Lecture 19

Conditional Independence, Bayesian networks intro
Assignment 4 will be out on next week.

- Due Tuesday April 3
Recap lecture 18

- Marginal Independence
- Conditional Independence
- Bayesian Networks Introduction
Recap: Conditioning

• Conditioning: revise beliefs based on new observations

• We need to integrate two sources of knowledge
  • Prior probability distribution $P(X)$: all background knowledge
  • New evidence $e$

• Combine the two to form a posterior probability distribution
  • The conditional probability $P(h|e)$
Recap: Conditional probability

Definition (conditional probability)
The conditional probability of formula $h$ given evidence $e$ is

$$P(h|e) = \frac{P(h \land e)}{P(e)}$$

E.g. $P(T = \text{hot}|W = \text{sunny}) = \frac{P(T=\text{hot} \land W=\text{sunny})}{P(W=\text{sunny})}$

<table>
<thead>
<tr>
<th>Possible world</th>
<th>Weather</th>
<th>Temperature</th>
<th>$\mu(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>sunny</td>
<td>hot</td>
<td>0.10</td>
</tr>
<tr>
<td>$w_2$</td>
<td>sunny</td>
<td>mild</td>
<td>0.20</td>
</tr>
<tr>
<td>$w_3$</td>
<td>sunny</td>
<td>cold</td>
<td>0.10</td>
</tr>
<tr>
<td>$w_4$</td>
<td>cloudy</td>
<td>hot</td>
<td>0.05</td>
</tr>
<tr>
<td>$w_5$</td>
<td>cloudy</td>
<td>mild</td>
<td>0.35</td>
</tr>
<tr>
<td>$w_6$</td>
<td>cloudy</td>
<td>cold</td>
<td>0.20</td>
</tr>
</tbody>
</table>

| $T$ | $P(T|W=\text{sunny})$ |
|-----|-----------------------|
| hot | 0.10/0.40=0.25        |
| mild| 0.20/0.40=0.50        |
| cold| 0.10/0.40=0.25        |
Conditional Probability among Random Variables

\[ P(X \mid Y) = \frac{P(X, Y)}{P(Y)} \]

\[ P(X \mid Y) = P(\text{Temperature} \mid \text{Weather}) = \frac{P(\text{Temperature} \land \text{Weather})}{P(\text{Weather})} \]

It expresses the conditional probability of each possible value for \( X \) given each possible value for \( Y \)

\[ P(X \mid Y) = \frac{P(X, Y)}{P(Y)} \]

Example:
Temperature \{hot, cold\}; Weather = \{sunny, cloudy\}

\[
\begin{array}{c|cc}
T = \text{hot} & T = \text{cold} \\
\hline
W = \text{sunny} & P(\text{hot} \mid \text{sunny}) & P(\text{cold} \mid \text{sunny}) \\
W = \text{cloudy} & P(\text{hot} \mid \text{cloudy}) & P(\text{cold} \mid \text{cloudy}) \\
\end{array}
\]

Which of the following is true?
A. The probabilities in each row should sum to 1
B. The probabilities in each column should sum to 1
C. Both of the above
D. None of the above
Conditional Probability among Random Variables

\[ P(X | Y) = \frac{P(X, Y)}{P(Y)} \]

It expresses the conditional probability of each possible value for \( X \) given each possible value for \( Y \)

\[ P(X | Y) = P(Temperature | Weather) = \frac{P(Temperature \land Weather)}{P(Weather)} \]

Example:
Temperature \{hot, cold\}; Weather = \{sunny, cloudy\}

\[
\begin{array}{c|c|c}
   & T = \text{hot} & T = \text{cold} \\
\hline
W = \text{sunny} & P(\text{hot}|\text{sunny}) & P(\text{cold}|\text{sunny}) \\
W = \text{cloudy} & P(\text{hot}|\text{cloudy}) & P(\text{cold}|\text{cloudy}) \\
\end{array}
\]

A. The probabilities in each row should sum to 1

These are two JPDs!
Recap: Inference by Enumeration

• Great, we can compute arbitrary probabilities now!

• Given
  • Prior joint probability distribution (JPD) on set of variables \( X \)
  • specific values \( e \) for the evidence variables \( E \) (subset of \( X \))

• We want to compute
  • posterior joint distribution of query variables \( Y \) (a subset of \( X \)) given evidence \( e \)

• Step 1: Condition to get distribution \( P(X|e) \)
• Step 2: Marginalize to get distribution \( P(Y|e) \)
Inference by Enumeration: example

- Given \( P(W,C,T) \) as JPD below, and evidence \( e : \text{“Wind=yes”} \)
- What is the probability that it is cold? I.e., \( P(T=\text{cold} | W=\text{yes}) \)
- Step 1: condition to get distribution \( P(C, T | W=\text{yes}) \)

<table>
<thead>
<tr>
<th>Windy W</th>
<th>Cloudy C</th>
<th>Temperature T</th>
<th>( P(W, C, T) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>no</td>
<td>hot</td>
<td>0.04</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td>mild</td>
<td>0.09</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td>cold</td>
<td>0.07</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>hot</td>
<td>0.01</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>mild</td>
<td>0.10</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>cold</td>
<td>0.12</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>hot</td>
<td>0.06</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>mild</td>
<td>0.11</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>cold</td>
<td>0.03</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>hot</td>
<td>0.04</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>mild</td>
<td>0.25</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>cold</td>
<td>0.08</td>
</tr>
</tbody>
</table>

\[
P(C = c \land T = t | W = \text{yes}) = \frac{P(C=c \land T=t \land W=\text{yes})}{P(W=\text{yes})}
\]
Inference by Enumeration: example

- Given $P(W,C,T)$ as JPD in previous slide, and evidence $e$: “Wind=yes”
- What is the probability that it is cold? I.e., $P(T=\text{cold} \mid W=\text{yes})$
- Step 2: marginalize over Cloudy to get distribution $P(T \mid W=\text{yes})$

<table>
<thead>
<tr>
<th>Cloudy</th>
<th>Temperature</th>
<th>$P(C, T \mid W=\text{yes})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sunny</td>
<td>hot</td>
<td>0.10</td>
</tr>
<tr>
<td>sunny</td>
<td>mild</td>
<td>0.21</td>
</tr>
<tr>
<td>sunny</td>
<td>cold</td>
<td>0.16</td>
</tr>
<tr>
<td>cloudy</td>
<td>hot</td>
<td>0.02</td>
</tr>
<tr>
<td>cloudy</td>
<td>mild</td>
<td>0.23</td>
</tr>
<tr>
<td>cloudy</td>
<td>cold</td>
<td>0.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temperature</th>
<th>$P(T \mid W=\text{yes})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.10+0.02 = 0.12</td>
</tr>
<tr>
<td>mild</td>
<td>0.21+0.23 = 0.44</td>
</tr>
<tr>
<td>cold</td>
<td>0.16+0.28 = 0.44</td>
</tr>
</tbody>
</table>

- This is a probability distribution: it defines the probability of all the possible values of Temperature (three here), given the observed value for Windy (yes).
- Because this is a probability distribution, the sum of all its values is...
Recap: Inference by Enumeration

- Great, we can compute arbitrary probabilities now!
- Given
  - Prior joint probability distribution (JPD) on \( X \)
  - specific values \( e \) for the evidence variables \( E \) (subset of \( X \))
- We want to compute
  - posterior joint distribution of query variables \( Y \) (a subset of \( X \)) given evidence \( e \)

- Step 1: Condition to get distribution \( P(X|e) \)
- Step 2: Marginalize to get distribution \( P(Y|e) \)

Generally applicable, but memory-heavy and slow
We will see a better way to do probabilistic inference
Bayes rule and Chain Rule

Theorem (Bayes theorem, or Bayes rule)

\[ P(h|e) = \frac{P(e|h) \times P(h)}{P(e)} \]

\[ P(\text{fire} | \text{alarm}) = \]
Bayes rule and Chain Rule

Theorem (Bayes theorem, or Bayes rule)

\[ P(h|e) = \frac{P(e|h) \times P(h)}{P(e)} \]

E.g., \( P(fire|alarm) = \frac{P(alarm|fire) \times P(fire)}{P(alarm)} \)
Product Rule

- By definition, we know that:

\[ P(f_2 \mid f_1) = \frac{P(f_2 \land f_1)}{P(f_1)} \]

- We can rewrite this to

\[ P(f_2 \land f_1) = P(f_2 \mid f_1) \times P(f_1) \]

- In general

**Theorem (Product Rule)**

\[ P(f_n \land \cdots \land f_{i+1} \land f_i \land \cdots \land f_1) = P(f_n \land \cdots \land f_{i+1} \mid f_i \land \cdots \land f_1) \times P(f_i \land \cdots \land f_1) \]
Chain Rule

• We know

\[ P(f_2 \land f_1) = P(f_2|f_1) \times P(f_1) \]

• In general:

\[ P(f_n \land f_{n-1} \land \ldots \land f_1) = \ldots = \prod_{i=1}^{n} \]
Chain Rule

- We know

\[ P(f_2 \land f_1) = P(f_2|f_1) \times P(f_1) \]

- In general:

\[
P(f_n \land f_{n-1} \land \cdots \land f_1) \\
= P(f_n|f_{n-1} \land \cdots \land f_1) \times P(f_{n-1} \land \cdots \land f_1) \\
= \cdots \\
= \prod_{i=1}^{n}
\]
Chain Rule

- We know
  \[ P(f_2 \land f_1) = P(f_2 | f_1) \times P(f_1) \]

- In general:
  \[
P(f_n \land f_{n-1} \land \cdots \land f_1) \\
  = P(f_n | f_{n-1} \land \cdots \land f_1) \times P(f_{n-1} \land \cdots \land f_1) \\
  = P(f_n | f_{n-1} \land \cdots \land f_1) \times P(f_{n-1} | f_{n-2} \land \cdots \land f_1) \\
  = \cdots \\
  = \prod_{i=1}^{n}
  \]
Chain Rule

• We know

\[ P(f_2 \land f_1) = P(f_2|f_1) \times P(f_1) \]

• In general:

\[
P(f_n \land f_{n-1} \land \cdots \land f_1) \\
= P(f_n|f_{n-1} \land \cdots \land f_1) \times P(f_{n-1} \land \cdots \land f_1) \\
= P(f_n|f_{n-1} \land \cdots \land f_1) \times P(f_{n-1}|f_{n-2} \land \cdots \land f_1) \times P(f_{n-2} \land \cdots \land f_1) \\
= \ldots \\
= \prod_{i=1}^{n} f_i
\]
Chain Rule

- We know

\[ P(f_2 \land f_1) = P(f_2|f_1) \times P(f_1) \]

- In general:

\[
P(f_n \land f_{n-1} \land ... \land f_1)
= P(f_n|f_{n-1} \land ... \land f_1) \times P(f_{n-1} \land ... \land f_1)
= P(f_n|f_{n-1} \land ... \land f_1) \times P(f_{n-1}|f_{n-2} \land ... \land f_1) \times P(f_{n-2} \land ... \land f_1)
= ... \]

\[= \prod_{i=1}^{n} P(f_i|f_{i-1} \land ... \land f_1) \]

Theorem: Chain Rule

\[ P(f_1 \land ... \land f_n) = \prod_{i=1}^{n} P(f_i|f_{i-1} \land ... \land f_1) \]
Chain Rule example

\[
P(f_1 \land \ldots \land f_n) = \prod_{i=1}^{n} P(f_i | f_{i-1} \land \ldots \land f_1)
\]

\[
P(A,B,C,D) = P(D|A,B,C) \times P(A,B,C) = P(D|A,B,C) \times P(C|A,B) \times P(A,B)
\]

\[
= P(D|A,B,C) \times P(C|B,A) \times P(B|A) \times P(A)
\]

\[
= P(A)P(B|A)P(C|A,B)P(D|A,B,C)
\]
Why does the chain rule help us?

We will see how, under specific circumstances (variables independence), this rule helps gain compactness

- We can represent the JPD as a product of marginal distributions
- We can simplify some terms when the variables involved are marginally independent or conditionally independent
Lecture Overview

- Recap lecture 18
- Marginal Independence
- Conditional Independence
- Bayesian Networks Introduction
Marginal Independence

Definition (Marginal independence)
Random variable $X$ is (marginally) independent of random variable $Y$, written $X \perp Y$, if for all $x \in \text{dom}(X)$, $y_j \in \text{dom}(Y)$ and $y_k \in \text{dom}(Y)$, the following equation holds:

\[
P(X = x | Y = y_j) = P(X = x | Y = y_k) = P(X = x)
\]

• Intuitively: if $X \perp Y$, then
  • learning that $Y = y$ does not change your belief in $X$
  • and this is true for all values $y$ that $Y$ could take

• For example, weather is marginally independent of the result of a coin toss
Examples for marginal independence

Definition (Marginal independence)
Random variable $X$ is (marginally) independent of random variable $Y$ if, for all $x_i \in \text{dom}(X)$, $y_j \in \text{dom}(Y)$ and $y_k \in \text{dom}(Y)$, the following equation holds:

$$P(X = x_i \mid Y = y_j) = P(X = x_i \mid Y = y_k) = P(X = x_i)$$

- Is Temperature marginally independent of Weather (see previous example)?

<table>
<thead>
<tr>
<th>Weather $W$</th>
<th>Temperature $T$</th>
<th>$P(W,T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sunny</td>
<td>hot</td>
<td>0.10</td>
</tr>
<tr>
<td>sunny</td>
<td>mild</td>
<td>0.20</td>
</tr>
<tr>
<td>sunny</td>
<td>cold</td>
<td>0.10</td>
</tr>
<tr>
<td>cloudy</td>
<td>hot</td>
<td>0.05</td>
</tr>
<tr>
<td>cloudy</td>
<td>mild</td>
<td>0.35</td>
</tr>
<tr>
<td>cloudy</td>
<td>cold</td>
<td>0.20</td>
</tr>
</tbody>
</table>
• Is Temperature marginally independent of Weather (see previous example)

A. yes  B. no

<table>
<thead>
<tr>
<th>Weather W</th>
<th>Temperature T</th>
<th>P(W,T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sunny</td>
<td>hot</td>
<td>0.10</td>
</tr>
<tr>
<td>sunny</td>
<td>mild</td>
<td>0.20</td>
</tr>
<tr>
<td>sunny</td>
<td>cold</td>
<td>0.10</td>
</tr>
<tr>
<td>cloudy</td>
<td>hot</td>
<td>0.05</td>
</tr>
<tr>
<td>cloudy</td>
<td>mild</td>
<td>0.35</td>
</tr>
<tr>
<td>cloudy</td>
<td>cold</td>
<td>0.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T$</th>
<th>$P(T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.15</td>
</tr>
<tr>
<td>mild</td>
<td>0.55</td>
</tr>
<tr>
<td>cold</td>
<td>0.30</td>
</tr>
</tbody>
</table>

| $T$       | $P(T|W=sunny)$ |
|-----------|----------------|
| hot       | 0.25           |
| mild      | 0.50           |
| cold      | 0.25           |
Examples for marginal independence

Definition (Marginal independence)
Random variable $X$ is (marginally) independent of random variable $Y$ if, for all $x_i \in \text{dom}(X)$, $y_j \in \text{dom}(Y)$ and $y_k \in \text{dom}(Y)$, the following equation holds:

\[
P(X = x_i | Y = y_j) = P(X = x_i | Y = y_k) = P(X = x_i)
\]

Is Weather marginally independent of temperature?

- No. We saw before that knowing the Temperature changes our belief on the weather

- E.g. $P(\text{hot}) = 0.15$
  $P(\text{hot} | \text{sunny}) = 0.25$

<table>
<thead>
<tr>
<th>$T$</th>
<th>$P(T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.15</td>
</tr>
<tr>
<td>mild</td>
<td>0.55</td>
</tr>
<tr>
<td>cold</td>
<td>0.30</td>
</tr>
</tbody>
</table>

| $T$       | $P(T|W=\text{sunny})$ |
|-----------|------------------------|
| hot       | 0.25                   |
| mild      | 0.50                   |
| cold      | 0.25                   |
Examples for marginal independence

Definition (Marginal independence)
Random variable $X$ is (marginally) independent of random variable $Y$ if, for all $x_i \in \text{dom}(X)$, $y_j \in \text{dom}(Y)$ and $y_k \in \text{dom}(Y)$, the following equation holds:

$$P(X = x_i | Y = y_j) = P(X = x_i | Y = y_k) = P(X = x_i)$$

Is Weather marginally independent of Temperature?

• We could have answered this question even without having the relevant probability distributions.
  ✓ Meteorological knowledge tells us that the weather influences the temperature, so information on what the weather is like should change one’s belief on the temperature

• If fact, for knowledge representation purposes, the evaluation for independence among variables will generally need to be made without numbers, based on pre-existing domain knowledge or assumptions
Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in \text{dom}(X)$, $y_j \in \text{dom}(Y)$ and $y_k \in \text{dom}(Y)$, the following equation holds:

$$P(X = x_i | Y = y_j) = P(X = x_i | Y = y_k) = P(X = x_i)$$

• Intuitively (without numbers):
  • Boolean random variable “Canucks win the Stanley Cup this season”
  • Numerical random variable “Canucks’ revenue last season”?
  • Are the two marginally independent?

A. yes  B. no
Examples for marginal independence

Definition (Marginal independence)
Random variable $X$ is (marginally) independent of random variable $Y$ if, for all $x_i \in \text{dom}(X)$, $y_j \in \text{dom}(Y)$ and $y_k \in \text{dom}(Y)$, the following equation holds:

$$P(X = x_i | Y = y_j) = P(X = x_i | Y = y_k) = P(X = x_i)$$

• Intuitively (without numbers):
  • Boolean random variable “Canucks win the Stanley Cup this season”
  • Numerical random variable “Canucks’ revenue last season”?
  • Are the two marginally independent?
    - No! Without revenue they cannot afford to keep their best players
Exploiting marginal independence

Recall the product rule

\[ p(X=x \land Y=y) = p(X=x \mid Y=y) \times p(Y=y) \]

If X and Y are marginally independent,

\[ p(X=x \mid Y=y) = p(X=x) \]

Thus we have

\[ p(X=x \land Y=y) = p(X=x) \times p(Y=y) \]

In distribution form

\[ p(X,Y) = p(X) \times p(Y) \]
Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, X_n)$ have?
Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, X_n)$ have?
  - One entry for each possible world: $2^n$

- How many entries would all the marginal distributions have combined?

  A. $2^n$  
  B. $2n$  
  C. $2+n$  
  D. $n^2$
Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, X_n)$ have?
  - One entry for each possible world: $2^n$

- How many entries would all the marginal distributions have combined?
  - Each of the $n$ tables only has two entries $P(X_1 = \text{true})$ and $P(X_1 = \text{false})$
  - So, in total: $2n$. Exponentially fewer than the JPD!
Given the binary variables A,B,C,D,
To specify $P(A,B,C,D)$ one needs the JDP below

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>P(A,B,C,D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

To specify $P(A) \times P(B) \times P(C) \times P(D)$
one needs the JDPs below

\[
\begin{array}{|c|c|}
\hline
A & P(A) \\
\hline
T & \ \\
F & \ \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|}
\hline
B & P(B) \\
\hline
T & \ \\
F & \ \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|}
\hline
C & P(C) \\
\hline
T & \ \\
F & \ \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|}
\hline
D & P(D) \\
\hline
T & \ \\
F & \ \\
\hline
\end{array}
\]
Lecture Overview

• Recap lecture 18
• Marginal Independence
• Conditional Independence
• Bayesian Networks Introduction
Conditional Independence

Definition (Conditional independence)
Random variable $X$ is (conditionally) independent of random variable $Y$ given random variable $Z$ if, for all $x_i \in \text{dom}(X)$, $y_j \in \text{dom}(Y)$, $y_k \in \text{dom}(Y)$ and $z_m \in \text{dom}(Z)$ the following equation holds:

$$P(X = x_i | Y = y_j, Z = z_m) = P(X = x_i | Y = y_k, Z = z_m) = P(X = x_i | Z = z_m)$$

• Intuitively: if $X$ and $Y$ are conditionally independent given $Z$, then
  • learning that $Y=y$ does not change your belief in $X$ when we already know $Z=z$
  • and this is true for all values $y$ that $Y$ could take and all values $z$ that $Z$ could take
Example for Conditional Independence

- Whether light $l_1$ is lit ($\text{Lit-} l_1)$ and the position of switch $s_2$ ($\text{Up-} s_2$) are not marginally independent
  - The position of the switch determines whether there is power in the wire $w_0$ connected to the light

- However, whether light $l_1$ is lit is conditionally independent from the position of switch $s_2$ given whether there is power in wire $w_0$ ($\text{Power-} w_0$)
  - Once we know $\text{Power-} w_0$, learning values for $\text{Up-} s_2$ does not change our beliefs about $\text{Lit-} l_1$
  - I.e., $\text{Lit-} l_1$ is conditionally independent of $\text{Up-} s_2$ given $\text{Power-} w_0$
Another example of conditionally but not marginally independent variables

- **ExamGrade** and **AssignmentGrade** are not marginally independent
  - Students who do well on one typically do well on the other, and vice versa

- But, conditional on **UnderstoodMaterial**, they are independent
  - Variable **UnderstoodMaterial** is a common cause of variables **ExamGrade** and **AssignmentGrade**
  - Knowing **UnderstoodMaterial** shields any information we could get from **AssignmentGrade** toward **Exam Grade** (and vice versa)
Two variables can be marginally but not conditionally independent

- “Smoking At Sensor” S: resident smokes cigarette next to fire sensor
- “Fire” F: there is a fire somewhere in the building
- “Alarm” A: the fire alarm rings
- S and F are marginally independent
  - Learning S=true or S=false does not change your belief in F, and vice versa

- But they are not conditionally independent given alarm
  - They are alternative causes for the alarm ringing - evidence on one of the two causes reduces the belief on the other if the alarm rings
  - E.g., if the alarm rings and you learn S=true your belief in F decreases
Conditional vs. Marginal Independence

Two variables can be

Conditionally but not marginally independent
- ExamGrade and AssignmentGrade
- ExamGrade and AssignmentGrade given UnderstoodMaterial
- Lit-l1 and Up-s2
- Lit-l1 and Up-s2 given Power_w0

Marginally but not conditionally independent
- SmokingAtSensor and Fire
- SmokingAtSensor and Fire given Alarm

Both marginally and conditionally independent
- CanucksWinStanleyCup and Lit_l1
- CanucksWinStanleyCup and Lit_l1 given Power_w0

Neither marginally nor conditionally independent
- Temperature and Cloudiness
- Temperature and Cloudiness given Wind
Learning Goals For Probability so far

- Define and give examples of random variables, their domains and probability distributions
- Calculate the probability of a proposition $f$ given $\mu(w)$ for the set of possible worlds
  - Define a joint probability distribution (JPD)
  - Marginalize over specific variables to compute distributions over any subset of the variables
- Given a JPD
  - Marginalize over specific variables
  - Compute distributions over any subset of the variables
- Prove the formula to compute conditional probability $P(h|e)$
- Use inference by enumeration
  - to compute joint posterior probability distributions over any subset of variables given evidence
- Derive and use Bayes Rule
- Derive the Chain Rule
- Define and use marginal independence
- Define and use conditional independence