{argmax}_q P(Q = q|E_1 = e_1 \ldots) \]

Reminder: Alarm Network
Atomic Inference

- Given unlimited time, inference in BNs is easy
- Recipe:
  - State the marginal probabilities you want
  - Figure out ALL the atomic probabilities you need
  - Calculate and combine them
- Example:

\[
P(b|j, m) = \frac{P(b, j, m)}{P(j, m)}
\]

Example

\[
P(b|j, m) = \frac{P(b, j, m)}{P(j, m)}
\]

\[
P(b, j, m) = P(b, e, a, j, m) + P(b, \bar{e}, a, j, m) + P(b, e, \bar{a}, j, m) + P(b, \bar{e}, \bar{a}, j, m)
\]

\[
= \sum_{e,a} P(b, e, a, j, m)
\]
Example

\[ P(b, j, m) = \]
\[ P(b)P(e)P(a|b, e)P(j|a)P(m|a) + \]
\[ P(b)P(e)P(\bar{a}|b, e)P(j|\bar{a})P(m|\bar{a}) + \]
\[ P(b)P(\bar{e})P(a|b, \bar{e})P(j|a)P(m|a) + \]
\[ P(b)P(\bar{e})P(\bar{a}|b, \bar{e})P(j|\bar{a})P(m|\bar{a}) \]

Example

\[ P(b|j, m) = \frac{P(b, j, m)}{P(j, m)} \]
\[ P(b, j, m) = \sum_{e,a} P(b, c, a, j, m) \]
\[ P(\bar{b}, j, m) = \sum_{e,a} P(\bar{b}, c, a, j, m) \]

\[
\begin{pmatrix}
P(b, j, m) \\
P(\bar{b}, j, m)
\end{pmatrix}
\xrightarrow{\text{Normalize}}
\begin{pmatrix}
P(b|j, m) \\
P(\bar{b}|j, m)
\end{pmatrix}
Inference by Enumeration

- Atomic inference is extremely slow!
- Slightly clever way to save work:
  - Move the sums as far right as possible
  - Example:

\[
P(b, j, m) = \sum_{e, a} P(b, e, a, j, m) \\
= \sum_{e, a} P(b) P(e) P(a|b, e) P(j|a) P(m|a) \\
= P(b) \sum_{e} P(e) \sum_{a} P(a|b, e) P(j|a) P(m|a)
\]

Example

\[
P(b, j, m) = P(b) P(e) P(a|b, e) P(j|a) P(m|a) + \\
= P(b) P(e) P(\bar{a}|b, e) P(j|\bar{a}) P(m|\bar{a}) + \\
= P(b) P(\bar{e}) P(a|b, \bar{e}) P(j|a) P(m|a) + \\
= P(b) P(\bar{e}) P(\bar{a}|b, \bar{e}) P(j|\bar{a}) P(m|\bar{a})
\]
**Evaluation Tree**

- View the nested sums as a computation tree:

```
                      P(b) = 0.01
                      /       |
       P(c) = 0.02 /     \   P(c̄) = 0.98
        /     \       /     \
   P(a | b,c) = 0.95 /     \   P(a | b,c̄) = 0.05
     /     \       /     \    /     \    /     \    /     \
P(j | a) = 0.90  P(j̄ | a) = 0.10  P(j | a) = 0.90  P(j̄ | a) = 0.10
     /     \       /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \    /     \  
P(m | a) = 0.70  P(m | a) = 0.30  P(m | a) = 0.70  P(m | a) = 0.30
```

- Still repeated work: calculate $P(m \mid a) P(j \mid a)$ twice, etc.

---

**Variable Elimination: Idea**

- Lots of redundant work in the computation tree!

- We can save time if we cache all partial results

- This is the basic idea behind variable elimination
Basic Objects

- Track objects called factors
- Initial factors are local CPTs

\[
\begin{align*}
P(B) & \quad f_B(B) \\
P(J|A) & \quad f_J(A, J) \\
P(A|B, E) & \quad f_A(A, B, E)
\end{align*}
\]

- During elimination, create new factors
- Anatomy of a factor:

\[
f_{A\overline{B}\overline{C}}(D, E)
\]

Variables introduced

Variables summed out

Argument variables, always non-evidence variables

4 numbers, one for each value of D and E

Basic Operations

- First basic operation: join factors
- Combining two factors:
  - Just like a database join
  - Build a factor over the union of the domains
- Example:

\[
f_1(A, B) \times f_2(B, C) \rightarrow f_3(A, B, C)
\]

\[
f_3(a, b, c) = f_1(a, b) \cdot f_2(b, c)
\]

"\[ P(a, b|c) = P(a|b) \cdot P(b|c) \]"
Basic Operations

- Second basic operation: marginalization
- Take a factor and sum out a variable
  - Shrinks a factor to a smaller one
  - A projection operation
- Example:

\[
    f_{\overline{AB}}(b) = \sum_a f_{AB}(a, b)
\]

“\( P(b) = \sum_a P(a, b) \)”

Example

\[
    P(b, j, m)
    = \frac{P(b)}{B} \sum_e \frac{P(e)}{E} \sum_a \frac{P(a|b, e)}{A} \frac{P(j|a)}{J} \frac{P(m|a)}{M}
    = f_B(b) \sum_e f_E(e) \sum_a f_A(a, b, e) f_J(a) f_M(a)
    = f_B(b) \sum_e f_E(e) \sum_a f_{AJM}(a, b, e)
    = f_B(b) \sum_e f_E(e) f_{\overline{AJM}}(b, e)
\]
Example

\[ P(b, j, m) = f_B(b) \sum_e f_E(e) f_{\overline{A}J_{M}}(b, e) \]
\[ = f_B(b) \sum_e f_{\overline{A}E_{J_{M}}}(b, e) \]
\[ = f_B(b) f_{\overline{A}E_{J_{M}}}(b) \]
\[ = f_{\overline{A}B\overline{E}_{J_{M}}}(b) \]

General Variable Elimination

- **Query:** \[ P(Q|E_1 = e_1, \ldots E_k = e_k) \]
- **Start with initial factors:**
  - Local CPTs (but instantiated by evidence)
- **While there are still hidden variables (not Q or evidence):**
  - Pick a hidden variable H
  - Join all factors mentioning H
  - Project out H
- **Join all remaining factors and normalize**
Example

\[ P(B|j, m) \propto P(B, j, m) \]

\[
\begin{array}{cccccc}
P(B) & P(E) & P(A|B, E) & P(j|A) & P(m|A) \\
 f_B(B) & f_E(E) & f_A(A, B, E) & f_J(A) & f_M(A)
\end{array}
\]

Choose A

\[
\begin{array}{ccc}
 f_A(A, B, E) \\
f_J(A) \\
f_M(A)
\end{array} \quad \times \quad \begin{array}{c}
 f_{AJM}(A, B, E) \\
\Sigma
\end{array} \quad \begin{array}{c}
 f_{\bar{A}JM}(B, E)
\end{array}
\]

\[
\begin{array}{ccc}
f_B(B) & f_E(E) & f_{AJM}(B, E)
\end{array}
\]

Example

\[
\begin{array}{ccc}
f_B(B) & f_E(E) & f_{\bar{A}JM}(B, E)
\end{array}
\]

Choose E

\[
\begin{array}{ccc}
f_E(E) \\
f_{\bar{AJM}}(B, E) \\
\times
\end{array} \quad \begin{array}{c}
 f_{\bar{AJEM}}(B, E) \\
\Sigma
\end{array} \quad \begin{array}{c}
 f_{\bar{AEJM}}(B)
\end{array}
\]

\[
\begin{array}{c}
f_B(B) \\
f_{\bar{AEJM}}(B)
\end{array}
\]

Finish

\[
\begin{array}{ccc}
f_B(B) \\
f_{\bar{AEJM}}(B) \\
\times
\end{array} \quad \begin{array}{c}
 f_{\bar{ARFM}}(B) \\
\text{Normalize}
\end{array} \quad \begin{array}{c}
 P(B|j, m)
\end{array}
\]