Outline

- Bayesian Networks Inference
  - Exact Inference: Variable Elimination
  - Approximate Inference: Sampling
Bayes Net
Representation

- Burglary
- Earthqk
- John calls
- Mary calls
- Alarm
Bayes’ Net Semantics

- Let’s formalize the semantics of a Bayes’ net
- A set of nodes, one per variable $X$
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
    $$P(X|a_1 \ldots a_n)$$
- CPT: conditional probability table

A Bayes net = Topology (graph) + Local Conditional Probabilities
Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars \{Z\}?
  - Yes, if X and Y “separated” by Z
  - Look for active paths from X to Y
  - No active paths = independence!

- A path is active if each triple is active:
  - Causal chain \(A \rightarrow B \rightarrow C\) where B is unobserved (either direction)
  - Common cause \(A \leftarrow B \rightarrow C\) where B is unobserved
  - Common effect (aka v-structure) \(A \rightarrow B \leftarrow C\) where B or one of its descendents is observed

- All it takes to block a path is a single inactive segment
Bayes Net Joint Distribution

\[
P(+b, -e, +a, -j, +m) =
P(+b)P(-e)P(+a| + b, -e)P(-j| + a)P(+m| + a) =
\]
Bayes Net Joint Distribution

\[
P(+b, -e, +a, -j, +m) = P(+b)P(-e)P(+a| +b, -e)P(-j| +a)P(+m| +a) =
\]
\[
0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

- We generally compute conditional probabilities
  - \( P(\text{on time } | \text{ no reported accidents}) = 0.90 \)
  - These represent the agent’s beliefs given the evidence

- Probabilities change with new evidence:
  - \( P(\text{on time } | \text{ no accidents, 5 a.m.}) = 0.95 \)
  - \( P(\text{on time } | \text{ no accidents, 5 a.m., raining}) = 0.80 \)
  - Observing new evidence causes beliefs to be updated
Inference

- Inference: calculating some useful quantity from a joint probability distribution

Examples:

- Posterior probability

\[
P(Q|E_1 = e_1, \ldots E_k = e_k)
\]

- Most likely explanation:

\[
\arg\max_Q P(Q = q|E_1 = e_1 \ldots)
\]
Inference by Enumeration

- **General case:**
  - Evidence variables: \( E_1 \ldots E_k = e_1 \ldots e_k \)
  - Query* variable: \( Q \)
  - Hidden variables: \( H_1 \ldots H_r \)

\[
\begin{aligned}
X_1, X_2, \ldots X_n
\end{aligned}
\]

All variables

- We want: \( P(Q|e_1 \ldots e_k) \)
- First, select the entries consistent with the evidence
- Second, sum out \( H \) to get joint of Query and evidence:

\[
P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(Q, h_1 \ldots h_r, e_1 \ldots e_k) \]

\[
X_1, X_2, \ldots X_n
\]

- Finally, normalize the remaining entries to conditionalize

- **Obvious problems:**
  - Worst-case time complexity \( O(d^n) \)
  - Space complexity \( O(d^n) \) to store the joint distribution
Inference in BN by Enumeration

- Given unlimited time, inference in BNs is easy
- Reminder of inference by enumeration by example:

\[
P(B \mid +j, +m) \propto_B P(B, +j, +m)
\]

\[
= \sum_{e,a} P(B, e, a, +j, +m)
\]

\[
= \sum_{e,a} P(B)P(e)P(a \mid B, e)P(+j \mid a)P(+m \mid a)
\]

\[
= P(B)P(+e)P(+a \mid B, +e)P(+j \mid +a)P(+m \mid +a) + P(B)P(+e)P(-a \mid B, +e)P(+j \mid -a)P(+m \mid -a)
\]

\[
P(B)P(+e)P(+a \mid B, -e)P(+j \mid +a)P(+m \mid +a) + P(B)P(-e)P(-a \mid B, -e)P(+j \mid -a)P(+m \mid -a)
\]
Inference by Enumeration

\[ P(\text{AntiLock} | \text{observed variables}) = ? \]
Variable Elimination

- Why is inference by enumeration so slow?
  - You join up the whole joint distribution before you sum out the hidden variables
  - You end up repeating a lot of work!

- Idea: interleave joining and marginalizing!
  - Called “Variable Elimination”
  - Still NP-hard, but usually much faster than inference by enumeration

- We’ll need some new notation to define VE
**Review**

- **Joint distribution:** $P(X,Y)$
  - Entries $P(x,y)$ for all $x$, $y$
  - Sums to 1

- **Selected joint:** $P(x,Y)$
  - A slice of the joint distribution
  - Entries $P(x,y)$ for fixed $x$, all $y$
  - Sums to $P(x)$

---

**$P(T, W)$**

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

---

**$P(cold, W)$**

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Review

- Family of conditionals: $P(X | Y)$
  - Multiple conditionals
  - Entries $P(x | y)$ for all $x, y$
  - Sums to $|Y|$

- Single conditional: $P(Y | x)$
  - Entries $P(y | x)$ for fixed $x$, all $y$
  - Sums to 1

---

**$P(W|T)$**

<table>
<thead>
<tr>
<th></th>
<th>$W$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**$P(W|hot)$**

<table>
<thead>
<tr>
<th></th>
<th>$W$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

**$P(W|cold)$**

<table>
<thead>
<tr>
<th></th>
<th>$W$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Review

Specified family: \( P(y \mid X) \)
- Entries \( P(y \mid x) \) for fixed \( y \), but for all \( x \)
- Sums to … who knows!

In general, when we write \( P(Y_1 \ldots Y_N \mid X_1 \ldots X_M) \)
- It is a “factor,” a multi-dimensional array
- Its values are all \( P(y_1 \ldots y_N \mid x_1 \ldots x_M) \)
- Any assigned \( X \) or \( Y \) is a dimension missing (selected) from the array

\[
\begin{array}{ccc}
T & W & P \\
\text{hot} & \text{rain} & 0.2 \\
\text{cold} & \text{rain} & 0.6 \\
\end{array}
\]

\[
P(\text{rain} \mid T) = \begin{cases} 
P(\text{rain} \mid \text{hot}) \\
P(\text{rain} \mid \text{cold}) 
\end{cases}
\]
Inference

- Inference is expensive with enumeration

- Variable elimination:
  - Interleave joining and marginalization: Store initial results and then join with the rest
Example: Traffic Domain

- Random Variables
  - R: Raining
  - T: Traffic
  - L: Late for class!

- First query: $P(L)$

\[
P(l) = \sum_{t} \sum_{r} P(l|t)P(t|r)P(r)
\]
Variable Elimination Outline

- Maintain a set of tables called **factors**
- Initial factors are local CPTs (one per node)

| $P(R)$ | $P(T|R)$ | $P(L|T)$ |
|--------|----------|----------|
| $+r$   | $+t$     | 0.8      |
| $+r$   | $-t$     | 0.2      |
| $-r$   | $+t$     | 0.1      |
| $-r$   | $-t$     | 0.9      |

- Any known values are selected
  - E.g. if we know $L = +l$, the initial factors are

| $P(R)$ | $P(T|R)$ | $P(\!+\!l|T)$ |
|--------|----------|---------------|
| $+r$   | $+t$     | 0.8           |
| $+r$   | $-t$     | 0.2           |
| $-r$   | $+t$     | 0.1           |
| $-r$   | $-t$     | 0.9           |

- VE: Alternately join factors and eliminate variables
Operation 1: Join Factors

- First basic operation: **joining factors**
- Combining factors:
  - Just like a database join
  - Get all factors over the joining variable
  - Build a new factor over the union of the variables involved

**Example: Join on R**

\[
P(R) \times P(T|R) \rightarrow P(R, T)
\]

\[
\begin{array}{c|c}
+ r & 0.1 \\
- r & 0.9 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
+ r & + t & 0.8 \\
+ r & - t & 0.2 \\
- r & + t & 0.1 \\
- r & - t & 0.9 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
+ r & + t & 0.08 \\
+ r & - t & 0.02 \\
- r & + t & 0.09 \\
- r & - t & 0.81 \\
\end{array}
\]

- Computation for each entry: pointwise products

\[\forall r, t : P(r, t) = P(r) \cdot P(t|r)\]
Example: Multiple Joins

\[ P(R) \]

<table>
<thead>
<tr>
<th>+r</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-r</td>
<td>0.9</td>
</tr>
</tbody>
</table>

\[ P(T|R) \]

<table>
<thead>
<tr>
<th>+r</th>
<th>+t</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>-t</td>
<td>0.2</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>0.1</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>0.9</td>
</tr>
</tbody>
</table>

\[ P(L|T) \]

<table>
<thead>
<tr>
<th>+t</th>
<th>+l</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>-l</td>
<td>0.7</td>
</tr>
<tr>
<td>-t</td>
<td>+l</td>
<td>0.1</td>
</tr>
<tr>
<td>-t</td>
<td>-l</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Join R

\[ P(R,T) \]

<table>
<thead>
<tr>
<th>+r</th>
<th>+t</th>
<th>0.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>-t</td>
<td>0.02</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>0.09</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>0.81</td>
</tr>
</tbody>
</table>

\[ P(L|T) \]

<table>
<thead>
<tr>
<th>+t</th>
<th>+l</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>-l</td>
<td>0.7</td>
</tr>
<tr>
<td>-t</td>
<td>+l</td>
<td>0.1</td>
</tr>
<tr>
<td>-t</td>
<td>-l</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Example: Multiple Joins

\[
P(R, T) = \begin{array}{c|c|c}
+r & +t & 0.08 \\
+r & -t & 0.02 \\
-r & +t & 0.09 \\
-r & -t & 0.81 \\
\end{array}
\]

\[
P(L|T) = \begin{array}{c|c|c}
+t & +l & 0.3 \\
+t & -l & 0.7 \\
-t & +l & 0.1 \\
-t & -l & 0.9 \\
\end{array}
\]

\[
P(R, T, L) = \begin{array}{c|c|c|c}
+r & +t & +l & 0.024 \\
+r & +t & -l & 0.056 \\
+r & -t & +l & 0.002 \\
+r & -t & -l & 0.018 \\
-r & +t & +l & 0.027 \\
-r & +t & -l & 0.063 \\
-r & -t & +l & 0.081 \\
-r & -t & -l & 0.729 \\
\end{array}
\]
Operation 2: Eliminate

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

\[
P(R, T)\\ 
\begin{array}{ccc}
+r & +t & 0.08 \\
+r & -t & 0.02 \\
-r & +t & 0.09 \\
-r & -t & 0.81 \\
\end{array}\\ 
\text{sum } R\\ 
P(T)\\ 
\begin{array}{c}
+t & 0.17 \\
-t & 0.83 \\
\end{array}
Multiple Elimination

\[ P(R, T, L) \]

\[
\begin{array}{ccc}
+r & +t & +l & 0.024 \\
+r & +t & -l & 0.056 \\
+r & -t & +l & 0.002 \\
+r & -t & -l & 0.018 \\
-r & +t & +l & 0.027 \\
-r & +t & -l & 0.063 \\
-r & -t & +l & 0.081 \\
-r & -t & -l & 0.729 \\
\end{array}
\]

Sum out R

\[ P(T, L) \]

\[
\begin{array}{cc}
+t & +l & 0.051 \\
+t & -l & 0.119 \\
-t & +l & 0.083 \\
-t & -l & 0.747 \\
\end{array}
\]

Sum out T

\[ P(L) \]

\[
\begin{array}{c}
+l & 0.134 \\
-l & 0.886 \\
\end{array}
\]
P(L) : Marginalizing Early!

\[ P(R) \]
\[
\begin{array}{c|c}
+r & 0.1 \\
-r & 0.9 \\
\end{array}
\]

Join R

\[ P(T|R) \]
\[
\begin{array}{c|cc}
+r & +t & 0.8 \\
+r & -t & 0.2 \\
-r & +t & 0.1 \\
-r & -t & 0.9 \\
\end{array}
\]

Sum out R

\[ P(R,T) \]
\[
\begin{array}{c|cc}
+r & +t & 0.08 \\
+r & -t & 0.02 \\
-r & +t & 0.09 \\
-r & -t & 0.81 \\
\end{array}
\]

\[ P(T) \]
\[
\begin{array}{c|c}
+t & 0.17 \\
-t & 0.83 \\
\end{array}
\]

\[ P(L|T) \]
\[
\begin{array}{c|cc}
+t & +l & 0.3 \\
+t & -l & 0.7 \\
-t & +l & 0.1 \\
-t & -l & 0.9 \\
\end{array}
\]

\[ P(L|T) \]
\[
\begin{array}{c|cc}
+t & +l & 0.3 \\
+t & -l & 0.7 \\
-t & +l & 0.1 \\
-t & -l & 0.9 \\
\end{array}
\]

\[ P(L) \]
\[
\begin{array}{c|cc}
+t & +l & 0.3 \\
+t & -l & 0.7 \\
-t & +l & 0.1 \\
-t & -l & 0.9 \\
\end{array}
\]
Marginalizing Early (aka VE*)

\[ P(T, L) \]

<table>
<thead>
<tr>
<th></th>
<th>+l</th>
<th>-l</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>0.051</td>
<td>0.119</td>
</tr>
<tr>
<td>-t</td>
<td>0.083</td>
<td>0.747</td>
</tr>
</tbody>
</table>

\[ P(L) \]

<table>
<thead>
<tr>
<th></th>
<th>+l</th>
<th>-l</th>
</tr>
</thead>
<tbody>
<tr>
<td>+l</td>
<td>0.134</td>
<td>0.886</td>
</tr>
<tr>
<td>-l</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* VE is variable elimination
Traffic Domain

\[ P(L) = ? \]

### Inference by Enumeration

\[
= \sum_t \sum_r P(L|t)P(r)P(t|r)
\]

- Join on \( r \)
- Join on \( t \)
- Eliminate \( r \)
- Eliminate \( t \)

### Variable Elimination

\[
= \sum_t P(L|t) \sum_r P(r)P(t|r)
\]

- Join on \( r \)
- Eliminate \( r \)
- Join on \( t \)
- Eliminate \( t \)
Marginalizing Early

\[ P(R) \]

<table>
<thead>
<tr>
<th>+r</th>
<th>-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

\[ P(T|R) \]

<table>
<thead>
<tr>
<th>+r</th>
<th>+t</th>
<th>-r</th>
<th>-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.2</td>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

\[ P(L|T) \]

<table>
<thead>
<tr>
<th>+t</th>
<th>+l</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

\[ P(R, T) \]

<table>
<thead>
<tr>
<th>+r</th>
<th>+t</th>
<th>-r</th>
<th>-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>0.02</td>
<td>0.09</td>
<td>0.81</td>
</tr>
</tbody>
</table>

\[ P(T) \]

<table>
<thead>
<tr>
<th>+t</th>
<th>-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.17</td>
<td>0.83</td>
</tr>
</tbody>
</table>

\[ P(L|T) \]

<table>
<thead>
<tr>
<th>+t</th>
<th>+l</th>
<th>-t</th>
<th>-l</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.7</td>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Join R

Sum out R

Join T

Sum out T
### Evidence

- If evidence, start with factors that select that evidence
  - No evidence uses these initial factors:

|   | $P(R)$ | $P(T|R)$ | $P(L|T)$ |
|---|---|---|---|
| +r | 0.1 | +r | 0.8 |
| -r | 0.9 | +t | 0.2 |
| -r | 0.9 | -t | 0.1 |
|   |   | +t | 0.3 |
|   |   | +l | 0.7 |
|   |   | -t | 0.1 |
|   |   | -l | 0.9 |

- Computing $P(L | +r)$, the initial factors become:

|   | $P(\pm r)$ | $P(T | \pm r)$ | $P(L | T)$ |
|---|---|---|---|
| +r | 0.1 | +r | 0.8 |
|   |   | +t | 0.2 |
| -r | 0.9 | -t | 0.1 |
| -r | 0.9 | +t | 0.3 |
|   |   | +l | 0.7 |
| -r | 0.9 | -t | 0.1 |
|   |   | -l | 0.9 |

- We eliminate all vars other than query + evidence
Result will be a selected joint of query and evidence

- E.g. for $P(L \mid +r)$, we’d end up with:

<table>
<thead>
<tr>
<th>$P(+r, L)$</th>
<th>Normalize</th>
<th>$P(L \mid +r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+l</td>
<td>0.026</td>
</tr>
<tr>
<td>+r</td>
<td>-l</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+l</td>
<td></td>
<td>0.26</td>
</tr>
<tr>
<td>-l</td>
<td></td>
<td>0.74</td>
</tr>
</tbody>
</table>

To get our answer, just normalize this!

That’s it!
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
  - Eliminate (sum out) H

- Join all remaining factors and normalize
Variable Elimination Bayes Rule

\[ P(B) \]

\[
\begin{array}{c|c}
B & P \\
\hline
+b & 0.1 \\
-b & 0.9 \\
\end{array}
\]

\[ P(A|B) \rightarrow P(a|B) \]

\[
\begin{array}{c|c|c}
B & A & P \\
\hline
+b & +a & 0.8 \\
-b & -a & 0.2 \\
\hline
+b & -a & 0.1 \\
-b & +a & 0.9 \\
\hline
+b & -a & 0.9 \\
\end{array}
\]

\[ P(a, B) \]

\[
\begin{array}{c|c|c}
A & B & P \\
\hline
+a & +b & 0.08 \\
+a & -b & 0.09 \\
\end{array}
\]

\[ P(B|a) \]

\[
\begin{array}{c|c|c}
A & B & P \\
\hline
+a & +b & 8/17 \\
+a & -b & 9/17 \\
\end{array}
\]

Start / Select

Join on B

Normalize
Example

Query: \[ P(B|j, m) \]

Choose A

\[
\begin{align*}
P(A|B, E) & \quad P(j|A) \\
P(m|A) & \quad \sum \quad P(j, m, A|B, E)
\end{align*}
\]

\[
\begin{align*}
P(B) \quad P(E) \quad P(j, m|B, E)
\end{align*}
\]
Example

\[
P(B) \quad P(E) \quad P(j, m|B, E)
\]

Choose E

\[
P(E) \quad P(j, m, E|B) \quad \sum \quad P(j, m|B)
\]

Finish with B

\[
P(B) \quad P(j, m|B) \quad \times \quad P(j, m, B) \quad \text{Normalize} \quad P(B|j, m)
\]
Variable Elimination

\[ P(B, j, m) = \sum_{A,E} P(b, j, m, A, E) = \]

\[ \sum_{A,E} P(B)P(E)P(A \mid B, E)P(m \mid A)P(j \mid A) \]

\[ \sum_{E} P(B)P(E) \sum_{A} P(A \mid B, E)P(m \mid A)P(j \mid A) \]

\[ = \sum_{E} P(B)P(E) \sum_{A} P(m, j, A \mid B, E) \]

\[ = \sum_{E} P(B)P(E)P(m, j \mid B, E) = P(B) \sum_{E} P(m, j, E \mid B) \]

\[ = P(B)P(m, j \mid B) \]
Another Example

Query: \( P(X_3|Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) \)

Start by inserting evidence, which gives the following initial factors:

\[
p(Z)p(X_1|Z)p(X_2|Z)p(X_3|Z)p(y_1|X_1)p(y_2|X_2)p(y_3|X_3)
\]

Eliminate \( X_1 \), this introduces the factor \( f_1(Z, y_1) = \sum_{x_1} p(x_1|Z)p(y_1|x_1) \), and we are left with:

\[
p(Z)f_1(Z, y_1)p(X_2|Z)p(X_3|Z)p(y_2|X_2)p(y_3|X_3)
\]

Eliminate \( X_2 \), this introduces the factor \( f_2(Z, y_2) = \sum_{x_2} p(x_2|Z)p(y_2|x_2) \), and we are left with:

\[
p(Z)f_1(Z, y_1)f_2(Z, y_2)p(X_3|Z)p(y_3|X_3)
\]

Eliminate \( Z \), this introduces the factor \( f_3(y_1, y_2, X_3) = \sum_z p(z)f_1(z, y_1)f_2(z, y_2)p(X_3|z) \), and we are left:

\[
p(y_3|X_3), f_3(y_1, y_2, X_3)
\]

No hidden variables left. Join the remaining factors to get:

\[
f_4(y_1, y_2, y_3, X_3) = P(y_3|X_3)f_3(y_1, y_2, X_3).
\]

Normalizing over \( X_3 \) gives \( P(X_3|y_1, y_2, y_3) \).

Computational complexity critically depends on the largest factor being generated in this process. Size of factor \( \leq \) number of entries in table. In example above (assuming binary) all factors generated are of size 2 --- as they all only have one variable (\( Z, Z, \) and \( X_3 \) respectively).
Variable Elimination Ordering

- For the query $P(X_n | y_1, ..., y_n)$ work through the following two different orderings as done in previous slide: $Z, X_1, ..., X_{n-1}$ and $X_1, ..., X_{n-1}, Z$. What is the size of the maximum factor generated for each of the orderings?

  - Answer: $2^{n+1}$ versus $2^2$ (assuming binary)

- In general: the ordering can greatly affect efficiency.
The computational and space complexity of variable elimination is determined by the largest factor.

The elimination ordering can greatly affect the size of the largest factor.
- E.g., previous slide’s example $2^n$ vs. 2

Does there always exist an ordering that only results in small factors?
- No!
Exact Inference: Variable Elimination

- Remaining Issues:
  - Complexity: exponential in tree width (size of the largest factor created)
  - Best elimination ordering? NP-hard problem

- What you need to know:
  - Should be able to run it on small examples, understand the factor creation / reduction flow
  - Better than enumeration: saves time by marginalizing variables as soon as possible rather than at the end

- We have seen a special case of VE already
  - HMM Forward Inference
Variable Elimination

- Interleave joining and marginalizing
- $d^k$ entries computed for a factor over $k$ variables with domain sizes $d$
- Ordering of elimination of hidden variables can affect size of factors generated
- Worst case: running time exponential in the size of the Bayes’ net