Probability
Probability

Inference.
Probability
Inference.
Begin: Bayes Networks
Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

\[ P(hot) = P(T = hot) \]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

\[ P(\text{hot}) = P(T = \text{hot}) \]
\[ P(\text{cold}) = P(T = \text{cold}) \]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

\[
P(\text{hot}) = P(T = \text{hot})
\]
\[
P(\text{cold}) = P(T = \text{cold})
\]
\[
P(\text{rain}) = P(W = \text{rain})
\]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

- \( P(hot) = P(T = hot) \)
- \( P(cold) = P(T = cold) \)
- \( P(rain) = P(W = rain) \)

...
Probability Distributions

Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

\[
P(\text{hot}) = P(T = \text{hot})
\]

\[
P(\text{cold}) = P(T = \text{cold})
\]

\[
P(\text{rain}) = P(W = \text{rain})
\]

\[
\ldots
\]

If domains don’t overlap.
Probability Distributions

Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

- \( P(\text{hot}) = P(T = \text{hot}) \)
- \( P(\text{cold}) = P(T = \text{cold}) \)
- \( P(\text{rain}) = P(W = \text{rain}) \)
  
  \[ \cdots \]

If domains don’t overlap.

A probability (lower case value) is a single number:

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Probability Distributions

Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Shorthand notation:

\[ P(\text{hot}) = P(T = \text{hot}) \]
\[ P(\text{cold}) = P(T = \text{cold}) \]
\[ P(\text{rain}) = P(W = \text{rain}) \]

... If domains don’t overlap.

A probability (lower case value) is a single number:

\[ P(W = \text{rain}) = 0.1. \]
Probability Distributions

Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

\[
\begin{align*}
P(\text{hot}) &= P(T = \text{hot}) \\
P(\text{cold}) &= P(T = \text{cold}) \\
P(\text{rain}) &= P(W = \text{rain}) \\
\end{align*}
\]

\[\vdots\]

If domains don’t overlap.

A probability (lower case value) is a single number:

\[
P(W = \text{rain}) = 0.1.
\]

A distribution is a TABLE of probabilities of values:

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Probability Distributions

Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

- \( P(\text{hot}) = P(T = \text{hot}) \)
- \( P(\text{cold}) = P(T = \text{cold}) \)
- \( P(\text{rain}) = P(W = \text{rain}) \)

If domains don’t overlap.

A probability (lower case value) is a single number:

\[ P(W = \text{rain}) = 0.1. \]

A distribution is a TABLE of probabilities of values:

Must have: \( \forall x, P(X = x) \geq 0, \)
Probability Distributions

Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

\[ P(\text{hot}) = P(T = \text{hot}) \]
\[ P(\text{cold}) = P(T = \text{cold}) \]
\[ P(\text{rain}) = P(W = \text{rain}) \]
\[ \ldots \]
If domains don’t overlap.

A probability (lower case value) is a single number:

\[ P(W = \text{rain}) = 0.1. \]

A distribution is a TABLE of probabilities of values:

Must have: \( \forall x, P(X = x) \geq 0, \)
and \( \sum_x P(X = x) \)
Probability Distributions

Unobserved random variables have distributions:

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Shorthand notation:

- \( P(hot) = P(T = hot) \)
- \( P(cold) = P(T = cold) \)
- \( P(rain) = P(W = rain) \)
- \( \cdots \)

If domains don’t overlap.

A probability (lower case value) is a single number:

\( P(W = rain) = 0.1. \)

A distribution is a TABLE of probabilities of values:

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteor</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Must have: \( \forall x, P(X = x) \geq 0, \)

and \( \sum_x P(X = x) = 1. \)
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:
Joint Distributions

Set of random variables: \( X_1, \ldots, X_n \)
Joint Distribution:

\[
P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\
\text{or } P(x_1, x_2, \ldots, x_n).
\]
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:
\[ P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \]
or $P(x_1, x_2, \ldots, x_n)$. 

\[ P(T, W) \]
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:

$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

or $P(x_1, x_2, \ldots, x_n)$.

<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>
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$
or $P(x_1, x_2, \ldots, x_n)$.

- Must obey:

$$P(T, W)
\begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array}
$$

Same table:

$$W \times T \quad \text{hot} \quad \text{cold}
\begin{array}{|c|c|c|}
\hline
W \times T & \text{hot} & \text{cold} \\
\hline
\text{sun} & 0.4 & 0.2 \\
\text{rain} & 0.1 & 0.3 \\
\hline
\end{array}$$
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:

$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

or $P(x_1, x_2, \ldots, x_n)$.

- Must obey:

  $P(x_1, x_2, \ldots, x_n) \geq 0$

$P(T, W)$

<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>

Same table:

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

Set of random variables: \( X_1, \ldots, X_n \)

Joint Distribution:

\[
P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)
\]
or

\[
P(x_1, x_2, \ldots, x_n).
\]

Must obey:

\[
P(x_1, x_2, \ldots, x_n) \geq 0
\]

\[
\sum_{x_1, x_2, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1
\]
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

or $P(x_1, x_2, \ldots, x_n)$.

- Must obey:

  $$P(x_1, x_2, \ldots, x_n) \geq 0$$

  $$\sum_{x_1, x_2, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1$$

\[
\begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array}
\]

Same table:

\[
\begin{array}{|c|c|c|}
\hline
W \times T & \text{hot} & \text{cold} \\
\hline
\text{sun} & 0.4 & 0.2 \\
\text{rain} & 0.1 & 0.3 \\
\hline
\end{array}
\]
Joint Distributions

Set of random variables: \( X_1, \ldots, X_n \)

Joint Distribution:

\[
P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)
\]
or \( P(x_1, x_2, \ldots, x_n) \).

- Must obey:
  \[
P(x_1, x_2, \ldots, x_n) \geq 0
  \]

\[
\sum_{x_1, x_2, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1
\]

Size of distribution if \( n \) variables with domain sizes \( d \)?

\[
P(T, W)
\]

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

Same table:

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

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:

$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

or $P(x_1, x_2, \ldots, x_n)$.

- Must obey:
  
  $P(x_1, x_2, \ldots, x_n) \geq 0$

  $$\sum_{x_1, x_2, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1$$

Size of distribution if $n$ variables with domain sizes $d$? $d^n$

<table>
<thead>
<tr>
<th>$P(T, W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>cold</td>
</tr>
<tr>
<td>cold</td>
</tr>
</tbody>
</table>

Same table:

<table>
<thead>
<tr>
<th>$W \times T$</th>
<th>hot</th>
<th>cold</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:

$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

or $P(x_1, x_2, \ldots, x_n)$.

- Must obey:

$P(x_1, x_2, \ldots, x_n) \geq 0$

$\sum_{x_1, x_2, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1$

Size of distribution if $n$ variables with domain sizes $d$? $d^n$

- For all but the smallest distributions, impractical to write out!

\[
P(T, W)
\begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array}
\]

Same table:

\[
\begin{array}{|c|c|c|}
\hline
W \times T & \text{hot} & \text{cold} \\
\hline
\text{sun} & 0.4 & 0.2 \\
\text{rain} & 0.1 & 0.3 \\
\hline
\end{array}
\]
Joint Distributions

Set of random variables: $X_1, \ldots, X_n$

Joint Distribution:

$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

or $P(x_1, x_2, \ldots, x_n)$.

- Must obey:
  
  $P(x_1, x_2, \ldots, x_n) \geq 0$

  $\sum_{x_1, x_2, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1$

Size of distribution if n variables with domain sizes d? $d^n$

- For all but the smallest distributions, impractical to write out!

<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>

Same table:

<table>
<thead>
<tr>
<th>$W \times T$</th>
<th>hot</th>
<th>cold</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>rain</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Conditional Probabilities

A simple relation between joint and conditional probabilities

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

\[ P(b) = P(w = s, T = c) + P(w = r, T = c) = 0.2 + 0.3 = 0.5 \]

\[ P(w = s | T = c) = \frac{P(w = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = \frac{2}{5} \]
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

\[
P(a | b) = \frac{P(a, b)}{P(b)}.
\]

\[
P(b) = P(w = s, T = c) + P(w = r, T = c) = 0.2 + 0.3 = 0.5.
\]

\[
P(a = s | T = c) = \frac{P(a = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = \frac{2}{5}.
\]
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$. 

$$P(a|b) = \frac{P(a,b)}{P(b)}$$
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b)$$
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

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

The probability of event \( a \) given event \( b \).

Probability of a given \( b \).
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

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

The probability of event \( a \) given event \( b \).

Probability of a given \( b \).

Natural?
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Probability of a given $b$.

Natural? Yes!
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Probability of $a$ given $b$.

Natural? Yes!

<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>
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Probability of $a$ given $b$.

Natural? Yes!

<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>

$$P(W = s| T = c) = \frac{P(w=s,T=c)}{P(T=c)}.$$
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Probability of $a$ given $b$.

Natural? Yes!

<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>

$$P(W = s | T = c) = \frac{P(w=s,T=c)}{P(T=c)}.$$  

$$P(T = c) = P(W = s, T = c) + P(W = r, T = c).$$
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Probability of a given $b$.

Natural? Yes!

<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>

$$P(W = s | T = c) = \frac{P(w=s, T=c)}{P(T=c)}.$$  

$$P(T = c) = P(W = s, T = c) + P(W = r, T = c)$$  

$$= 0.2 + 0.3$$
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$ P(a|b) = \frac{P(a,b)}{P(b)} $$

Probability of $a$ given $b$.

Natural? Yes!

<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>

$$ P(W = s|T = c) = \frac{P(w=s,T=c)}{P(T=c)} = 0.2 + 0.3 = 0.5 $$
Conditional Probabilities

A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

$$P(a \mid b) = \frac{P(a, b)}{P(b)}$$

The probability of event $a$ given event $b$.

Probability of a given $b$.

Natural? Yes!

<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>

$$P(W = s \mid T = c) = \frac{P(w=s, T=c)}{P(T=c)}.$$  

$$P(T = c) = P(W = s, T = c) + P(W = r, T = c) = 0.2 + 0.3 = 0.5$$  

$$P(W = s \mid T = c) = \frac{P(w=s, T=c)}{P(T=c)}.$$
A simple relation between joint and conditional probabilities

- In fact, this is taken as the definition of a conditional probability

The probability of event $a$ given event $b$.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Probability of $a$ given $b$.

Natural? Yes!

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>$W$</td>
<td>$P$</td>
</tr>
<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>

$$P(W = s | T = c) = \frac{P(w = s, T = c)}{P(T = c)}.$$

$$P(T = c) = P(W = s, T = c) + P(W = r, T = c) = 0.2 + 0.3 = 0.5$$

$$P(W = s | T = c) = \frac{P(w = s, T = c)}{P(T = c)} = \frac{2}{5} = 2/5.$$
Conditional Distributions

Conditional distributions are probability distributions over some variables given fixed values of others.
Conditional Distributions

Conditional distributions are probability distributions over some variables given fixed values of others.

Joint Distribution

<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>
Conditional distributions are probability distributions over some variables given fixed values of others.

### Joint Distribution

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>W</td>
<td>P</td>
</tr>
<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>
Conditional Distributions

Conditional distributions are probability distributions over some variables given fixed values of others.

<table>
<thead>
<tr>
<th>Joint Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>cold</td>
</tr>
<tr>
<td>cold</td>
</tr>
</tbody>
</table>

Conditional Distributions

\[ P(W|T = \text{hot}) \]
Conditional Distributions

Conditional distributions are probability distributions over some variables given fixed values of others.

**Joint Distribution**

<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>

**Conditional Distributions**

\[ P(W|T=\text{hot}) \]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>cold</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Conditional Distributions

Conditional distributions are probability distributions over some variables given fixed values of others.

<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>

Conditional Distributions

\[ P(W|T = \text{hot}) \]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>cold</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\[ P(W|T = \text{cold}) \]
Conditional Distributions

Conditional distributions are probability distributions over some variables given fixed values of others.

### Joint Distribution

<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>

### Conditional Distributions

- $P(W | T = \text{hot})$
  - $W$ | $P$
  - sun: 0.8
  - cold: 0.2

- $P(W | T = \text{cold})$
  - $W$ | $P$
  - sun: 0.4
  - cold: 0.6
Normalization Trick

**Table**

<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>

Why does this work? Sum of selection is \( P(\text{evidence}) \)!

\[
P(x_1 | x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{\sum x_1 P(x_1, x_2)}{P(x_2)}
\]
Normalization Trick

<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>

SELECT the joint probabilities matching the evidence

Why does this work? Sum of selection is $P(\text{evidence})$! ($P(T=c)$, here)

$$P(x_1 | x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{\sum x_1 P(x_1, x_2)}{\sum x_2 P(x_1, x_2)}$$
## Normalization Trick

<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>

SELECT the joint probabilities matching the evidence:

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>cold</td>
<td>sun</td>
<td></td>
</tr>
</tbody>
</table>

NORMALIZE the selection (make it sum to one):

<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>

\[
P(x_1 | x_2) = \frac{Pr(x_1, x_2)}{\sum x_1 Pr(x_1, x_2)}
\]

Why does this work? Sum of selection is \( P(\text{evidence}) \)!
Normalization Trick

<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>

SELECT the joint probabilities matching the evidence

<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></td>
</tr>
</tbody>
</table>
Normalization Trick

<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>

SELECT the joint probabilities matching the evidence

<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>

Why does this work? Sum of selection is $P(\text{evidence})!$ ($P(T=c)$, here)

$$P(x_1 | x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{\sum x_1 P(x_1, x_2)}{P(x_2)}$$
Normalization Trick

<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>

SELECT the joint probabilities matching the evidence

<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>

NORMALIZE the selection (make it sum to one)
Normalization Trick

<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>

SELECT the joint probabilities matching the evidence

<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>

• NORMALIZE the selection (make it sum to one)

Why does this work? Sum of selection is $P(evidence) = \sum_x P(x_1, x_2)$.
**Normalization Trick**

**SELECT** the joint probabilities matching the evidence

<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>

**NORMALIZE** the selection (make it sum to one)

<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>

Why does this work? Sum of selection is $P(\text{evidence})$!

$$P(x_1 | x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{\sum_{x_1} P(x_1, x_2)}{P(x_2)}$$
### Normalization Trick

<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>

**SELECT the joint probabilities matching the evidence**

<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>

**NORMALIZE the selection (make it sum to one)**

<table>
<thead>
<tr>
<th>T</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></td>
</tr>
</tbody>
</table>

Why does this work? Sum of selection is \( P(\text{evidence}) \), here:

\[
P(x_1 | x_2) = \frac{P(x_1, x_2)}{P(x_2)} = P(x_1, x_2) \sum_{x_1} P(x_1, x_2)
\]
Normalization Trick

<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>

**SELECT** the joint probabilities matching the evidence

<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>

**NORMALIZE** the selection (make it sum to one)

<table>
<thead>
<tr>
<th>T</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>
Normalization Trick

<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>

SELECT the joint probabilities matching the evidence

<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>

NORMALIZE the selection (make it sum to one)

<table>
<thead>
<tr>
<th>T</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>

Why does this work? Sum of selection is P(evidence)! (P(T=c), here)
Normalization Trick

<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>

SELECT the joint probabilities matching the evidence

<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>

NORMALIZE the selection (make it sum to one)

<table>
<thead>
<tr>
<th>T</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>

Why does this work? Sum of selection is $P(\text{evidence})!$ ($P(T=c)$, here)

$$P(x_1|x_2) = \frac{Pr(x_1,x_2)}{Pr(x_2)} = \frac{Pr(x_1,x_2)}{\sum_{x_1} Pr(x_1,x_2)}$$
**Quiz: Normalization Trick**

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Select the joint probabilities matching the evidence. Then normalize the selection (make it sum to one).
Quiz: Normalization Trick

SELECT the joint probabilities matching the evidence

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Quiz: Normalization Trick

SELECT the joint probabilities matching the evidence

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

NORMALIZE the selection (make it sum to one)

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>
**Quiz: Normalization Trick**

SELECT the joint probabilities matching the evidence

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

NORMALIZE the selection (make it sum to one)

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>
### Quiz: Normalization Trick

**SELECT** the joint probabilities matching the evidence

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**NORMALIZE** the selection (make it sum to one)

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Quiz: Normalization Trick

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

SELECT the joint probabilities matching the evidence

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

NORMALIZE the selection (make it sum to one)

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.8</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Quiz: Normalization Trick

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**SELECT** the joint probabilities matching the evidence

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

- **NORMALIZE** the selection (make it sum to one)

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.8</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.2</td>
</tr>
</tbody>
</table>
(Dictionary) To bring or restore to a normal condition (sum to one).
To Normalize

(Dictionary) To bring or restore to a normal condition (sum to one).

Procedure:
- Step 1: Compute $Z = \text{sum over all entries}$
To Normalize

(Dictionary) To bring or restore to a normal condition (sum to one).

Procedure:
- Step 1: Compute $Z = \text{sum over all entries}$
- Step 2: Divide every entry by $Z$
(Dictionary) To bring or restore to a normal condition (sum to one).

Procedure:
- Step 1: Compute $Z = \text{sum over all entries}$
- Step 2: Divide every entry by $Z$

Example 1:

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>
To Normalize

(Dictionary) To bring or restore to a normal condition (sum to one).

Procedure:
- Step 1: Compute \( Z = \text{sum over all entries} \)
- Step 2: Divide every entry by \( Z \)

Example 1:

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

Normalize
To Normalize

(Dictionary) To bring or restore to a normal condition (sum to one).

Procedure:
- Step 1: Compute $Z = \text{sum over all entries}$
- Step 2: Divide every entry by $Z$

Example 1:

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

Normalize

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>rain</td>
<td>0.6</td>
</tr>
</tbody>
</table>
To Normalize

(Dictionary) To bring or restore to a normal condition (sum to one).

Procedure:
- Step 1: Compute \( Z = \text{sum over all entries} \)
- Step 2: Divide every entry by \( Z \)

Example 1:

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

Normalize:

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

Example 2:

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>20</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>5</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>10</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>15</td>
</tr>
</tbody>
</table>

Normalize
To Normalize

(Dictionary) To bring or restore to a normal condition (sum to one).

Procedure:
- Step 1: Compute $Z = \text{sum over all entries}$
- Step 2: Divide every entry by $Z$

Example 1:

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

Normalize

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

Example 2:

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>20</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>5</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>10</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>15</td>
</tr>
</tbody>
</table>

Normalize

<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>
Probabilistic Inference

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

We generally compute conditional probabilities:

\[ P(\text{on time} | \text{no reported accidents}) = 0.90 \]

Represent the agent's beliefs given the evidence:

Probabilities change with new evidence:

\[ P(\text{on time} | \text{no accidents}, a_m) = 0.95 \]

\[ P(\text{on time} | \text{no accidents}, a_m, \text{raining}) = 0.80 \]

Observing new evidence causes beliefs to be updated.
Probabilistic Inference

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

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

We generally compute conditional probabilities

- \( P(\text{on time}|\text{no reported accidents}) = 0.90 \)
Probabilistic Inference

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

We generally compute conditional probabilities

- \( P(\text{on time}|\text{no reported accidents}) = 0.90 \)
- Represent the agent’s beliefs given the evidence
Probabilistic Inference

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

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

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

We generally compute conditional probabilities

- $P(\text{on time}|\text{no reported accidents}) = 0.90$
- Represent the agent’s beliefs given the evidence

Probabilities change with new evidence:

- $P(\text{on time}|\text{no accidents, 5 a.m.}) = 0.95$
Probabilistic Inference

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

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

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

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

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

Probabilities change with new evidence:
- \( P(\text{on time}|\text{no accidents}, 5 \text{ a.m.}) = 0.95 \)
- \( P(\text{on time}|\text{no accidents}, 5 \text{ a.m.}, \text{raining}) = 0.80 \)
- Observing new evidence causes beliefs to be updated
Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

We generally compute conditional probabilities
- $P(on\ time|no\ reported\ accidents) = 0.90$
- Represent the agent’s beliefs given the evidence

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

General case:

- Evidence variables:
  \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
Inference by Enumeration

General case:
- Evidence variables: \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)

* Works fine with multiple query variables, too

Step 1: Entries consistent with evidence

Step 2: Sum out \( H \)

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

Step 3: Normalize

\[
Z = \sum_q P(q, e_1, \ldots, e_k)
\]

\[
P(Q| e_1, \ldots, e_k) = \frac{1}{Z} P(Q, e_1, \ldots, e_k)
\]
Inference by Enumeration

General case:
- Evidence variables: \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)
- Hidden variables: \( H_1, \ldots, H_r. \)
Inference by Enumeration

General case:
- Evidence variables: 
  \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)
- Hidden variables: \( H_1, \ldots, H_r \).
Inference by Enumeration

General case:

- Evidence variables:
  \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)
- Hidden variables: \( H_1, \ldots, H_r \).

We Want:

\[
\text{We Want:}
\]

\[
\text{We Want:}
\]

\[
P(Q | e_1, \ldots, e_k).
\]

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

\[
Z = \sum_q P(q, e_1, \ldots, e_k)
\]

\[
P(Q | e_1, \ldots, e_k) = \frac{1}{Z} P(Q, e_1, \ldots, e_k).
\]
Inference by Enumeration

General case:
- Evidence variables: $\bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k$
- Query* variable: $Q$
- Hidden variables: $H_1, \ldots, H_r$.

We Want:
$$P(Q|e_1, \ldots, e_k).$$
Inference by Enumeration

General case:
- Evidence variables: $\tilde{E}_1, \ldots, \tilde{E}_k = e_1, \ldots, e_k$
- Query* variable: $Q$
- Hidden variables: $H_1, \ldots, H_r$

We Want:

$$P(Q|e_1, \ldots, e_k).$$

* Works fine with multiple query variables, too
Inference by Enumeration

General case:
- Evidence variables: $\tilde{E}_1, \ldots, \tilde{E}_k = e_1, \ldots, e_k$
- Query* variable: $Q$
- Hidden variables: $H_1, \ldots, H_r$

We Want:
$P(Q|e_1, \ldots, e_k)$.
* Works fine with multiple query variables, too

Step 1:
Entries consistent with evidence
Inference by Enumeration

General case:
- Evidence variables: $\bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k$
- Query* variable: $Q$
- Hidden variables: $H_1, \ldots, H_r$.

We Want:

$P(Q|e_1, \ldots, e_k)$.

* Works fine with multiple query variables, too

Step 1:
Entries consistent with evidence

<table>
<thead>
<tr>
<th>$x$</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>0.05</td>
</tr>
<tr>
<td>-2</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Step 2: Sum out H

Step 3: Normalize
Inference by Enumeration

General case:
- Evidence variables: \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)
- Hidden variables: \( H_1, \ldots, H_r \).

We Want:
\[
P(Q|e_1, \ldots, e_k).
\]
* Works fine with multiple query variables, too

Step 1:
Entries consistent with evidence

<table>
<thead>
<tr>
<th>( n )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>0.05</td>
</tr>
<tr>
<td>-2</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Step 2: Sum out H
Inference by Enumeration

General case:
- Evidence variables: \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)
- Hidden variables: \( H_1, \ldots, H_r \).

We Want:

\[ P(Q|e_1, \ldots, e_k). \]

* Works fine with multiple query variables, too

Step 1:
Entries consistent with evidence

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

Step 2: Sum out H
Inference by Enumeration

General case:
- Evidence variables: \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)
- Hidden variables: \( H_1, \ldots, H_r \).

We Want:
\[
P(Q|e_1, \ldots, e_k).
\]
* Works fine with multiple query variables, too

Step 1:
Entries consistent with evidence

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

Step 2: Sum out \( H \)

\[
P(Q) = \frac{\sum_{q, e_1, \ldots, e_k} P(Q, e_1, \ldots, e_k)}{Z}
\]

Step 3: Normalize
**Inference by Enumeration**

**General case:**
- Evidence variables: \( \bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k \)
- Query* variable: \( Q \)
- Hidden variables: \( H_1, \ldots, H_r \).

**We Want:**
\[
P(Q|e_1, \ldots, e_k).
\]
* Works fine with multiple query variables, too

**Step 1:**
Entries consistent with evidence

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>0.08</td>
</tr>
<tr>
<td>-2</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Step 2:** Sum out \( H \)

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

**Step 3:** Normalize

\[
Z = \sum_q P(q, e_1, \ldots, e_k)
\]
Inference by Enumeration

General case:
- Evidence variables: $\bar{E}_1, \ldots, \bar{E}_k = e_1, \ldots, e_k$
- Query* variable: $Q$
- Hidden variables: $H_1, \ldots, H_r$.

We Want:
$$P(Q|e_1, \ldots, e_k).$$
* Works fine with multiple query variables, too

Step 1:
Entries consistent with evidence

Step 2: Sum out $H$

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

Step 3: Normalize

$$Z = \sum_q P(q, e_1, \ldots, e_k)$$
$$P(Q|e_1, \ldots, e_k) = \frac{1}{Z} P(Q, e_1, \ldots, e_k)$$
Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>
### Inference by Enumeration

**Table:**

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

**Question:**

\[
P(W = \text{sun})?
\]

**Solution:**

1. **Sum out S, T.**
   - Get 0.65

2. **Sum out T over S = winter.**
   - Get 0.25 for W = sun.
   - Get 0.25 for W = rain.

3. **Normalize:**
   - \( \frac{1}{2} \)

4. **Calculate:**
   - Get 0.10 for W = sun.
   - Get 0.05 for W = rain.

5. **Normalize:**
   - \( \frac{2}{3} \),

**Final answer:**

\[
P(W = \text{sun}) = \frac{2}{3}.
\]
Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\[ P(W = \text{sun})? \]
Sum out S, T.
Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\[ P(W = \text{sun})? \]
Sum out S, T.
Get 0.65
Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\[
P(W = \text{sun})? \\
\text{Sum out S, T.} \\
\text{Get 0.65}
\]

\[
P(W = \text{sun}|\text{winter})? \\
\]

\[
\text{Sum out T over } S=\text{winter.} \\
\text{Get .25 for } W=\text{sun}. \\
\text{Get .25 for } W=\text{rain}. \\
\text{Normalize: } 1/2 \\
P(W = \text{sun}|\text{winter}, \text{hot})? \\
\text{Get .10 for } W=\text{sun}. \\
\text{Get .05 for } W=\text{rain}. \\
\text{Normalize: } 2/3.
\]
Inference by Enumeration

\[ P(W = \text{sun})? \]
Sum out S, T.
Get 0.65

\[ P(W = \text{sun}|\text{winter})? \]
Sum out T over S=winter.
Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\[ P(W = \text{sun})? \]

Sum out S, T.

Get 0.65

\[ P(W = \text{sun}|\text{winter})? \]

Sum out T over S=winter.

Get .25 for W=sun.
Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

$P(W = \text{sun})$?
Sum out S, T.
Get 0.65

$P(W = \text{sun}|\text{winter})$?
Sum out T over S=winter.
Get .25 for W=sun.
Get .25 for W=rain.
Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\[ P(W = \text{sun})? \]
Sum out S, T.
Get 0.65

\[ P(W = \text{sun}|\text{winter})? \]
Sum out T over S=winter.
Get .25 for W=sun.
Get .25 for W=rain.
Normalize: 1/2
### Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

**$P(W = \text{sun})$?**
- Sum out $S, T$.
- Get 0.65

**$P(W = \text{sun}|winter)$?**
- Sum out $T$ over $S=\text{winter}$.
  - Get .25 for $W=\text{sun}$.
  - Get .25 for $W=\text{rain}$.
- Normalize: $1/2$

**$P(W = \text{sun}|winter, hot)$?**
### Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

**P(W = sun)?**
- Sum out S, T.
- Get 0.65

**P(W = sun|winter)?**
- Sum out T over S=winter.
  - Get .25 for W=sun.
  - Get .25 for W=rain.
- Normalize: 1/2

**P(W = sun|winter, hot)?**
- Get .10 for W=sun.
Inference by Enumeration

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>T</td>
<td>W</td>
<td>P</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

\[ P(W = \text{sun})? \]
Sum out S, T.
Get 0.65

\[ P(W = \text{sun} | \text{winter})? \]
Sum out T over S=winter.
Get .25 for W=sun.
Get .25 for W=rain.
Normalize: 1/2

\[ P(W = \text{sun} | \text{winter, hot})? \]
Get .10 for W=sun.
Get .05 for W=rain.
Inference by Enumeration

\[
P(W = \text{sun})? \\
\text{Sum out S, T.} \\
\text{Get 0.65}
\]

\[
P(W = \text{sun}|\text{winter})? \\
\text{Sum out T over S=winter.} \\
\text{Get .25 for W=sun.} \\
\text{Get .25 for W=rain.} \\
\text{Normalize: 1/2}
\]

\[
P(W = \text{sun}|\text{winter, hot})? \\
\text{Get .10 for W=sun.} \\
\text{Get .05 for W=rain.} \\
\text{Normalize: 2/3.}
\]

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>T</td>
<td>W</td>
<td>P</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>
### Inference by Enumeration

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

- **$P(W = \text{sun})$?**
  - Sum out S, T.
  - Get 0.65

- **$P(W = \text{sun}|\text{winter})$?**
  - Sum out T over S=winter.
  - Get .25 for $W=\text{sun}$.
  - Get .25 for $W=\text{rain}$.
  - Normalize: $\frac{1}{2}$

- **$P(W = \text{sun}|\text{winter, hot})$?**
  - Get .10 for $W=\text{sun}$.
  - Get .05 for $W=\text{rain}$.
  - Normalize: $\frac{2}{3}$.  

The table above shows the probabilities of different weather conditions in summer and winter. The probabilities are calculated by summing over the seasons or the weather conditions as required.
Inference by Enumeration

Obvious problems:

Worst-case time complexity $O(d^n)$

Space complexity $O(d^n)$ to store the joint distribution
Inference by Enumeration

Obvious problems:
- Worst-case time complexity $O(d^n)$
Inference by Enumeration

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

Obvious problems:
- Worst-case time complexity $O(d^n)$
- Space complexity $O(d^n)$ to store the joint distribution
The Product Rule

Sometimes have conditional distributions but want the joint
The Product Rule

Sometimes have conditional distributions but want the joint

\[ P(x|y) = \frac{P(x,y)}{P(y)} \]
The Product Rule

Sometimes have conditional distributions but want the joint

\[ P(x|y) = \frac{P(x,y)}{P(y)} \iff P(x, y) = P(x|y)P(y) \]
The Product Rule

Sometimes have conditional distributions but want the joint

\[ P(x|y) = \frac{P(x,y)}{P(y)} \iff P(x, y) = P(x|y)P(y) \]
The Product Rule
The Product Rule

Example:

\[
P(W)
\]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>
The Product Rule

Example:

| \( P(W) \) | \( P(W) \) | \( P(D|W) \) |
|-----------|-----------|-----------|
| W         | P         | D         |
| sun       | 0.8       | wet       |
| rain      | 0.2       | dry       |

\[
P(D|W) = P(wet | sun) \cdot P(wet) + P(dry | sun) \cdot P(dry) + P(wet | rain) \cdot P(wet) + P(dry | rain) \cdot P(dry)
\]
The Product Rule

Example:

| $P(W)$ | $P(D|W)$ | $P(D, W)$ |
|--------|----------|-----------|
| W      | D        | W         | P         |
| sun    | wet      | sun       | 0.1       |
|        | dry      | sun       | 0.9       |
| rain   | wet      | rain      | 0.7       |
|        | dry      | rain      | 0.3       |

$P(W)$

$P(D|W)$

$P(D, W)$
The Product Rule

Example:

\[ P(W) \]

\[
\begin{array}{|c|c|}
\hline
W & P \\
\hline
\text{sun} & 0.8 \\
\text{rain} & 0.2 \\
\hline
\end{array}
\]

\[ P(D|W) \]

\[
\begin{array}{|c|c|c|}
\hline
D & W & P \\
\hline
\text{wet} & \text{sun} & 0.1 \\
\text{dry} & \text{sun} & 0.9 \\
\text{wet} & \text{rain} & 0.7 \\
\text{dry} & \text{rain} & 0.3 \\
\hline
\end{array}
\]

\[ P(D, W) \]

\[
\begin{array}{|c|c|c|}
\hline
D & W & P \\
\hline
\text{wet} & \text{sun} & \\
\hline
\end{array}
\]
The Product Rule

Example:

| $P(W)$ | $P(D|W)$ | $P(D, W)$ |
|--------|----------|------------|
| W | P | D | W | P |
| sun | 0.8 | wet | sun | 0.1 |
| rain | 0.2 | dry | sun | 0.9 |
| wet | rain | 0.7 |
| dry | rain | 0.3 |
| wet | sun | 0.08 |
| dry | sun | 0.06 |
The Product Rule

Example:

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

| $P(D|W)$ | D       | W       | P  |
|----------|---------|---------|----|
| wet      | sun     | 0.1     |
| dry      | sun     | 0.9     |
| wet      | rain    | 0.7     |
| dry      | rain    | 0.3     |

<table>
<thead>
<tr>
<th>$P(D, W)$</th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Product Rule

Example:

<table>
<thead>
<tr>
<th>P(W)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>P</td>
</tr>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

| P(D|W)  | D   | W   | P   |
|-------|-----|-----|-----|
| wet   | sun | 0.1 |
| dry   | sun | 0.9 |
| wet   | rain| 0.7 |
| dry   | rain| 0.3 |

<table>
<thead>
<tr>
<th>P(D,W)</th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Product Rule

Example:

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

$P(D|W)$

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

$P(D, W)$

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.08</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.72</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.14</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.06</td>
</tr>
</tbody>
</table>
The Chain Rule

More generally, can always write any joint distribution as an incremental product of conditional distributions
The Chain Rule

More generally, can always write any joint distribution as an incremental product of conditional distributions

\[ P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \]
The Chain Rule

More generally, can always write any joint distribution as an incremental product of conditional distributions

- \( P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \)
- \( P(x_1, \ldots, x_n) = \prod_i P(x_i|x_1, \ldots, x_{i-1}) \)
The Chain Rule

More generally, can always write any joint distribution as an incremental product of conditional distributions

\[ P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \]

\[ P(x_1, \ldots, x_n) = \prod_{i} P(x_i|x_1, \ldots, x_{i-1}) \]

Why is this always true?
The Chain Rule

More generally, can always write any joint distribution as an incremental product of conditional distributions

- $P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$
- $P(x_1, \ldots, x_n) = \prod_i P(x_i|x_1, \ldots, x_{i-1})$

Why is this always true?

   Induction.
Bayes Rule
Bayes’ Rule

Two ways to factor a joint distribution over two variables:
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) \]
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?
- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Foundation of many systems we'll see later (e.g. ASR, MT)
- In the running for most important AI equation!

That's my rule!
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?
- Lets us build one conditional from its reverse
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?
- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?
- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Foundation of many systems we’ll see later (e.g. ASR, MT)
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?

- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Foundation of many systems we’ll see later (e.g. ASR, MT)
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?
- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Foundation of many systems we’ll see later (e.g. ASR, MT)

In the running for most important AI equation!
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?
- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Foundation of many systems we’ll see later (e.g. ASR, MT)

In the running for most important AI equation!
Bayes’ Rule

Two ways to factor a joint distribution over two variables:

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

Dividing, we get:

\[ P(x|y) = \frac{P(y|x)}{P(y)} P(x) \]

Why is this at all helpful?
- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Foundation of many systems we’ll see later (e.g. ASR, MT)

In the running for most important AI equation!

That’s my rule!
Example: Diagnostic probability from causal probability:

\[
P(\text{cause} \mid \text{effect}) = P(\text{effect} \mid \text{cause}) P(\text{cause}) P(\text{effect})
\]

Example: \(t\)M: meningitis, \(t\)S: stiff neck

Given:

\[
P(+\text{m}) = 0.0001, \quad P(+\text{s} \mid +\text{m}) = 0.8, \quad P(+\text{s} \mid -\text{m}) = 0.01
\]

Bayes' Rule

\[
P(+\text{m} \mid +\text{s}) = \frac{P(+\text{s} \mid +\text{m}) P(+\text{m})}{P(+\text{s} \mid +\text{m}) P(+\text{m}) + P(+\text{s} \mid -\text{m}) P(-\text{m})}
\]

\[
= \frac{(0.8)(0.0001)}{(0.8)(0.0001) + (0.01)(0.9999)}
\]

Note: posterior probability of meningitis still very small

Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause | effect) = \frac{P(effect | cause)P(cause)}{P(effect)} \]

Example: 
M: meningitis, S: stiff neck

Given: 
\[ P(+)m = 0.0001, \quad P(+)s | +m = 0.8, \quad P(+)s | -m = 0.01 \]

Bayes’ Rule

\[ P(+)m | (+)s = \frac{P(+)s | (+)m P(+)m}{P(+)s} = \frac{P(+)s | (+)m P(+)m}{P(+)s | (+)m P(+)m + P(+)s | -m Pr(-m)} = (0.8)(0.0001) \]

Note: posterior probability of meningitis still very small

Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|\text{effect}) = \frac{P(\text{effect}|cause)P(cause)}{P(\text{effect})} \]

Example:
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(cause|effect)} \]

Example:
- M: meningitis, S: stiff neck
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:
- M: meningitis, S: stiff neck
- Givens.

Note: posterior probability of meningitis still very small
Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- M: meningitis, S: stiff neck
- Givens.

\[ P(+m) = 0.0001, P(+s|+m) = 0.8, P(+s|-m) = 0.01 \]
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- M: meningitis, S: stiff neck
- Givens.

\[ P(+m) = 0.0001, P(+s|m) = 0.8, P(+s|\neg m) = 0.01 \]

Bayes Rule

Note: posterior probability of meningitis still very small

Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- M: meningitis, S: stiff neck
- Givens.

\[ P(+m) = 0.0001, \ P(+s|+m) = 0.8, \ P(+s|-m) = 0.01 \]

Bayes Rule

\[ P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} \]

Note: posterior probability of meningitis still very small
Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- M: meningitis, S: stiff neck
- Givens.
  \[ P(+m) = 0.0001, P(+s|m) = 0.8, P(+s|m) = 0.01 \]
- Bayes Rule

\[ P(+m|+s) = \frac{P(+s|m)P(+m)}{P(+s)} = \frac{P(+s|m)P(+m)}{P(+s|m)P(+m)+P(+s|m)Pr(-m)} \]

Note: posterior probability of meningitis still very small

Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- M: meningitis, S: stiff neck
- Givens.

\[ P(+m) = 0.0001, P(+s|m) = 0.8, P(+s|\neg m) = 0.01 \]

- Bayes Rule

\[
P(+m|+s) = \frac{P(+s|m)P(+m)}{P(+s)} = \frac{P(+s|m)P(+m)}{P(+s|m)P(+m) + P(+s|\neg m)Pr(\neg m)} \]

\[= \frac{(0.8)(0.0001)}{(0.8)(0.0001)+(0.01)(0.9999)} \]

Note: posterior probability of meningitis still very small

Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- **M**: meningitis, **S**: stiff neck
- **Givens.**

\[ P(+m) = 0.0001, P(+s|m) = 0.8, P(+s|\neg m) = 0.01 \]

- **Bayes Rule**

\[
P(+m|+s) = \frac{P(+s|m)P(+m)}{P(+s)} = \frac{P(+s|m)P(+m)}{P(+s|m)P(+m) + P(+s|\neg m)Pr(\neg m)}
\]

\[
= \frac{(0.8)(0.0001)}{(0.8)(0.0001) + (0.01)(0.9999)} = \text{small}
\]
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:
- M: meningitis, S: stiff neck
- Givens.
  \[ P(+m) = 0.0001, P(+s|+m) = 0.8, P(+s|-m) = 0.01 \]
- Bayes Rule
  \[ P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m)+P(+s|-m)Pr(-m)} = \frac{0.8(0.0001)}{(0.8)(0.0001)+(0.01)(0.9999)} = small \]
- Note: posterior probability of meningitis still very small

Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- M: meningitis, S: stiff neck
- Givens.

\[ P(+m) = 0.0001, P(+s|m) = 0.8, P(+s|m) = 0.01 \]

- Bayes Rule

\[ P(+m|+s) = \frac{P(+s|m)P(+m)}{P(+s)} = \frac{P(+s|m)P(+m)}{P(+s|m)P(+m)+P(+s|\neg m)\Pr(\neg m)} \]

\[ = \frac{(0.8)(0.0001)}{(0.8)(0.0001)+(0.01)(0.9999)} = small \]

- Note: posterior probability of meningitis still very small
- Note: you should still get stiff necks checked out! Why?
Inference with Bayes’ Rule

Example: Diagnostic probability from causal probability:

\[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

Example:

- M: meningitis, S: stiff neck
- Givens.

\[ P(+m) = 0.0001, P(+s|m) = 0.8, P(+s|\neg m) = 0.01 \]

- Bayes Rule

\[ P(+m|+s) = \frac{P(+s|m)P(+m)}{P(+s)} = \frac{P(+s|m)P(+m)}{P(+s|m)P(+m) + P(+s|\neg m)Pr(\neg m)} = \text{small} \]

- Note: posterior probability of meningitis still very small
- Note: you should still get stiff necks checked out! Why?
Quiz: Bayes’ Rule

Given:

<table>
<thead>
<tr>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Pr(D—W)

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

What is $P(\text{sun} | \text{dry})$?

$P(\text{sun} | \text{dry}) = \frac{P(\text{dry} | \text{sun}) \times P(\text{sun})}{P(\text{dry})}$

$= \frac{0.72 \times 0.8}{0.72 + 0.06}$

$= \frac{0.576}{0.78}$

$= 0.7256$
Quiz: Bayes’ Rule

Given:

<table>
<thead>
<tr>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\[ P(sun|dry) \]

Pr(D—W)

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

What is \[ P(sun|dry) \]?
Quiz: Bayes’ Rule

Given:

<table>
<thead>
<tr>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Pr(D—W)

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

What is \( P(\text{sun}|\text{dry})? \)

\[
= P(\text{dry}|\text{sun}) \times P(\text{sun}) / P(\text{dry}) = \frac{72}{78} = 12 / 13
\]
Quiz: Bayes’ Rule

Given:

<table>
<thead>
<tr>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pr(D—W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>wet</td>
</tr>
<tr>
<td>dry</td>
</tr>
<tr>
<td>wet</td>
</tr>
<tr>
<td>dry</td>
</tr>
</tbody>
</table>

What is \( P(\text{sun}|\text{dry}) \)?

\[
P(\text{sun}|\text{dry}) = \frac{P(\text{dry}|\text{sun}) \times P(\text{sun})}{P(\text{dry})} = \frac{72}{78} = \frac{12}{13}
\]

\[
P(\text{dry}) = Pr(\text{dry}, \text{sun}) + Pr(\text{dry}, \text{rain})
\]
Quiz: Bayes’ Rule

Given:

<table>
<thead>
<tr>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Pr(D—W)

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

What is $P(\text{sun}|\text{dry})$?

$$= P(\text{dry}|\text{sun}) \times P(\text{sun})/P(\text{dry}) = \frac{72}{78} = \frac{12}{13}$$

$P(\text{dry}) = Pr(\text{dry}, \text{sun}) + Pr(\text{dry}, \text{rain})$

$$= Pr(\text{dry}|\text{sun}) \times Pr(\text{sun}) + Pr(\text{dry}|\text{rain}) Pr(\text{rain}) = 0.72 + 0.06 = 0.78.$$
Quiz: Bayes’ Rule

Given:

$$\begin{array}{|c|c|}
\hline
R & P \\
\hline
\text{sun} & 0.8 \\
\text{rain} & 0.2 \\
\hline
\end{array}$$

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

What is $P(\text{sun}|\text{dry})$?

$$= P(\text{dry}|\text{sun}) \times P(\text{sun}) / P(\text{dry}) = \frac{72}{78} = 12/13$$

$P(\text{dry}) = Pr(\text{dry}, \text{sun}) + Pr(\text{dry}, \text{rain})$

$$= Pr(\text{dry}|\text{sun}) \times Pr(\text{sun}) + Pr(\text{dry}|\text{rain}) Pr(\text{rain}) = 0.72 + 0.06 = 0.78.$$
Let's say we have two distributions:

- Prior distribution over ghost location: $P(G)$
- Sensor model: $P(R|G)$

Given: we know what our sensors do

- $R =$ color measured at $(1,1)$
- E.g. $P(R = \text{yellow} | G = (1,1)) = 0.1$. 

We can calculate the posterior distribution $P(G|R)$ over ghost locations given a reading using Bayes' rule:

$$P(G|R) \propto P(R|G) P(G)$$
Let’s say we have two distributions:
- Prior distribution over ghost location: $P(G)$
Let’s say we have two distributions:

- Prior distribution over ghost location: \( P(G) \)
- Let’s say this is uniform
Let’s say we have two distributions:

- Prior distribution over ghost location: $P(G)$
- Let’s say this is uniform
- Sensor model: $P(R|G)$
Let’s say we have two distributions:

- Prior distribution over ghost location: $P(G)$
- Let’s say this is uniform
- Sensor model: $P(R|G)$
- Given: we know what our sensors do
Let’s say we have two distributions:
- Prior distribution over ghost location: $P(G)$
- Let’s say this is uniform
- Sensor model: $P(R|G)$
- Given: we know what our sensors do
- $R =$ color measured at (1,1)
Let’s say we have two distributions:
- Prior distribution over ghost location: \( P(G) \)
- Let’s say this is uniform
- Sensor model: \( P(R|G) \)
- Given: we know what our sensors do
- \( R \) = color measured at \((1,1)\)
- E.g. \( P(R = yellow|G = (1,1)) = 0.1 \)
Let’s say we have two distributions:
- Prior distribution over ghost location: \( P(G) \)
- Let’s say this is uniform
- Sensor model: \( P(R|G) \)
- Given: we know what our sensors do
- \( R = \) color measured at (1,1)
- E.g. \( P(R = \text{yellow} | G = (1,1)) = 0.1 \)
Let’s say we have two distributions:
- Prior distribution over ghost location: $P(G)$
- Let’s say this is uniform
- Sensor model: $P(R|G)$
- Given: we know what our sensors do
- $R =$ color measured at (1,1)
- E.g. $P(R = \text{yellow} | G = (1,1)) = 0.1$

We can calculate the posterior distribution $P(G|r)$ over ghost locations given a reading using Bayes’ rule:
Let’s say we have two distributions:

- Prior distribution over ghost location: \( P(G) \)
- Let’s say this is uniform
- Sensor model: \( P(R|G) \)
- Given: we know what our sensors do
- \( R = \) color measured at (1,1)
- E.g. \( P(R = yellow|G = (1,1)) = 0.1 \)

We can calculate the posterior distribution \( P(G|r) \) over ghost locations given a reading using Bayes’ rule:

\[
P(g|r) \propto P(r|g)P(g)
\]
Ghostbusters, Revisited

Let’s say we have two distributions:
- Prior distribution over ghost location: $P(G)$
- Let’s say this is uniform
- Sensor model: $P(R|G)$
- Given: we know what our sensors do
- $R =$ color measured at (1,1)
- E.g. $P(R = \text{yellow}|G = (1,1)) = 0.1$

We can calculate the posterior distribution $P(G|r)$ over ghost locations given a reading using Bayes’ rule:

$$P(g|r) \propto P(r|g)P(g)$$

[Demo: Ghostbuster – with probability (L12D2) ]
Video of Demo Ghostbusters with Probability
Next up: bayes nets.

Bayes’ Nets
Bayes’ Nets

Probabilistic Models

Models describe how (a portion of) the world works


"All models are wrong; but some are useful." – George E. P. Box

What do we do with probabilistic models?

- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
- Example: value of information
Probabilistic Models


Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable

“All models are wrong; but some are useful.” – George E. P. Box

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
- Example: value of information
Probabilistic Models


Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.” – George E. P. Box
Probabilistic Models


Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.” – George E. P. Box
Probabilistic Models

Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.” – George E. P. Box

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
Probabilistic Models

Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.” – George E. P. Box

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
Probabilistic Models


Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.” – George E. P. Box

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
Probabilistic Models


Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.” – George E. P. Box

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
- Example: value of information
Probabilistic Models


Models describe how (a portion of) the world works

Models are always simplifications
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.” – George E. P. Box

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
- Example: value of information
Probability Recap

Probability Recap


Conditional probability:

\[
P(x \mid y) = \frac{P(x, y)}{P(y)}.
\]

Product rule:

\[
P(x, y) = P(x \mid y)P(y).
\]

Bayes Rule:

\[
P(y \mid x) = \frac{P(x \mid y)P(y)}{P(x)}.
\]

Chain rule:

\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2) \cdots P(x_n \mid x_1, x_2, \ldots, x_{n-1}).
\]

\[
X, Y\text{ independent if and only if: }\forall x, y: P(x, y) = P(x)P(y).
\]

\[X \text{ and } Y\text{ are conditionally independent given } Z\text{ if and only if: }\forall x, y, z: P(x, y \mid z) = P(x \mid z)P(y \mid z).
\]
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).
Probability Recap


Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule:
Probability Recap


Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.
Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule:
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
Probability Recap


Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)
\]

Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:

\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)
\]
Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x, y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots
\]
Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule: $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

Chain rule:

$$P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots$$
$$= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).$$

Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule: $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

Chain rule:

$$P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots$$

$$= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).$$

$X, Y$ independent if and only if:
Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule: $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

Chain rule:

\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots
= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).
\]

$X, Y$ independent if and only if: $\forall x, y : P(x, y) = P(x)P(y)$. 

Probability Recap


Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots \]
\[
= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).
\]

\( X, Y \) independent if and only if: \( \forall x, y : P(x, y) = P(x)P(y) \).

\( X \) and \( Y \) are conditionally independent given \( Z \) if and only if:
Probability Recap


Conditional probability: \( P(x|y) = \frac{P(x,y)}{P(y)} \).

Product rule: \( P(x,y) = P(x|y)P(y) \).

Bayes Rule: \( P(y|x) = \frac{P(x|y)P(y)}{P(x)} \).

Chain rule:
\[
P(x_1, x_2, \ldots, x_n) = P(x_1) P(x_2|x_1) P(x_3|x_1, x_2) \ldots \\
= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}).
\]

\( X, Y \) independent if and only if: \( \forall x, y : P(x,y) = P(x)P(y) \).

\( X \) and \( Y \) are conditionally independent given \( Z \) if and only if:
\[
\forall x, y, z : P(x,y|z) = P(x|z)P(y|z).
\]
Ghostbusters Chain Rule


Each sensor depends only on where the ghost is. That means, the two sensors are conditionally independent, given the ghost position.

\[ P(T,B,G) = P(G)P(T|G)P(B|G) \]

Givens:

\[ P(+g) = 0.5 \]
\[ P(-g) = 0.5 \]
\[ P(+t|+g) = 0.8 \]
\[ P(+t|-g) = 0.4 \]
\[ P(+b|+g) = 0.4 \]
\[ P(+b|-g) = 0.8 \]
Each sensor depends only on where the ghost is.
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[ P( +g ) = 0.5 \]
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[ P( +g ) = 0.5 \]
\[ P( -g ) = 0.5 \]
Ghostbusters Chain Rule

Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[
\begin{align*}
P(+g) &= 0.5 \\
P(-g) &= 0.5 \\
P(+t | +g) &= 0.8 \\
P(+t | -g) &= 0.4 \\
P(+b | +g) &= 0.4 \\
P(+b | -g) &= 0.8 \\
\end{align*}
\]
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

\[
P(+g) = 0.5 \\
P(-g) = 0.5 \\
P(+t|+g) = 0.8 \\
P(+t|-g) = 0.4 \\
P(+b|+g) = 0.4 \\
P(+b|-g) = 0.8 \\
\]

\[
P(T, B, G) = P(G)P(T|G)P(B|G) \\
\]
Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

Givens:

P( +g ) = 0.5
P( -g ) = 0.5
P( +t | +g ) = 0.8
P( +t | -g ) = 0.4
P( +b | +g ) = 0.4
P( +b | -g ) = 0.8

\[ P(T, B, G) = P(G)P(T|G)P(B|G) \]
Bayes’Nets: Big Picture

Encoding Complex Distributions

In 12 Easy Steps!
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly.
- Hard to learn (estimate) anything empirically about more than a few variables at a time.

Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities).

- More properly called graphical models.
- We describe how variables locally interact.
- Local interactions chain together to give global, indirect interactions.
- For now, we'll be vague about how these interactions are specified.
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly.
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time
Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly.
- Hard to learn (estimate) anything empirically about more than a few variables at a time.
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)

- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For now, we’ll be vague about how these interactions are specified
Bayes’ Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For now, we’ll be vague about how these interactions are specified
Example Bayes’ Net: Insurance
Example Bayes’ Net: Car
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate "direct influence" between variables
- Formally: encode conditional independence (more later)

For now: imagine that arrows mean direct causation (in general, they don't!)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate "direct influence" between variables
- Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don't!)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate "direct influence" between variables
- Formally: encode conditional independence

For now: imagine that arrows mean direct causation (in general, they don't!)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence (more later)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence (more later)
Graphical Model Notation

Nodes: variables (with domains)
- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions
- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence (more later)

For now: imagine that arrows mean direct causation (in general, they don’t!)
Example: Coin Flips

$X_1, X_2, \ldots, X_n$
Example: Coin Flips

$N$ independent coin flips
Example: Coin Flips

N independent coin flips
No interactions between variables: absolute independence

$X_1 \quad X_2 \quad \ldots \quad X_n$
Example: Traffic

Variables:
- R: It rains
Example: Traffic

Variables:
- R: It rains
- T: There is traffic
Example: Traffic

Variables:
- R: It rains
- T: There is traffic
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

$$R$$

$$T$$
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic
Example: Traffic

Variables:
- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic

Why is an agent using model 2 better?
Example: Traffic II

Let’s build a causal graphical model!
Example: Traffic II

Let’s build a causal graphical model!
Example: Traffic II

Let’s build a causal graphical model!

Variables

- T: Traffic
Example: Traffic II

Let's build a causal graphical model!

Variables

- T: Traffic
- R: It rains
Example: Traffic II

Let's build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
Example: Traffic II

Let’s build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity
Example: Traffic II

Let's build a causal graphical model!

Variables
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity
Example: Alarm Network
Example: Alarm Network

Variables

- B: Burglary
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
Example: Alarm Network

Variables

- B: Burglary
- A: Alarm goes off
- M: Mary calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!
Example: Alarm Network

Variables
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!
Bayes’ Net Semantics
Bayes’ Net Semantics

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

CPT: conditional probability table

Description of a noisy “causal” process

A Bayes net = Topology (graph) + Local Conditional Probabilities
Bayes’ Net Semantics

A set of nodes, one per variable $X$

$A_1, \ldots, A_n$
Bayes’ Net Semantics

A set of nodes, one per variable $X$
A directed, acyclic graph

$A_1 ... A_n$
Bayes’ Net Semantics

A set of nodes, one per variable $X$
A directed, acyclic graph
A conditional distribution for each node
Bayes’ Net Semantics

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
Bayes’ Net Semantics

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
- CPT: conditional probability table
Bayes’ Net Semantics

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
- CPT: conditional probability table
- Description of a noisy “causal” process
Bayes’ Net Semantics

- 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
  - CPT: conditional probability table
  - Description of a noisy “causal” process
Bayes’ Net Semantics

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
- CPT: conditional probability table
- Description of a noisy “causal” process

A Bayes net = Topology (graph) + Local Conditional Probabilities
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | Parents(X_i)) \]
Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i|\text{Parents}(X_i)) \]

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | \text{Parents}(X_i)) \]

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions
- As a product of local conditional distributions
  \[ P(x_1, \ldots, x_n) = \prod_i \Pr(x_i|\text{Parents}(X_i)) \]
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
  \[ P(x_1, \ldots, x_n) = \prod_i \Pr(x_i|\text{Parents}(X_i)) \]
Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | \text{Parents}(X_i)) \]

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

\[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | \text{Parents}(X_i)) \]

Example:

\[ Pr(G = b | T = r, B = r) = Pr(G = r | T = r) P(S = s | B = r) \]
Probabilities in Bayesnets

Bayes’ nets implicitly encode joint distributions

- As a product of local conditional distributions
  \[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | Parents(X_i)) \]

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
  \[ P(x_1, \ldots, x_n) = \prod_i Pr(x_i | Parents(X_i)) \]

Example:
\[
Pr(G = b | T = r, B = r) = Pr(G = r | T = r)P(S = s | B = r)
\]
\[
Pr(G = b | T = r, B = r) = Pr(G = r | T = r, B = r)P(S = s | B = r)
\]
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i)$$

Assume conditional independences:

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i))$$

Consequence:

Not every BN can represent every joint distribution.

The topology enforces certain conditional independencies!
Why are we guaranteed that setting results is joint distribution?
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_{i-1}) \]
Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i)$$
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i) \]

Assume conditional independences:
Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):
\[ P(x_1, \ldots, x_n) = \prod_i P(x_i|x_1, \ldots, x_i) \]

Assume conditional independences:
\[ P(x_1, \ldots, x_n) = \prod_i P(x_i|\text{parents}(X_i)) \]
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i)$$

Assume conditional independences:

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(X_i))$$

**Consequence:** Not every BN can represent every joint distribution.  
- The topology enforces certain conditional independencies!
Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i|x_1, \ldots, x_i) \]

Assume conditional independences:

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i|\text{parents}(X_i)) \]

**Consequence:** Not every BN can represent every joint distribution.

- The topology enforces certain conditional independencies!
Example: Coin Flips

$X_1$

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.
Example: Coin Flips

Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.
Example: Traffic

\[
\begin{align*}
P(+r,-t) &= P(+r) \times P(-t | +r) \\
&= \frac{1}{4} \times \frac{1}{4} \\
&= \frac{1}{16}
\end{align*}
\]
Example: Traffic

\[
P(+r, -t) = P(+r) \times P(-t | +r) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}
\]

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>3/4</td>
<td></td>
</tr>
</tbody>
</table>
Example: Traffic

\[ P(+r, -t) = \]

| R  | T   | P(T|R) |
|----|-----|--------|
| +r | +t  | 3/4    |
| +r | -t  | 1/4    |
| -r | +t  | 1/2    |
| -r | -t  | 1/2    |
Example: Traffic

\[ P(+r,-t) = P(+r) \times P(-t \mid +r) \]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
<th>P(T-R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>3/4</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>1/2</td>
</tr>
</tbody>
</table>

\[ R \rightarrow T \]
Example: Traffic

\[ P(+r,-t) = P(+r) \times P(-t \mid +r) \]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
<th>P(T→R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>3/4</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>1/2</td>
</tr>
</tbody>
</table>
Example: Traffic

\[ P(+r,-t) = P(+r) \times P(-t \mid +r) \]
\[ = \frac{1}{4} \times \frac{1}{4} \]
\[ = \frac{1}{16} \]
Example: Traffic

\[
P(+r,-t) = P(+r) \times P(-t \mid +r) = (1/4) \times (1/4) = 1/16
\]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
<th>P(T—R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>3/4</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>1/2</td>
</tr>
</tbody>
</table>
Example: Traffic

\[ P(+r, -t) = P(+r) \times P(-t \mid +r) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
<th>P(T—)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>3/4</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>1/2</td>
</tr>
</tbody>
</table>
Example: Alarm Network

- Burglary
- Earthqk
- Alarm
- John calls
- Mary calls

Table of Conditional Probabilities:

| Event | P(B)  | P(\text{Earthqk}) | P(J | A)  | P(M | A)  | P(E)  | P(A | B, E) |
|-------|-------|-------------------|-------|-------|--------|-------|------------|
|       | +b    | +j                |       | +a    | +m    | +e    | +b +e +a   |
|       | -b    | -j                |       | -a    | -m    | -e    | -b -e -a   |

P(B) = 0.001, P(\text{Earthqk}) = 0.999, P(J | A) = 0.9, P(M | A) = 0.7, P(E) = 0.002, P(A | B, E) = 0.95
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B  | E  | A  | P(A|B,E) |
|----|----|----|---------|
| +b | +e | +a | 0.95    |
| +b | +e | -a | 0.05    |
| +b | -e | +a | 0.94    |
| +b | -e | -a | 0.06    |
| -b | +e | +a | 0.29    |
| -b | +e | -a | 0.71    |
| -b | -e | +a | 0.001   |
| -b | -e | -a | 0.999   |

Burglary ➔ Alarm ➔ Earthquake

John calls ➔ Alarm ➔ Mary calls
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

| A  | J  | P(J|A) |
|----|----|------|
| +a | +j | 0.9  |
| +a | -j | 0.1  |
| -a | +j | 0.05 |
| -a | -j | 0.95 |

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B  | E  | A  | P(A|B,E) |
|----|----|----|---------|
| +b | +e | +a | 0.95    |
| +b | +e | -a | 0.05    |
| +b | -e | +a | 0.94    |
| +b | -e | -a | 0.06    |
| -b | +e | +a | 0.29    |
| -b | +e | -a | 0.71    |
| -b | -e | +a | 0.001   |
| -b | -e | -a | 0.999   |
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

| A   | J   | P(J|A) |
|-----|-----|------|
| +a  | +j  | 0.9  |
| +a  | -j  | 0.1  |
| -a  | +j  | 0.05 |
| -a  | -j  | 0.95 |

| A   | M   | P(M|A) |
|-----|-----|------|
| +a  | +m  | 0.7  |
| +a  | -m  | 0.3  |
| -a  | +m  | 0.01 |
| -a  | -m  | 0.99 |

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| B   | E   | A   | P(A|B,E) |
|-----|-----|-----|---------|
| +b  | +e  | +a  | 0.95    |
| +b  | +e  | -a  | 0.05    |
| +b  | -e  | +a  | 0.94    |
| +b  | -e  | -a  | 0.06    |
| -b  | +e  | +a  | 0.29    |
| -b  | +e  | -a  | 0.71    |
| -b  | -e  | +a  | 0.001   |
| -b  | -e  | -a  | 0.999   |
Example: Traffic

Causal direction

\[ P(T|R) = \frac{1}{4} \]

\[ P(R,T) = \frac{3}{16} \]
Example: Traffic

Causal direction

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>3/4</td>
<td></td>
</tr>
</tbody>
</table>
Example: Traffic

Causal direction

\[
P(T | R) = \begin{array}{c|c}
R & T \\
+\text{r} & 1/4 \\
-\text{r} & 3/4 \\
\end{array}
\]
Example: Traffic

Causal direction

\[
\begin{array}{c|c|c|c|c|c}
R & T & +r & 1/4 & -r & 3/4 \\
\hline
+r & +t & 3/4 & -r & -t & 1/2 \\
- & - & 1 & + & + & 2 \\
\end{array}
\]
Example: Traffic

Causal direction

\[ R \quad \rightarrow \quad T \]

\[
\begin{array}{|c|c|}
\hline
R & T \\
+ r & 1/4 \\
- r & 3/4 \\
\hline
\end{array}
\]

\[
P(T|R) \\
\begin{array}{|c|c|c|}
+ r & + t & 3/4 \\
+ r & - t & 1/4 \\
- r & + t & 1/2 \\
- r & - t & 1/2 \\
\hline
\end{array}
\]
Example: Traffic

Causal direction

\[
\begin{array}{c|c}
R & T \\
+r & 1/4 \\
-r & 3/4 \\
\end{array}
\]

\[
P(T|R)
\]

\[
\begin{array}{c|c|c}
 & +t & \frac{3}{4} \\
+T & +t & 1/4 \\
T & +t & 1/2 \\
\end{array}
\]

\[
P(R,T)
\]

\[
\begin{array}{c|c|c}
 & +t & \frac{1}{2} \\
+T & +t & 1/2 \\
T & +t & 1/2 \\
\end{array}
\]
Example: Traffic

Causal direction

\[ P(T|R) \]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
<th>( \frac{3}{4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>1/2</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>1/4</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>1/2</td>
</tr>
</tbody>
</table>

\[ P(R,T) \]

<table>
<thead>
<tr>
<th>R</th>
<th>T</th>
<th>( \frac{3}{16} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td>3/16</td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>1/16</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>6/16</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>6/16</td>
</tr>
</tbody>
</table>
Example: Reverse Traffic

Reverse causality?

\[ P(R|T) = \begin{cases} 
+R & \frac{1}{3} \\
-T & \frac{2}{3} \\
\end{cases} \quad 
\begin{cases} 
+T & \frac{3}{16} \\
-R & \frac{1}{16} \\
\end{cases} \quad 
\begin{cases} 
+R & \frac{6}{16} \\
-T & \frac{6}{16} \\
\end{cases} \]
Example: Reverse Traffic

Reverse causality?

\[
P(R | T) = \begin{cases} 
+\tau & \frac{9}{16} \\
-\tau & \frac{7}{16} 
\end{cases}
\]
Example: Reverse Traffic

Reverse causality?

\[
P(R|T)
\]

\[
\begin{array}{|c|c|}
\hline
+t & 9/16 \\
-t & 7/16 \\
\hline
\end{array}
\]
Example: Reverse Traffic

Reverse causality?

\[
\begin{align*}
P(R|T) & \\
+ t & 9/16 \\
- t & 7/16 \\
\end{align*}
\]

\[
\begin{align*}
+ t & + r & 1/3 \\
+ t & - r & 2/3 \\
- t & + r & 1/7 \\
- t & - r & 6/7 \\
\end{align*}
\]
Example: Reverse Traffic

Reverse causality?

\[
P(R|T) = \begin{array}{c|c|c}
+t & +r & 1/3 \\
-t & +r & 1/7 \\
+t & -r & 2/3 \\
-t & -r & 6/7
\end{array}
\]

\[
P(R,T) = \begin{array}{c|c|c}
+t & 9/16 \\
-t & 7/16
\end{array}
\]
Example: Reverse Traffic

Reverse causality?

\[
P(R | T) = \begin{array}{c|c|c}
+ t & + r & 1/3 \\
- t & - r & 2/3 \\
+ t & + r & 1/7 \\
- t & - r & 6/7 \\
\end{array}
\]

\[
P(R, T) = \begin{array}{c|c|c}
+ t & 9/16 \\
- t & 7/16 \\
\end{array}
\]
Example: Reverse Traffic

Reverse causality?

\[
P(R|T) = \begin{array}{c|cc}
+\text{t} & +\text{r} & 1/3 \\
-\text{t} & -\text{r} & 2/3 \\
\end{array}
\]

\[
P(R,T) = \begin{array}{c|ccc}
+\text{r} & +\text{t} & 3/16 \\
+\text{r} & -\text{t} & 1/16 \\
-\text{r} & +\text{t} & 6/16 \\
-\text{r} & -\text{t} & 6/16 \\
\end{array}
\]
Causality?

When Bayes' nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables Traffic and Drips
  - End up with arrows that reflect correlation, not causation

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)

\[ P(x_i | x_1, \ldots, x_{i-1}) = P(x_i | \text{parents}(X_i)) \]

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about

$$P(x_i | x_1, \ldots, x_{i-1}) = P(x_i | \text{parents}(X_i))$$

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts
When Bayes’ nets reflect the true causal patterns:

- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts
Causality?

When Bayes’ nets reflect the true causal patterns:

- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

$$P(x_i|x_1,\ldots,x_{i-1}) = P(x_i|\text{parents}(X_i))$$
When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

\[ P(x_i|x_1, \ldots, x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

What do the arrows really mean?
- Topology may happen to encode causal structure
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

\[ P(x_i|x_1, \ldots, x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
Causality?

When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

BNs need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

\[ P(x_i|x_1, \ldots, x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

First assembled BNs using an intuitive notion of conditional independence as causality.

Then saw that key property is conditional independence

Main goal: answer queries about conditional independence and influence

After that: how to answer numerical queries (inference)
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
- First assembled BNs using an intuitive notion of conditional independence as causality.
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
- First assembled BNs using an intuitive notion of conditional independence as causality.
- Then saw that key property is conditional independence
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
  - First assembled BNs using an intuitive notion of conditional independence as causality.
  - Then saw that key property is conditional independence
  - Main goal: answer queries about conditional independence and influence
Bayes’ Nets

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
  - First assembled BNs using an intuitive notion of conditional independence as causality.
  - Then saw that key property is conditional independence
  - Main goal: answer queries about conditional independence and influence
So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
  - First assembled BNs using an intuitive notion of conditional independence as causality.
  - Then saw that key property is conditional independence
  - Main goal: answer queries about conditional independence and influence

After that: how to answer numerical queries (inference)