Reasoning under Uncertainty: Marginalization, Conditional Prob., and Bayes

Computer Science cpsc322, Lecture 25

(Textbook Chpt 6.1.3.1-2)

Nov, 6, 2013
Lecture Overview

– Recap Semantics of Probability
– Marginalization
– Conditional Probability
– Chain Rule
– Bayes' Rule
Recap: Possible World Semantics for Probabilities

Probability is a formal measure of subjective uncertainty.

- Random variable and probability distribution

  \[
  \begin{align*}
  X & \quad \text{dom}(X) = \{x_1, x_2, x_3\} \\
  x_2 & \quad \Rightarrow \quad P(x_2) \quad \sum = 1 \\
  x_3 & \quad \Rightarrow \quad P(x_3) \\
  \end{align*}
  \]

- Model Environment with a set of random vars

  \[
  X \quad Y \quad \text{binary} \\
  \mu(w) = 1 \quad \text{formula} \\
  \sum_{w \in W} \mu(w) = 1
  \]

- Probability of a proposition \( f \)

  \[
  P(f) = \sum_{w \models f} \mu(w)
  \]
Joint Distribution and Marginalization

Given a joint distribution, e.g. $P(X,Y,Z)$ we can compute distributions over any smaller sets of variables.

$$P(X,Y) = \sum_{z \in \text{dom}(Z)} P(X,Y,Z = z)$$

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>catch</th>
<th>$\mu(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>.108</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>.012</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>.072</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>.008</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>.016</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>.064</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>.144</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>.576</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>catch</th>
<th>$\neg$ catch</th>
<th>$P(\text{cavity}, \text{toothache})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>.108</td>
<td>.012</td>
<td>.12</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>.072</td>
<td>.008</td>
<td>.08</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>.016</td>
<td>.064</td>
<td>.08</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>.144</td>
<td>.576</td>
<td>.72</td>
</tr>
</tbody>
</table>
Joint Distribution and Marginalization

Given a joint distribution, e.g. \( P(X, Y, Z) \) we can compute distributions over any smaller sets of variables.

\[
P(X, Z) = \sum_{y \in \text{dom}(Y)} P(X, Z, Y = y)
\]

### Table A

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>catch</th>
<th>( \mu(w) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>.108</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>.012</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>.072</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>.008</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>.016</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>.064</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>.144</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>.576</td>
</tr>
</tbody>
</table>

### Table B

<table>
<thead>
<tr>
<th>cavity</th>
<th>catch</th>
<th>( P(\text{cavity}, \text{catch}) )</th>
<th>( P(\text{cavity}, \text{catch}) )</th>
<th>( P(\text{cavity}, \text{catch}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>.12</td>
<td>.18</td>
<td>.18</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>.08</td>
<td>.02</td>
<td>.72</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
## Joint Distribution and Marginalization

Given a joint distribution, e.g. $P(X, Y, Z)$, we can compute distributions over any smaller sets of variables.

$$P(X, Z) = \sum_{y \in \text{dom}(Y)} P(X, Y = y, Z)$$

### Table A

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>catch</th>
<th>$\mu(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>0.108</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>0.012</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>0.072</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>0.008</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>0.016</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>0.064</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>0.144</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>0.576</td>
</tr>
</tbody>
</table>

### Table B

<table>
<thead>
<tr>
<th>cavity</th>
<th>catch</th>
<th>$P(\text{cavity}, \text{catch})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.12</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.08</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>...</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>...</td>
</tr>
</tbody>
</table>

### Table C

<table>
<thead>
<tr>
<th>cavity</th>
<th>catch</th>
<th>$P(\text{cavity}, \text{catch})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.72</td>
</tr>
</tbody>
</table>
Why is it called Marginalization?

Why is it called Marginalization?

\[
P(X) = \sum_{y \in \text{dom}(Y)} P(X, Y = y)
\]

\[
P(X) = \sum_{y \in \text{dom}(Y)} P(X, Y = y)
\]
Lecture Overview

– Recap Semantics of Probability
– Marginalization
– **Conditional Probability**
– Chain Rule
– Bayes' Rule
– Independence
Conditioning
(Conditional Probability)

• We model our environment with a set of random variables.

• Assume have the joint, we can compute the probability of... any formula

• Are we done with reasoning under uncertainty?

• What can happen?

• Think of a patient showing up at the dentist office. Does she have a cavity?
Conditioning
(Conditional Probability)

• Probabilistic conditioning specifies how to revise beliefs based on new information.

• You build a probabilistic model (for now the joint) taking all background information into account. This gives the prior probability.

• All other information must be conditioned on.

• If evidence \( e \) is all of the information obtained subsequently, the conditional probability \( P(h|e) \) of \( h \) given \( e \) is the posterior probability of \( h \).

\[
P(\text{cavity} = T \mid \text{toothache} = T)
\]
Conditioning Example

• Prior probability of having a cavity
  \( P(\text{cavity} = T) \)

• Should be revised if you know that there is toothache
  \( P(\text{cavity} = T \mid \text{toothache} = T) \)

• It should be revised again if you were informed that the probe did not catch anything
  \( P(\text{cavity} = T \mid \text{toothache} = T, \text{catch} = F) \)

• What about?
  \( P(\text{cavity} = T \mid \text{sunny} = T) \)
How can we compute $P(h|e)$

- What happens in term of possible worlds if we know the value of a random var (or a set of random vars)?
- Some worlds are ruled out. The other become more likely.

$$\Sigma = P(e) = 0.2$$

$$e = (cavity = T)$$

$$\begin{align*}
M(w) &= \begin{cases} 
\mu(w) & \text{if } w \in e \\
0 & \text{if } w \notin e 
\end{cases} \\
M_e(w) &= \frac{M(w)}{P(e)}
\end{align*}$$

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>catch</th>
<th>$\mu(w)$</th>
<th>$\mu_e(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>0.108</td>
<td>0.54</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>0.012</td>
<td>0.06</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>0.072</td>
<td>0.36</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>0.008</td>
<td>0.04</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>0.016</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>0.064</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>0.144</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>0.576</td>
<td>0</td>
</tr>
</tbody>
</table>
How can we compute \( P(h|e) \)?

\[
P(h|e) = \sum_{w \models h} \mu_e(w)
\]

\[
P(\text{toothache} = F | \text{cavity} = T) = \sum_{w \models \text{toothache} = F} \mu_{\text{cavity}=T}(w)
\]

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>catch</th>
<th>( \mu(w) )</th>
<th>( \mu_{\text{cavity}=T}(w) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>.108</td>
<td>.54</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>.012</td>
<td>.06</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>.072</td>
<td>.36</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>.008</td>
<td>.04</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>.016</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>.064</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>.144</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>.576</td>
<td>0</td>
</tr>
</tbody>
</table>
Semantics of Conditional Probability

\[ \mu_e(w) = \begin{cases} \frac{1}{P(e)} \times \mu(w) & \text{if } w \models e \\ 0 & \text{if } w \not\models e \end{cases} \]

- The conditional probability of formula \( h \) given evidence \( e \) is

\[ P(h \mid e) = \sum_{w \models h} \mu_e(w) = \sum \frac{1}{P(e)} \times \mu(w) = \frac{1}{P(e)} \sum_{\omega \models h \land e} \mu(w) \]

A

B
### Semantics of Conditional Prob.: Example

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>catch</th>
<th>$\mu(w)$</th>
<th>$\mu_e(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>.108</td>
<td>.54</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>.012</td>
<td>.06</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>.072</td>
<td>.36</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>.008</td>
<td>.04</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>.016</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>.064</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>.144</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>.576</td>
<td>0</td>
</tr>
</tbody>
</table>

$e = (cavity = T)$

$$P(h \mid e) = P(toothache = T \mid cavity = T) = \frac{\sum_{w \in h} \mu_e(w)}{P(e)}$$
**Conditional Probability among Random Variables**

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

\[
P(X | Y) = \frac{P(\text{toothache} \mid \text{cavity})}{P(\text{cavity})}
\]

<table>
<thead>
<tr>
<th></th>
<th>Toothache = T</th>
<th>Toothache = F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity = T</td>
<td>.12</td>
<td>.08</td>
</tr>
<tr>
<td>Cavity = F</td>
<td>.08</td>
<td>.72</td>
</tr>
</tbody>
</table>

**Probability Distributions**

\[
P(C_X | Y) = \begin{cases} 0.2 & \text{Cavity} = T \\ 0.8 & \text{Cavity} = F \end{cases}
\]

\[
P(C_X | Y) = \begin{cases} 0.2 & \text{Toothache} = T \\ 0.8 & \text{Toothache} = F \end{cases}
\]
Product Rule

- **Definition of conditional probability:**
  \[ P(X_1 \mid X_2) = \frac{P(X_1, X_2)}{P(X_2)} \]

- **Product rule** gives an alternative, more intuitive formulation:
  \[ P(X_1, X_2) = P(X_2) P(X_1 \mid X_2) = P(X_1) P(X_2 \mid X_1) \]

- **Product rule** general form:
  \[
P(X_1, \ldots, X_n) = P(X_1, \ldots, X_t, X_{t+1}, \ldots, X_n) \]
  \[ = P(X_1, \ldots, X_t) P(X_{t+1}, \ldots, X_n \mid X_1, \ldots, X_t) \]
Chain Rule

- **Product rule** general form:

\[
P(X_1, \ldots, X_n) =
\]

\[
= P(X_1, \ldots, X_t) P(X_{t+1} \ldots X_n | X_1, \ldots, X_t)
\]

- **Chain rule** is derived by successive application of product rule:

\[
P(X_1, \ldots X_{n-1}, X_n) =
\]

\[
= P(X_1, \ldots X_{n-1}) P(X_n | X_1, \ldots, X_{n-1})
\]

\[
= P(X_1, \ldots X_{n-2}) P(X_{n-1} | X_1, \ldots, X_{n-2}) P(X_n | X_1, \ldots, X_{n-1})
\]

\[
= P(X_1) P(X_2 | X_1) \ldots P(X_{n-1} | X_1, \ldots, X_{n-2}) P(X_n | X_1, \ldots, X_{n-1})
\]

\[
= \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})
\]
Chain Rule: Example

\[ P(\text{cavity}, \text{toothache}, \text{catch}) = \]
\[ P(\text{cavity}) \times P(\text{toothache} | \text{cavity}) \times \]
\[ \times P(\text{catch} | \text{cavity}, \text{toothache}) \]

\[ P(\text{toothache}, \text{catch}, \text{cavity}) = \]
\[ P(\text{toothache}) \times P(\text{catch} | \text{toothache}) \times P(\text{cavity} | \text{toothache}, \text{catch}) \]

In how many other ways can this joint be decomposed using the chain rule?

A. 4  B. 1  C. 8  D. 0
Chain Rule: Example

\[ P(\text{cavity} , \text{toothache} , \text{catch}) = \]
\[ P(\text{cavity}) \times P(\text{toothache} | \text{cavity}) \times \]
\[ \times P(\text{catch} | \text{cavity} , \text{toothache}) \]

\[ P(\text{toothache} , \text{catch} , \text{cavity}) = \]
\[ P(\text{toothache}) \times P(\text{catch} | \text{toothache}) \times P(\text{cavity} | \text{toothache} , \text{catch}) \]

these and the other four decompositions are OK
Lecture Overview

– Recap Semantics of Probability
– Marginalization
– Conditional Probability
– Chain Rule
– Bayes' Rule
– Independence
Using conditional probability

• Often you have causal knowledge (forward from cause to evidence):
  – For example
    ✓ P(symptom | disease)
    ✓ P(light is off | status of switches and switch positions)
    ✓ P(alarm | fire)
  – In general: P(evidence e | hypothesis h)

• ... and you want to do evidential reasoning (backwards from evidence to cause):
  – For example
    ✓ P(disease | symptom)
    ✓ P(status of switches | light is off and switch positions)
    ✓ P(fire | alarm)
  – In general: P(hypothesis h | evidence e)
Bayes Rule

• By definition, we know that:
  \[ P(h \mid e) = \frac{P(h \wedge e)}{P(e)} \quad P(e \mid h) = \frac{P(e \wedge h)}{P(h)} \]

• We can rearrange terms to write
  \[ P(h \wedge e) = P(h \mid e) \times P(e) \quad (1) \]
  \[ P(e \wedge h) = P(e \mid h) \times P(h) \quad (2) \]

• But
  \[ P(h \wedge e) = P(e \wedge h) \quad (3) \]

• From (1) (2) and (3) we can derive

  \[ P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e)} \quad (3) \]
Example for Bayes rule

- On average, the alarm rings once a year
  - $P(\text{alarm}) = ?$

- If there is a fire, the alarm will almost always ring

- On average, we have a fire every 10 years

- The fire alarm rings. What is the probability there is a fire?
Example for Bayes rule

- On average, the alarm rings once a year
  \[ P(\text{alarm}) = \frac{1}{365} \]

- If there is a fire, the alarm will almost always ring
  \[ P(\text{alarm} | \text{fire}) = 0.999 \]

- On average, we have a fire every 10 years
  \[ P(\text{fire}) = \frac{1}{3650} \]

- The fire alarm rings. What is the probability there is a fire?
  - Take a few minutes to do the math!

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

iclicker

A. 0.999  B. 0.9  C. 0.0999  D. 0.1
Example for Bayes rule

- On average, the alarm rings once a year
  \[ P(\text{alarm}) = 1/365 \]

- If there is a fire, the alarm will almost always ring
  \[ P(\text{alarm}|\text{fire}) = 0.999 \]

- On average, we have a fire every 10 years
  \[ P(\text{fire}) = 1/3650 \]

- The fire alarm rings. What is the probability there is a fire?

\[
P(\text{fire}|\text{alarm}) = \frac{P(\text{alarm}|\text{fire}) \times P(\text{fire})}{P(\text{alarm})} = \frac{0.999 \times 1/3650}{1/365} = 0.0999
\]

- Even though the alarm rings the chance for a fire is only about 10%!
Learning Goals for today’s class

• You can:
• Given a joint, compute distributions over any subset of the variables
• Prove the formula to compute $P(h|e)$
• Derive the **Chain Rule** and the **Bayes Rule**
Next Class: Fri Nov 8
(postdoc Yashar Mehdad will sub for me)

• Marginal Independence
• Conditional Independence

Assignments

• I will post Assignment 3 this evening
• Assignment1 return today
• Assignment2: TAs are marking it
Plan for this week

- **Probability** is a rigorous formalism for uncertain knowledge

- **Joint probability distribution** specifies probability of every possible world

- Probabilistic 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
Conditional probability (irrelevant evidence)

• New evidence may be irrelevant, allowing simplification, e.g.,
  – \( P(\text{cavity} \ | \ \text{toothache, sunny}) = P(\text{cavity} \ | \ \text{toothache}) \)
  
  – We say that Cavity is conditionally independent from Weather (more on this next class)

• This kind of inference, sanctioned by domain knowledge, is crucial in probabilistic inference