Uncertainty: wrap up
& Decision Theory: Intro

CPSC 322 – Decision Theory 1

Textbook §6.4.1 & §9.2

March 30, 2011
Remarks on Assignment 4

• Question 2 (Bayesian networks)
  – “correctly represent the situation described above” means “do not make any independence assumptions that aren’t true”
    • Step 1: identify the causal network
    • Step 2: for each network, check if it entails (conditional or marginal) independencies the causal network does not entail. If so, it’s incorrect
  – Failing to entail some (or all) independencies does not make a network incorrect (only computationally suboptimal)

• Question 5 (Rainbow Robot)
  – If you got rainbowrobot.zip before Sunday, get the updated version: rainbowrobot_updated.zip (on WebCT)

• Question 4 (Decision Networks)
  – This is mostly Bayes rule and common sense
  – One could compute the answer algorithmically, but you don’t need to
Lecture Overview

Variable elimination: recap and some more details
• Variable elimination: pruning irrelevant variables
• Summary of Reasoning under Uncertainty
• Decision Theory
  – Intro
  – Time-permitting: Single-Stage Decision Problems
Recap: Factors and Operations on them

- A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$.

- Operation 1: assigning a variable in a factor
  - E.g., assign $X = t$

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>$f_1(X,Y,Z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>t</td>
<td>0.1</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>f</td>
<td>0.9</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>t</td>
<td>0.2</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>f</td>
<td>0.8</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>t</td>
<td>0.4</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>f</td>
<td>0.6</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>t</td>
<td>0.3</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>f</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Given $f_1(X,Y,Z)_{X=t} = f_2(Y,Z)$,

<table>
<thead>
<tr>
<th>Y</th>
<th>Z</th>
<th>$f_2(Y,Z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>0.1</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>0.9</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>0.2</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Factor of $Y,Z$
Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$
- Operation 1: assigning a variable in a factor
  - $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor

$$\sum_B f_3(A,B,C) = f_4(A,C)$$

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
<th>C</th>
<th>$f_3(A,B,C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>t</td>
<td>0.03</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>f</td>
<td>0.07</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>t</td>
<td>0.54</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>f</td>
<td>0.36</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>t</td>
<td>0.06</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>f</td>
<td>0.14</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>t</td>
<td>0.48</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>f</td>
<td>0.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>C</th>
<th>$f_4(A,C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>0.57</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>0.43</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>0.54</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Recap: Operation 3: multiplying factors

\[ f_5(A,B) \times f_6(B,C) = f_7(A,B,C), \text{ i.e.} \]

\[ f_5(A=a,B=b) \times f_6(B=b,C=c) = f_7(A=a,B=b,C=c) \]
Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- If we assign variable $A=a$ in factor $f_7(A,B)$, what is the correct form for the resulting factor?

  $f(A)$  $f(B)$  $f(A,B)$  A number
Recap: Factors and Operations on Them

• A factor is a function from a tuple of random variables to the real numbers \( \mathbb{R} \).

• Operation 1: assigning a variable in a factor
  – E.g., \( f_2(Y,Z) = f_1(X,Y,Z)_{X=t} \)

• Operation 2: marginalize out a variable from a factor
  – E.g., \( f_4(A,C) = \sum_B f_3(A,B,C) \)

• Operation 3: multiply two factors
  – E.g. \( f_7(A,B,C) = f_5(A,B) \times f_6(B,C) \)
    • That means, \( f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c) \)

• If we assign variable \( A=a \) in factor \( f_7(A,B) \), what is the correct form for the resulting factor?
  – \( f(B) \).
    When we assign variable \( A \) we remove it from the factor’s domain.
Recap: Factors and Operations on Them

• A factor is a function from a tuple of random variables to the real numbers \( \mathbb{R} \).

• Operation 1: assigning a variable in a factor
  – E.g., \( f_2(Y,Z) = f_1(X,Y,Z)_{X=t} \)

• Operation 2: marginalize out a variable from a factor
  – E.g., \( f_4(A,C) = \sum_B f_3(A,B,C) \)

• Operation 3: multiply two factors
  – E.g. \( f_7(A,B,C) = f_5(A,B) \times f_6(B,C) \)
    • That means, \( f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c) \)

• If we marginalize variable A out from factor \( f_7(A,B) \), what is the correct form for the resulting factor?
  - \( f(A) \)  \( f(B) \)  \( f(A,B) \)  A number
Recap: Factors and Operations on Them

• A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$

• Operation 1: assigning a variable in a factor
  – E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$

• Operation 2: marginalize out a variable from a factor
  – E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$

• Operation 3: multiply two factors
  – E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    • That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

• If we assign variable $A=a$ in factor $f_7(A,B)$, what is the correct form for the resulting factor?
  – $f(B)$.
    When we marginalize out variable $A$ we remove it from the factor’s domain
Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- If we multiply factors $f_4(X,Y)$ and $f_6(Z,Y)$, what is the correct form for the resulting factor?
  - $f(X,Z)$
  - $f(X)$
  - $f(X,Y)$
  - $f(X,Y,Z)$
Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers \( \mathbb{R} \).
- Operation 1: assigning a variable in a factor
  - E.g., \( f_2(Y,Z) = f_1(X,Y,Z)_{X=t} \)
- Operation 2: marginalize out a variable from a factor
  - E.g., \( f_4(A,C) = \sum_B f_3(A,B,C) \)
- Operation 3: multiply two factors
  - E.g. \( f_7(A,B,C) = f_5(A,B) \times f_6(B,C) \)
    - That means, \( f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c) \)

- If we multiply factors \( f_4(X,Y) \) and \( f_6(Z,Y) \), what is the correct form for the resulting factor?
  - \( f(X,Y,Z) \)
  - When multiplying factors, the resulting factor’s domain is the union of the multiplicands’ domains
Recap: Factors and Operations on Them

- A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$
- Operation 1: assigning a variable in a factor
  - E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$
- Operation 2: marginalize out a variable from a factor
  - E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$
- Operation 3: multiply two factors
  - E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    - That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

- What is the correct form for $\sum_B f_5(A,B) \times f_6(B,C)$
  - As usual, product before sum: $\sum_B ( f_5(A,B) \times f_6(B,C) )$
Recap: Factors and Operations on Them

• A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$

• Operation 1: assigning a variable in a factor
  – E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$

• Operation 2: marginalize out a variable from a factor
  – E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$

• Operation 3: multiply two factors
  – E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    • That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

• What is the correct form for $\sum_B f_5(A,B) \times f_6(B,C)$
  – As usual, product before sum: $\sum_B ( f_5(A,B) \times f_6(B,C) )$
  – Result of multiplication: $f(A,B,C)$. Then marginalize out $B$: $f'(A,C)$
Recap: Factors and Operations on Them

• A factor is a function from a tuple of random variables to the real numbers $\mathbb{R}$
  
• Operation 1: assigning a variable in a factor
  – E.g., $f_2(Y,Z) = f_1(X,Y,Z)_{X=t}$

• Operation 2: marginalize out a variable from a factor
  – E.g., $f_4(A,C) = \sum_B f_3(A,B,C)$

• Operation 3: multiply two factors
  – E.g. $f_7(A,B,C) = f_5(A,B) \times f_6(B,C)$
    • That means, $f_7(A=a,B=b,C=c) = f_5(A=a,B=b) \times f_6(B=b,C=c)$

• Operation 4: normalize the factor
  – Divide each entry by the sum of the entries. The result will sum to 1.

<table>
<thead>
<tr>
<th>A</th>
<th>$f_5(A,B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>0.4</td>
</tr>
<tr>
<td>f</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>$f_6(A,B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>0.4/(0.4+0.1) = 0.8</td>
</tr>
<tr>
<td>f</td>
<td>0.1/(0.4+0.1) = 0.2</td>
</tr>
</tbody>
</table>
Recap: the Key Idea of Variable Elimination

• An efficient way to sum out a variable $Z_k$ from a product $f_1 \times \ldots \times f_k$ of factors:
  – Partition the factors into
    • those that don't contain $Z_k$, say $f_1 \times \ldots \times f_i$
    • those that contain $Z_k$, say $f_{i+1} \times \ldots \times f_k$

• Since multiplication distributes over addition:

\[
\sum_{Z_k} f_1 \times \ldots \times f_k = f_1 \times \ldots \times f_i \times \left( \sum_{Z_k} f_{i+1} \times \ldots \times f_k \right)
\]

New factor! Let’s call it $f'$

• Store $f'$ explicitly, and discard $f_{i+1} \ldots f_k$
• Now we've summed out $Z_k$
Recap: Variable Elimination (VE) in BNs

- The joint probability distribution of a Bayesian network is
  \[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{pa}(X_i)) \]
  - We make a factor \( f_i \) for each conditional probability table \( P(X_i|\text{pa}(X_i)) \)
  - So we have \( P(X_1, \ldots, X_n) = \prod_{i=1}^{n} f_i \)

- The variable elimination algorithm computes \( P(Y| E_1=e_1, \ldots, E_j=e_j) \) as follows:
  - Assign \( E_1=e_1, \ldots, E_j=e_j \)
  - Sum out all non-query variables \( Z_1, \ldots, Z_k \), one at a time
    - To sum out \( Z_i \):
      - Multiply factors containing it \( Z_i \)
      - Then marginalize out \( Z_i \) from the product
    - The order in which we sum out variables is called our elimination ordering
  - Normalize the final factor \( f(Y) \).
    - The resulting factor is exactly \( P(Y| E_1=e_1, \ldots, E_j=e_j) \)
Recap: VE example: compute $P(G|H=h_1)$
Step 1: construct a factor for each cond. probability

$$P(G,H) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_7(H,G) f_8(I,G)$$
Recap: VE example: compute $P(G|H=h_1)$

Step 2: assign observed variables their observed value

$$P(G,H) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_7(H,G) f_8(I,G)$$

Assigning the variable $H=h_1$:

$$f_9(G) = f_7(H,G)_{H=h_1}$$
Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non-query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) \cdot f_1(B,A) \cdot f_2(C) \cdot f_3(D,B,C) \cdot f_4(E,C) \cdot f_5(F,D) \cdot f_6(G,F,E) \cdot f_9(G) \cdot f_8(I,G)$$

Elimination ordering: A, C, E, I, B, D, F
Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non-query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) \cdot f_1(B,A) \cdot f_2(C) \cdot f_3(D,B,C) \cdot f_4(E,C) \cdot f_5(F,D) \cdot f_6(G,F,E) \cdot f_8(I,G) \cdot f_9(G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C) \cdot f_3(D,B,C) \cdot f_4(E,C) \cdot f_5(F,D) \cdot f_6(G,F,E) \cdot f_8(I,G) \cdot f_9(G) \cdot f_{10}(B)$$

Summing out variable $A$:

$$\sum_A f_0(A) \cdot f_1(B,A) = f_{10}(B)$$

Elimination ordering: $A, C, E, I, B, D, F$
Recap: VE example: compute $P(G|H=h_1)$
Step 4: sum out non-query variables (one at a time)

\[
P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) \ f_1(B,A) \ f_2(C) \ f_3(D,B,C) \ f_4(E,C) \ f_5(F,D) \ f_6(G,F,E) \ f_9(G) \ f_8(I,G)
\]
\[
= \sum_{B,C,D,E,F,I} f_2(C) \ f_3(D,B,C) \ f_4(E,C) \ f_5(F,D) \ f_6(G,F,E) \ f_9(G) \ f_8(I,G) \ f_{10}(B)
\]
\[
= \sum_{B,D,E,F,I} f_5(F,D) \ f_6(G,F,E) \ f_9(G) \ f_8(I,G) \ f_{10}(B) \ f_{11}(B,D,E)
\]

Elimination ordering: A, C, E, I, B, D, F

Summing out variable $C$:
\[
\sum_C f_2(C) \ f_3(D,B,C) \ f_4(E,C) = f_{11}(B,D,E)
\]
Recap: VE example: compute $P(G|H=h_1)$
Step 4: sum out non-query variables (one at a time)

\[
P(G, H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F, D) f_6(G,F,E) f_9(G) f_8(I,G)
\]

\[
= \sum_{B,C,D,E,F,I} f_2(C) f_3(D,B,C) f_4(E,C) f_5(F, D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B)
\]

\[
= \sum_{B,D,E,F,I} f_5(F, D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B) f_{11}(B,D,E)
\]

\[
= \sum_{B,D,F,I} f_5(F, D) f_9(G) f_8(I,G) f_{10}(B) f_{12}(G,F,B,D)
\]

Summing out variable $E$:

\[
\sum_{E} f_6(G,F,E) f_{11}(B,D,E) = f_{12}(G,F,B,D)
\]

Elimination ordering: $A$, $C$, $E$, $I$, $B$, $D$, $F$
Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non-query variables (one at a time)

$$P(G, H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F, D) f_6(G,F,E) f_9(G) f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C) f_3(D,B,C) f_4(E,C) f_5(F, D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F, D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B) f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F, D) f_9(G) f_{10}(B) f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F, D) f_9(G) f_{10}(B) f_{12}(G,F,B,D) f_{13}(G)$$
Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non-query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B) f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F,D) f_9(G) f_8(I,G) f_{10}(B) f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F,D) f_9(G) f_{10}(B) f_{12}(G,F,B,D) f_{13}(G)$$

$$= \sum_{D,F} f_5(F,D) f_9(G) f_{11}(G,F) f_{12}(G) f_{14}(G,F,D)$$

Elimination ordering: $A, C, E, I, B, D, F$
Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non-query variables (one at a time)

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B) f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F,D) f_9(G) f_{10}(B) f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F,D) f_9(G) f_{10}(B) f_{12}(G,F,B,D) f_{13}(G)$$

$$= \sum_{D,F} f_5(F,D) f_9(G) f_{11}(G,F) f_{12}(G) f_{14}(G,F,D)$$

$$= \sum_{F} f_9(G) f_{11}(G,F) f_{12}(G) f_{15}(G,F)$$

Elimination ordering: A, C, E, I, B, D, F
Recap: VE example: compute $P(G|H=h_1)$

Step 4: sum out non-query variables (one at a time)

$$P(G, H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B) f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F,D) f_9(G) f_8(I,G) f_{10}(B) f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F,D) f_9(G) f_{10}(B) f_{12}(G,F,B,D) f_{13}(G)$$

$$= \sum_{D,F} f_5(F,D) f_9(G) f_{11}(G,F) f_{12}(G) f_{14}(G,F,D)$$

$$= \sum_{F} f_9(G) f_{11}(G,F) f_{12}(G) f_{15}(G,F)$$

$$= f_9(G) f_{12}(G) f_{16}(G)$$

Elimination ordering: A, C, E, I, B, D, F
Recap: VE example: compute $P(G|H=h_1)$
Step 5: multiply the remaining factors

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G)$$
$$= \sum_{B,C,D,E,F,I} f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_10(B)$$
$$= \sum_{B,D,E,F,I} f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_10(B) f_{11}(B,D,E)$$
$$= \sum_{B,D,F,I} f_5(F,D) f_9(G) f_8(I,G) f_10(B) f_{12}(G,F,B,D)$$
$$= \sum_{B,D,F} f_5(F,D) f_9(G) f_{10}(B) f_{12}(G,F,B,D) f_{13}(G)$$
$$= \sum_{D,F} f_5(F,D) f_9(G) f_{11}(G,F) f_{12}(G) f_{14}(G,F,D)$$
$$= \sum_{F} f_9(G) f_{11}(G,F) f_{12}(G) f_{15}(G,F)$$
$$= f_9(G) f_{12}(G) f_{16}(G)$$
$$= f_{17}(G)$$
Recap: VE example: compute $P(G|H=h_1)$

Step 6: normalize

$$P(G,H=h_1) = \sum_{A,B,C,D,E,F,I} f_0(A) f_1(B,A) f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G)$$

$$= \sum_{B,C,D,E,F,I} f_2(C) f_3(D,B,C) f_4(E,C) f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B)$$

$$= \sum_{B,D,E,F,I} f_5(F,D) f_6(G,F,E) f_9(G) f_8(I,G) f_{10}(B) f_{11}(B,D,E)$$

$$= \sum_{B,D,F,I} f_5(F,D) f_9(G) f_8(I,G) f_{10}(B) f_{12}(G,F,B,D)$$

$$= \sum_{B,D,F} f_5(F,D) f_9(G) f_{10}(B) f_{12}(G,F,B,D) f_{13}(G)$$

$$= \sum_{D,F} f_5(F,D) f_9(G) f_{11}(G,F) f_{12}(G) f_{14}(G,F,D)$$

$$= \sum_{F} f_9(G) f_{11}(G,F) f_{12}(G) f_{15}(G,F)$$

$$= f_9(G) f_{12}(G) f_{16}(G)$$

$$= f_{17}(G)$$

$$P(G = g \mid H = h_1) = \frac{f_{17}(g)}{\sum_{g' \in \text{dom}(G)} f_{17}(g')}$$
Lecture Overview

• Variable elimination: recap and some more details
• Variable elimination: pruning irrelevant variables
• Summary of Reasoning under Uncertainty
• Decision Theory
  – Intro
  – Time-permitting: Single-Stage Decision Problems
Recap: conditional independence in BNs

- Two variables $X$ and $Y$ are conditionally independent given a set of observed variables $E$, if and only if
  - There is no path along which information can flow from $X$ to $Y$
  - Information can flow along a path if it can flow through all the nodes in the path.

- Note: observation status of $A$ and $C$ does not matter.
Conditional independence in BNs

- Memoization trick:
  - Assume that whether kids are nice depends only on whether their parents are nice
  - Assume that people get married independent of their niceness
  - Then “child” in a Bayesian network translates to child in the real world

Your grandparent is nice, so your parent is likely to be nice, so you are likely to be nice. But if we know how nice your parent is, the grandparent’s niceness doesn’t provide extra information.

Nice people are likely to have nice siblings since they have the same parent. But if you know the parent’s niceness, then that explains everything.

The dad is nice, that tells us nothing about the mom. But if we know the kid is mean, the mom is likely mean.
Conditional independence in BNs example

- Is E marginally independent of C?
  - No. Information flows between them (through all nodes on the path).
Conditional independence in BNs example

- Is E marginally independent of C?
  - No. Information flows between them (through all nodes on the path).
- What if we observe A?
  - I.e., is E conditionally independent of C given A?
  - Yes. The observed node in a chain blocks information.
Conditional independence in BNs example

• Is E marginally independent of C?
  – No. Information flows between them (through all nodes on the path).

• What if we observe A?
  – I.e., is E conditionally independent of C given A?
  – Yes. The observed node in a chain blocks information.

• What if we add nodes F and G (observed)?
  – Now the information can flow again
  – So E and C are not conditionally independent given G and A
VE and conditional independence

• So far, we haven't use conditional independence in VE!
  – Before running VE, we can prune all variables $Z$ that are conditionally independent of the query $Y$ given evidence $E$: $Z \perp Y \mid E$

• Example: which variables can we prune for the query $P(G=g \mid C=c_1, F=f_1, H=h_1)$?
VE and conditional independence

• So far, we haven’t use conditional independence!
  – Before running VE, we can prune all variables \( Z \) that are conditionally independent of the query \( Y \) given evidence \( E \):
    \[ Z \perp Y \mid E \]

• Example: which variables can we prune for the query \( P(G=g \mid C=c_1, F=f_1, H=h_1) \) ?
  – A, B, and D. Both paths are blocked
    • \( F \) is an observed node in a chain structure
    • \( C \) is an observed common parent
  – Thus, we only need to consider this subnetwork
Variable Elimination: One last trick

- We can also prune unobserved leaf nodes
  - And we can do so recursively

E.g., which nodes can we prune if the query is \( P(A) \)?

Recursively prune unobserved leaf nodes:
we can prune all nodes other than \( A \)!
Complexity of Variable Elimination (VE)

• A factor over n binary variables has to store $2^n$ numbers
  – The initial factors are typically quite small
    (variables typically only have few parents in Bayesian networks)
  – But variable elimination constructs larger factors
    by multiplying factors together

• The complexity of VE is exponential in the maximum number of variables in any factor during its execution
  – This number is called the treewidth of a graph (along an ordering)
  – Elimination ordering influences treewidth

• Finding the best ordering is NP complete
  – I.e., the ordering that generates the minimum treewidth
  – Heuristics work well in practice (e.g. least connected variables first)
  – Even with best ordering, inference is sometimes infeasible
    • In those cases, we need approximate inference. See CS422 & CS540
Lecture Overview

• Variable elimination: recap and some more details
• Variable elimination: pruning irrelevant variables

Summary of Reasoning under Uncertainty

• Decision Theory
  – Intro
  – Time-permitting: Single-Stage Decision Problems
Big picture: Reasoning Under Uncertainty

Probability Theory

Bayesian Networks & Variable Elimination
- Monitoring (e.g. credit card fraud detection)
- Diagnostic systems (e.g. medicine)

Dynamic Bayesian Networks
- Bioinformatics
- Natural Language Processing

Hidden Markov Models & Filtering
- Motion Tracking, Missile Tracking, etc
- Email spam filters

Diagnostic systems (e.g. medicine)
One Realistic BNet: Liver Diagnosis

Source: Onisko et al., 1999

~60 nodes, max 4 parents per node
Course Overview

Environment
- Deterministic
- Stochastic

Problem Type
- Logic
- Planning
- Static
- Sequential

Constraint Satisfaction

Variables + Constraints
- Arc Consistency
- Search

Logics
- STRIPS
- Search
- As CSP (using arc consistency)

Bayesian Networks
- Variable Elimination

Decision Networks
- Variable Elimination

Markov Processes
- Value Iteration

Course Module
- Representation
  - Reasoning
  - Technique

This concludes the uncertainty module

Decision Theory

Static

Uncertainty
<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Logic</th>
<th>Static</th>
<th>Sequential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraint Satisfaction</td>
<td>STRIPS</td>
<td>As CSP (using arc consistency)</td>
<td>Planning</td>
</tr>
<tr>
<td>Deterministic Environment</td>
<td>Arc Consistency</td>
<td>Search</td>
<td>Value Iteration</td>
</tr>
<tr>
<td>Stochastic Environment</td>
<td>Bayesian Networks</td>
<td>Variable Elimination</td>
<td>Decision Theory</td>
</tr>
<tr>
<td>Search Variables + Constraints</td>
<td>Decision Networks</td>
<td>Variable Elimination</td>
<td>Markov Processes</td>
</tr>
</tbody>
</table>

But uncertainty is also at the core of decision theory: now we’re acting under uncertainty.
Lecture Overview

- Variable elimination: recap and some more details
- Variable elimination: pruning irrelevant variables
- Summary of Reasoning under Uncertainty

Decision Theory
- Intro
- Time-permitting: Single-Stage Decision Problems
Decisions Under Uncertainty: Intro

• Earlier in the course, we focused on decision making in deterministic domains
  – Search/CSPs: single-stage decisions
  – Planning: sequential decisions

• Now we face **stochastic domains**
  – so far we've considered how to represent and update beliefs
  – What if an agent has to make decisions under uncertainty?

• Making decisions under uncertainty is important
  – We mainly represent the world probabilistically so we can use our beliefs as the basis for making decisions
Decisions Under Uncertainty: Intro

• An agent's decision will depend on
  – What actions are available
  – What beliefs the agent has
  – Which goals the agent has

• Differences between deterministic and stochastic setting
  – Obvious difference in representation: need to represent our uncertain beliefs
  – Now we'll speak about representing actions and goals
    • Actions will be pretty straightforward: decision variables
    • Goals will be interesting: we'll move from all-or-nothing goals to a richer notion: rating how happy the agent is in different situations.
    • Putting these together, we'll extend Bayesian networks to make a new representation called decision networks
Lecture Overview

• Variable elimination: recap and some more details
• Variable elimination: pruning irrelevant variables
• Summary of Reasoning under Uncertainty
• Decision Theory
  – Intro
  – Time-permitting: Single-Stage Decision Problems
Delivery Robot Example

• Decision variable 1: the robot can choose to wear pads
  – Yes: protection against accidents, but extra weight
  – No: fast, but no protection

• Decision variable 2: the robot can choose the way
  – Short way: quick, but higher chance of accident
  – Long way: safe, but slow

• Random variable: is there an accident?

Agent decides
Chance decides
Possible worlds and decision variables

• A possible world specifies a value for each random variable and each decision variable.

• For each assignment of values to all decision variables
  – the probabilities of the worlds satisfying that assignment sum to 1.
Possible worlds and decision variables

• A **possible world** specifies a value for each random variable and each decision variable

• For each assignment of values to all decision variables
  – the probabilities of the worlds satisfying that assignment sum to 1.

![Diagram showing possible worlds and decision variables with conditional probabilities]
Possible worlds and decision variables

• A possible world specifies a value for each random variable and each decision variable.

• For each assignment of values to all decision variables:
  – the probabilities of the worlds satisfying that assignment sum to 1.

```
<table>
<thead>
<tr>
<th>World</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>w0</td>
<td>0.2</td>
</tr>
<tr>
<td>w1</td>
<td>0.8</td>
</tr>
<tr>
<td>w2</td>
<td>0.01</td>
</tr>
<tr>
<td>w3</td>
<td>0.99</td>
</tr>
<tr>
<td>w4</td>
<td>0.2</td>
</tr>
<tr>
<td>w5</td>
<td>0.8</td>
</tr>
<tr>
<td>w6</td>
<td>0.2</td>
</tr>
<tr>
<td>w7</td>
<td>0.8</td>
</tr>
</tbody>
</table>
```
Possible worlds and decision variables

- A possible world specifies a value for each random variable and each decision variable.
- For each assignment of values to all decision variables:
  - the probabilities of the worlds satisfying that assignment sum to 1.
Possible worlds and decision variables

- A **possible world** specifies a value for each random variable and each decision variable.
- For each assignment of values to all decision variables:
  - the probabilities of the worlds satisfying that assignment sum to 1.
Utility

- **Utility**: a measure of desirability of possible worlds to an agent
  - Let $U$ be a real-valued function such that $U(w)$ represents an agent's degree of preference for world $w$
  - Expressed by a number in $[0,100]$

- Simple goals can still be specified
  - Worlds that satisfy the goal have utility 100
  - Other worlds have utility 0

- Utilities can be more complicated
  - For example, in the robot delivery domains, they could involve
    - Amount of damage
    - Reached the target room?
    - Energy left
    - Time taken
Combining probabilities and utilities

• We can combine probability with utility
  – The expected utility of a probability distribution over possible worlds average utility, weighted by probabilities of possible worlds
  – What is the expected utility of Wearpads=yes, Way=short ?
    • It is $0.2 \times 35 + 0.8 \times 95 = 83$
Expected utility

• Suppose $U(w)$ is the utility of possible world $w$ and $P(w)$ is the probability of possible world $w$

Definition (expected utility)
The expected utility is

$$E[U] = \sum_w P(w)U(w)$$

Definition (expected utility)
The conditional expected utility given $e$ is

$$E[U|e] = \sum_w P(w|e)U(w)$$
Expected utility of a decision

- We write the expected utility of a decision as:

\[ E[U|D = d] = \sum_w P(w|D = d)U(w) \]

| Conditional probability | Utility | E[U|D] |
|-------------------------|---------|--------|
| 0.2                    | 35      | 83     |
| 0.8                    | 95      |        |
| 0.01                   | 30      | 74.55  |
| 0.99                   | 75      |        |
| 0.2                    | 3       | 80.6   |
| 0.8                    | 100     |        |
| 0.01                   | 0       | 79.2   |
| 0.99                   | 80      |        |

Diagram:

- Wear pads:
  - Short way: 0.2, Utility: 35, E[U|D] = 83
  - Long way: 0.8, Utility: 95

- Don't wear pads:
  - Short way: 0.01, Utility: 30, E[U|D] = 74.55
  - Long way: 0.99, Utility: 75
Optimal single-stage decision

- Given a single decision variable $D$
  - the agent can choose $D = d_i$ for any value $d_i \in \text{dom}(D)$

**Definition (optimal single-stage decision)**

An optimal single-stage decision is the decision $D = d_{max}$ whose expected value is maximal:

$$d_{max} \in \arg \max_{d_i \in \text{dom}(D)} E[U|D=d_i]$$
Learning Goals For Today’s Class

• Identify implied (in)dependencies in the network
• Variable elimination
  – Carry out variable elimination by using factor representation and using the factor operations
  – Use techniques to simplify variable elimination
• Define a Utility Function on possible worlds
• Define and compute optimal one-off decisions

• Assignment 4 is due on Monday
  – You should now be able to solve Questions 1, 2, 3, and 5
  – And basically Question 4

• Final exam: Monday, April 11