Marginal Independence and Conditional Independence

Computer Science cpsc322, Lecture 26

(Textbook Chpt 6.1-2)

Nov, 2013
Lecture Overview

• Recap with Example
  – Marginalization
  – Conditional Probability
  – Chain Rule

• Bayes’ Rule

• Marginal Independence
• Conditional Independence

our most basic and robust form of knowledge about uncertain environments.
Recap Joint Distribution

3 binary random variables: \( P(H, S, F) \)

- \( H \) \( \text{dom}(H) = \{h, \neg h\} \) has heart disease, does not have...
- \( S \) \( \text{dom}(S) = \{s, \neg s\} \) smokes, does not smoke
- \( F \) \( \text{dom}(F) = \{f, \neg f\} \) high fat diet, low fat diet
Recap Joint Distribution

**Joint Prob. Distribution (JPD)**

- 3 binary random variables: \( P(H, S, F) \)
  - \( H \) \( \text{dom}(H) = \{h, \neg h\} \): has heart disease, does not have...
  - \( S \) \( \text{dom}(S) = \{s, \neg s\} \): smokes, does not smoke
  - \( F \) \( \text{dom}(F) = \{f, \neg f\} \): high fat diet, low fat diet

\[
\begin{array}{c|c|c}
\text{f} & s & \neg s \\
\hline
\text{s} & .015 & .007 \\
\text{\neg s} & .21 & .51 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{\neg f} & s & \neg s \\
\hline
\text{s} & .005 & .003 \\
\text{\neg s} & .07 & .18 \\
\end{array}
\]

\( 2^3 - 1 \) \hspace{1cm} \( 2^K - 1 \)
Recap Marginalization

\[ P(H, S) = \sum_{x \in \text{dom}(F)} P(H, S, F = x) \]

\[ P(H, S) \]

\[ P(H) \]

\[ P(S) \]

\[ \begin{array}{cc|cc}
    f & \neg f \\
    \hline
    s & \neg s & 0.015 & 0.007 \\
    \hline
    \neg s & \neg s & 0.005 & 0.003 \\
\end{array} \]
Recap Conditional Probability

\[
P(S \mid H) = \frac{P(S, H)}{P(H)}
\]

<table>
<thead>
<tr>
<th></th>
<th>(s)</th>
<th>(\neg s)</th>
<th>(P(H))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(h)</td>
<td>.02</td>
<td>.01</td>
<td>.03</td>
</tr>
<tr>
<td>(\neg h)</td>
<td>.28</td>
<td>.69</td>
<td>.97</td>
</tr>
</tbody>
</table>

| \(P(S)\) | .30 | .70 |

\[
P(S \mid \neg h) = \frac{P(S, \neg h)}{P(\neg h)}
\]

<table>
<thead>
<tr>
<th></th>
<th>(s)</th>
<th>(\neg s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(h)</td>
<td>.666</td>
<td>.333</td>
</tr>
<tr>
<td>(\neg h)</td>
<td>.29</td>
<td>.71</td>
</tr>
</tbody>
</table>

Two probability distributions for \(S\)

\[P(H \mid S) = \frac{P(H, S)}{P(S)}\]

Do this as an exercise
Recap Conditional Probability (cont.)

\[ P(S \mid H) = \frac{P(S, H)}{P(H)} \]

Two key points we covered in the previous lecture:

- We derived this equality from a possible world semantics of probability.
- It is not a probability distribution but...
- One for each configuration of the conditioning var(s).

If conditioned by \( k \) binary vars, set \( 2^k \) prob. distributions.
Recap Chain Rule

\[ P(H, S, F) = P(H) \cdot P(S | H) \cdot P(F | H, S) \]

Bayes Theorem

\[ P(S | H) = \frac{P(S, H)}{P(H)} \]

\[ P(H | S) = \frac{P(S, H)}{P(S)} \]

Substitute

\[ P(H | S) \cdot P(S) = P(S, H) \]
Lecture Overview

• Recap with Example and Bayes Theorem
• **Marginal Independence**
• Conditional Independence
Do you always need to revise your beliefs?

No. when your knowledge of \( Y \)'s value doesn't affect your belief in the value of \( X \)

**DEF.** Random variable \( X \) is **marginal independent** of random variable \( Y \) if, for all \( x_i \in \text{dom}(X), y_k \in \text{dom}(Y), \)

\[
P( X= x_i \mid Y= y_k ) = P(X= x_i )
\]
Marginal Independence: Example

- $X$ and $Y$ are independent iff:
  \[ P(X) = P(X|Y) = \frac{P(X,Y)}{P(Y)} \]
  or
  \[ P(Y) = P(Y|X) = \frac{P(X,Y)}{P(X)} \]
  or
  \[ P(X,Y) = P(X)P(Y) \]

- That is new evidence $Y$ (or $X$) does not affect current belief in $X$ (or $Y$)

- Ex:
  \[ P(\text{Toothache, Catch, Cavity, Weather}) = P(\text{Toothache, Catch, Cavity}) \]
  \[ \cdot P(\text{weather}) \]

- JPD requiring $32$ entries is reduced to two smaller ones ($8$ and $4$)
In our example are Smoking and Heart Disease marginally Independent?

What our probabilities are telling us....?

\[
\begin{array}{c|cc}
P(H,S) & s & \neg s \\
\hline
h & .02 & .01 \\
\neg h & .28 & .69 \\
\end{array}
\]

\[
P(H) = \begin{cases} 
.03 \\
.97 
\end{cases}
\]

\[
P(S|H) = \begin{cases} 
.666 \\
.334 
\end{cases}
\]

\[
P(S) = \begin{cases} 
.30 \\
.70 
\end{cases}
\]

\[P(S|H) = P(S) \Rightarrow \text{No.}\]
Lecture Overview

• Recap with Example
• Marginal Independence
• Conditional Independence
Conditional Independence

- With marg. Independence, for $n$ independent random vars, $O(2^n) \rightarrow O(n)$

$$P(x_1, \ldots, x_n) = P(x_1) \times \ldots \times P(x_n)$$

- Absolute independence is powerful but when you model a particular domain, it is rare.

- Dentistry is a large field with hundreds of variables, few of which are independent (e.g., Cavity, Heart-disease).

- What to do?
Look for weaker form of independence

- $P(\text{Toothache, Cavity, Catch})$

- Are Toothache and Catch marginally independent?
  $P\left(\text{catch} \mid \text{toothache, cavity}\right) = P(\text{Toothache})$? NO

- BUT If I have a cavity, does the probability that the probe catches depend on whether I have a toothache?
  (1) $P(\text{catch} \mid \text{toothache, cavity}) = P(\text{catch} \mid \text{cavity})$  NO

- What if I haven't got a cavity?
  (2) $P(\text{catch} \mid \text{toothache, } \neg \text{cavity}) = P(\text{catch} \mid \neg \text{cavity})$

- Each is directly caused by the cavity, but neither has a direct effect on the other
Conditional independence

• In general, Catch is conditionally independent of Toothache given Cavity:

1. \( P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity}) \)

• Equivalent statements:

2. \( P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \)

3. \( P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \)

\[ P(x, y) = P(x) P(y) \]
Proof of equivalent statements

\[ P(X \mid Y, Z) = P(X \mid Z) \implies \]

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

\[ P(Y \mid X, Z) = P(Y \mid Z) \]

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

\[ = \frac{P(Y, Z)}{P(Z)} \cdot \frac{P(X, Z)}{P(Z)} = P(Y \mid Z) \cdot P(X \mid Z) \]
Conditional Independence: Formal Def.

Sometimes, two variables might not be marginally independent. However, they become independent after we observe some third variable.

**DEF.** Random variable $X$ is conditionally independent of random variable $Y$ given random variable $Z$ if, for all $x_i \in \text{dom}(X)$, $y_k \in \text{dom}(Y)$, $z_m \in \text{dom}(Z)$

$$P( X= x_i \mid Y= y_k , Z= z_m ) = P(X= x_i \mid Z= z_m )$$

That is, knowledge of $Y$'s value doesn't affect your belief in the value of $X$, given a value of $Z$. 
Conditional independence: Use

- Write out full joint distribution using **chain rule**:

\[
P(Cavity, Catch, Toothache) = P(Toothache | Catch, Cavity) \cdot P(Catch | Cavity) \cdot P(Cavity)
\]

\[
= P(Toothache | Catch) \cdot P(Catch | Cavity) \cdot P(Cavity)
\]

\[
= P(Toothache | Cavity) \cdot P(Catch | Cavity) \cdot P(Cavity)
\]

- The use of conditional independence often reduces the size of the representation of the joint distribution from exponential in \(n\) to linear in \(n\). What is \(n\)?

\[
\text{# of vars} = \frac{2^3 - 1}{2 - 1} + 2 + 1 = 5
\]

- **Conditional independence** is our most basic and robust form of **knowledge** about uncertain environments.
Conditional Independence Example 2

- Given whether there is/isn’t power in wire $w_0$, is whether light $l_1$ is lit or not, independent of the position of switch $s_2$?

\[ P(l_1 | s_2, w_0) = P(l_1 | w_0) \]

- yes!
Conditional Independence Example 3

- Is every other variable in the system independent of whether light $l_1$ is lit, given whether there is power in wire $w_0$?

\[ P(s_1 | l_1, w_0) = P(s_1 | w_0) \]

$w_1$

$w_2$

... $w_2$

... $w_2$

\[ \text{yes!} \]
Learning Goals for today’s class

• You can:
• Derive the Bayes Rule

• Define and use Marginal Independence
• Define and use Conditional Independence
Where are we? (Summary)

- Probability is a rigorous formalism for uncertain knowledge

- Joint probability distribution specifies probability of every possible world

- Queries can be answered by summing over possible worlds

- For nontrivial domains, we must find a way to reduce the joint distribution size

- Independence \((rare)\) and conditional independence \((frequent)\) provide the tools
Next Class

- Bayesian Networks (Chpt 6.3)

Start working on assignments!