Uncertainty

Chapter 13b
Inference by enumeration: 13.4

• Start with the joint distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>.108</td>
<td>.072</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.012</td>
<td>.008</td>
</tr>
<tr>
<td>cavity</td>
<td>.064</td>
<td>.144</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>.016</td>
<td>.576</td>
</tr>
</tbody>
</table>

• For any proposition \( \phi \), the probability that it is true is the sum of the probabilities of the atomic events that entail it:

\[
P(\phi) = \sum_{\omega : \omega \models \phi} P(\omega)
\]

(Q) What does the symbol \( \models \) mean?
Inference by enumeration

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>.108</td>
<td>.072</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.012</td>
<td>.008</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>.016</td>
<td>.144</td>
</tr>
<tr>
<td>catch</td>
<td>.064</td>
<td>.576</td>
</tr>
</tbody>
</table>

(Q) What is $P(\text{toothache})$?

$$P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

- This is called marginalization or summing out.
Inference by enumeration

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>.108</td>
<td>.072</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.012</td>
<td>.008</td>
</tr>
<tr>
<td>catch</td>
<td>.016</td>
<td>.144</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.064</td>
<td>.576</td>
</tr>
</tbody>
</table>

(Q) \( P(\text{cavity} \lor \text{toothache}) = 0.28 \). How do we arrive at that?

(Q) What is \( P(\text{cavity} \land \text{toothache}) \)?
Inference by enumeration

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>catch</td>
<td>¬ catch</td>
</tr>
<tr>
<td>cavity</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
</tbody>
</table>

- Conditional probabilities calculated by making a quotient.

\[
P(¬\text{cavity}|\text{toothache}) = \frac{P(¬\text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
\]

(Q) Write the probability expression for a cavity given that the probe catches.
(Q) If the probe catches, what is the probability of a cavity?
Denominator can be viewed as a normalization constant $\alpha$

$$P(Cavity|toothache) = \alpha P(Cavity, toothache)$$

- $Cavity$ is a variable (upper case), and $toothache$ a constant.

(Q) The result is a table. What size should the table be?

(Q) Compute normalized $P(Cavity|toothache)$ table $(\langle 0.6, 0.4 \rangle)$. 

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>$\neg$ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>$catch$</td>
<td>.108</td>
<td>.072</td>
</tr>
<tr>
<td>$\neg catch$</td>
<td>.012</td>
<td>.008</td>
</tr>
<tr>
<td>$cavity$</td>
<td>.016</td>
<td>.144</td>
</tr>
<tr>
<td>$\neg cavity$</td>
<td>.064</td>
<td>.576</td>
</tr>
</tbody>
</table>
Problems with Inference by Enumeration

Space and time issue:
1) Worst-case time complexity $O(d^n)$ where $d$ is the largest arity
2) Space complexity $O(d^n)$ to store the joint distribution
3) Difficulty computing numbers for $O(d^n)$ entries.

(Q) If there are 20 Boolean variables, how large is the joint probability table?
Independence

$A$ and $B$ are independent iff

$P(A|B) = P(A)$ or $P(B|A) = P(B)$ or $P(A, B) = P(A)P(B)$

- 32 entries reduced to 12
  - for $n$ independent biased coins, $2^n \rightarrow n$
  - Absolute independence powerful but rare

(Q) Draw the tables schematically (no need to fill in numbers).
(Q) How do we compute $P(\text{toothache}, \neg \text{catch}, \text{cavity}, \text{sunny})$?
Conditional independence

\( P(\text{Toothache}, \text{Cavity}, \text{Catch}) \) has \( 2^3 - 1 = 7 \) independent entries

If I have a cavity, the probability that the probe catches in it doesn’t depend on whether I have a toothache:

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

The same independence holds if I haven’t got a cavity:

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

Catch is conditionally independent of Toothache given Cavity:

\[
P(\text{Catch}|\text{Toothache}, \text{Cavity}) = P(\text{Catch}|\text{Cavity})
\]
Conditional independence contd.

Write out full joint distribution using chain rule:

\[
P(\text{Toothache, Catch, Cavity}) = P(\text{Toothache|Catch, Cavity})P(\text{Catch, Cavity}) \\
= P(\text{Toothache|Catch, Cavity})P(\text{Catch|Cavity})P(\text{Cavity}) \\
= P(\text{Toothache|Cavity})P(\text{Catch|Cavity})P(\text{Cavity})
\]

i.e., \(2 + 2 + 1 = 5\) independent numbers (ignoring redundant entries)

Example:

\[
P(\text{Toothache|Cavity}): \\
\begin{array}{cc}
\text{cavity} & \neg\text{cavity} \\
\text{toothache} & a & b \\
\neg\text{toothache} & 1 - a & 1 - b & \text{(redundant)}
\end{array}
\]

(Q) Draw \(P(\text{Catch|Cavity})\) and \(P(\text{Cavity})\) tables.
Write out full joint distribution using chain rule:

\[
P(\text{Toothache}, \text{Catch}, \text{Cavity})
= P(\text{Toothache}|\text{Catch}, \text{Cavity})P(\text{Catch}, \text{Cavity})
= P(\text{Toothache}|\text{Catch}, \text{Cavity})P(\text{Catch}|\text{Cavity})P(\text{Cavity})
= P(\text{Toothache}|\text{Cavity})P(\text{Catch}|\text{Cavity})P(\text{Cavity})
\]

I.e., \(2 + 2 + 1 = 5\) independent numbers (equations 1 and 2 remove 2)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in \(n\) to linear in \(n\).

**Conditional independence is our most basic and robust form of knowledge about uncertain environments.**
Bayes’ Rule

- Bayes’ rule gives us a way to calculate the posterior probability given evidence.

- Bayes’ rule can be derived by rearranging the terms of the product rule:

  Product rule: \( P(a \land b) = P(a|b)P(b) = P(b|a)P(a) \)

  \[ \Rightarrow \text{Bayes’ rule } P(a|b) = \frac{P(b|a)P(a)}{P(b)} \]

(Q) Why do we care?

Often, we can compute or estimate the terms on the right hand side, whereas the left hand side is what we want to compute.
Bayes’ Rule

Useful for assessing diagnostic probability from causal probability:

\[ P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)} \]

E.g., let \( M \) be meningitis, \( S \) be stiff neck:

\[ P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.001}{0.1} = 0.0008 \]

Note: posterior probability of meningitis still very small!
(Although 8 times more likely than before).

(Q) Discuss the right hand terms and explain how to estimate them.
(Q) Why is it easier to estimate the right hand terms than the left?
Bayes’ Rule and conditional independence

\[ P(Cavity|\text{toothache} \land \text{catch}) = \alpha P(\text{toothache} \land \text{catch}|Cavity)P(Cavity) = \alpha P(\text{toothache}|Cavity)P(\text{catch}|Cavity)P(Cavity) \]

This is an example of a naive Bayes model:

\[ P(\text{Cause}, \text{Effect}_1, \ldots, \text{Effect}_n) = P(\text{Cause})\prod_i P(\text{Effect}_i|\text{Cause}) \]

• Total number of parameters is linear in \( n \).
• Naive Bayes often works even if the vars not conditionally independent.

(Q) The top equations are in tabular form. Draw the tables.