Thanks to Andrew Moore for some course material.

Full Joint Probability Distributions

<table>
<thead>
<tr>
<th>Toothache</th>
<th>Cavity</th>
<th>Catch</th>
<th>P(Toothache, Cavity, Catch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>0.576</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>true</td>
<td>0.144</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>0.008</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>0.072</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>false</td>
<td>0.064</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>true</td>
<td>0.016</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>false</td>
<td>0.012</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
<td>0.108</td>
</tr>
</tbody>
</table>

"Catch" means the dentist’s probe catches in my teeth.

This cell means \( P(\text{Toothache} = \text{true}, \text{Cavity} = \text{true}, \text{Catch} = \text{true}) = 0.108 \)

The probabilities in the last column sum to 1.

Joint Probability Distribution

From the full joint probability distribution, we can calculate any probability involving the three random variables in this world.

\[
P(\text{Toothache} = \text{true} \text{ OR } \text{Cavity} = \text{true}) =
\]

\[
P(\text{Toothache} = \text{true}, \text{Cavity} = \text{false}, \text{Catch} = \text{false}) +
\]

\[
P(\text{Toothache} = \text{true}, \text{Cavity} = \text{false}, \text{Catch} = \text{true}) +
\]

\[
P(\text{Toothache} = \text{false}, \text{Cavity} = \text{true}, \text{Catch} = \text{false}) +
\]

\[
P(\text{Toothache} = \text{false}, \text{Cavity} = \text{true}, \text{Catch} = \text{true}) +
\]

\[
= 0.064 + 0.016 + 0.008 + 0.072 + 0.012 + 0.108 = 0.28
\]

Marginalization

We can even calculate marginal probabilities (the probability distribution over a subset of the variables) eg:

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

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

\[
P(\text{Toothache} = \text{true}, \text{Cavity} = \text{true}, \text{Catch} = \text{false}) +
\]

\[
P(\text{Toothache} = \text{false}, \text{Cavity} = \text{true}, \text{Catch} = \text{true}) +
\]

\[
P(\text{Toothache} = \text{false}, \text{Cavity} = \text{true}, \text{Catch} = \text{false}) +
\]

\[
= 0.108 + 0.012 + 0.072 + 0.008 = 0.2
\]
Marginalization

The general marginalization rule for any **sets** of variables $Y$ and $Z$:

$$P(Y) = \sum_z P(Y,z)$$

or

$$P(Y) = \sum_z P(Y | z) P(z)$$

$z$ is over all possible combinations of values of $Z$ (remember $Z$ is a set).

Marginalization

For continuous variables, marginalization involves taking the integral:

$$P(Y) = \int P(Y,z) dz$$

Normalization

$$P(Cavity = true | Toothache = true) = \frac{P(Cavity = true, Toothache = true)}{P(Toothache = true)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Note that $1/P(Toothache=true)$ remains constant in the two equations.

$$P(Cavity = false | Toothache = true) = \frac{P(Cavity = false, Toothache = true)}{P(Toothache = true)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

• In fact, $1/P(Toothache)$ can be viewed as a normalization constant for $P(Cavity | Toothache)$, ensuring it adds up to 1
• We will refer to normalization constants with the symbol $\alpha$

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

Inference

• Suppose you get a query such as

$$P(Cavity = true | Toothache = true)$$

Toothache is called the evidence variable because we observe it. More generally, it’s a set of variables.

Cavity is called the query variable (we’ll assume it’s a single variable for now)

There are also unobserved (aka hidden) variables like Catch

Inference

• We will write the query as $P(X | e)$

$X$ = Query variable (a single variable for now)

$E$ = Set of evidence variables

$e$ = the set of observed values for the evidence variables

$Y$ = Unobserved variables
Inference

We will write the query as $P(X \mid e)$

$P(X \mid e) = \alpha P(X, e) = \alpha \sum_{\mathcal{Y}} P(X, e, \mathcal{Y})$

$\mathcal{Y} = \text{Unobserved variables}$

$X$ = Query variable (a single variable for now)

$E$ = Set of evidence variables

$e$ = the set of observed values for the evidence variables

Summation is over all possible combinations of values of the unobserved variables $\mathcal{Y}$

Independence

• How do you avoid the exponential space and time complexity of inference?

• Use independence (aka factoring)

Independence

Suppose the full joint distribution now consists of four variables:

Toothache = \{true, false\}

Catch = \{true, false\}

Cavity = \{true, false\}

Weather = \{sunny, rain, cloudy, snow\}

There are now 32 entries in the full joint distribution table

Independence

Does the weather influence one’s dental problems?

Is $P(\text{Weather}=\text{cloudy} \mid \text{Toothache} = \text{toothache}, \text{Catch} = \text{catch}, \text{Cavity} = \text{cavity}) = P(\text{Weather}=\text{cloudy})$?

In other words, is Weather independent of Toothache, Catch and Cavity?

Independence

We say that variables $X$ and $Y$ are independent if any of the following hold:

(note that they are all equivalent)

$P(X \mid Y) = P(X)$ or

$P(Y \mid X) = P(Y)$ or

$P(X, Y) = P(X)P(Y)$
Why is independence useful?

Assume that Weather is independent of toothache, catch, cavity ie.
\[ P(\text{Weather} = \text{cloudy} | \text{Toothache} = \text{toothache}, \text{Catch} = \text{catch}, \text{Cavity} = \text{cavity}) = P(\text{Weather} = \text{cloudy}) \]

Now we can calculate:
\[ P(\text{Weather} = \text{cloudy}, \text{Toothache} = \text{toothache}, \text{Catch} = \text{catch}, \text{Cavity} = \text{cavity}) = P(\text{Weather} = \text{cloudy}) \times P(\text{Toothache} = \text{toothache}, \text{Catch} = \text{catch}, \text{Cavity} = \text{cavity}) \]

This table has 4 values

This table has 8 values

Why is independence useful?

If Weather was not independent of Toothache, Catch, and Cavity then you would have needed 32 values

Independence

Another example:

• Suppose you have n coin flips and you want to calculate the joint distribution \( P(C_1, \ldots, C_n) \)
• If the coin flips are not independent, you need \( 2^n \) values in the table
• If the coin flips are independent, then
\[ P(C_1, \ldots, C_n) = \prod_{i=1}^{n} P(C_i) \]

Bayes’ Rule

The product rule can be written in two ways:
\[ P(A, B) = P(A | B)P(B) \]
\[ P(A, B) = P(B | A)P(A) \]

You can combine the equations above to get:
\[ P(B | A) = \frac{P(A | B)P(B)}{P(A)} \]

Bayes’ Rule

More generally, the following is known as Bayes’ Rule:
\[ P(A | B) = \frac{P(B | A)P(A)}{P(B)} \]

Note that these are distributions

Sometimes, you can treat \( P(B) \) as a normalization constant \( \alpha \)
\[ P(A | B) = \alpha P(B | A)P(A) \]
More General Forms of Bayes Rule

If A takes 2 values:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)} \]

If A takes k values:

\[ P(A = v_i | B) = \frac{P(B | A = v_i)P(A = v_i)}{\sum_{k=1}^{k} P(B | A = v_k)P(A = v_k)} \]

Bayes Rule Example

Meningitis causes stiff necks with probability 0.5. The prior probability of having meningitis is 0.00002. The prior probability of having a stiff neck is 0.05. What is the probability of having meningitis given that you have a stiff neck?

Let \( m \) = patient has meningitis

Let \( s \) = patient has stiff neck

\( P(s|m) = 0.5 \)

\( P(m) = 0.00002 \)

\( P(s) = 0.05 \)

\[
\frac{P(m|s)P(m)}{P(s)} = \frac{(0.5)(0.00002)}{0.05} = 0.0002
\]

Bayes Rule With More Than One Piece of Evidence

Suppose you now have 2 evidence variables

Toothache = true and Catch = catch (note that Cavity is uninstantiated below)

\( P(Cavity | Toothache = true, Catch = true) = \alpha \)

\( P(Toothache = true, Catch = true | Cavity) P(Cavity) \)

In order to calculate \( P(Toothache = true, Catch = true | Cavity) \), you need a table of 4 probability values. With N evidence variables, you need \( 2^N \) probability values.
Conditional Independence

Are Toothache and Catch independent?

No – if probe catches in the tooth, it likely has a cavity which causes the toothache.

But given the presence or absence of the cavity, they are independent (since they are directly caused by the cavity but don’t have a direct effect on each other).

Conditional independence:

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

Conditional Independence

General form:

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

Or equivalently:

\[ P(A | B, C) = P(A | C) \quad \text{and} \quad P(B | A, C) = P(B | C) \]

How to think about conditional independence:

In \( P(A | B, C) = P(A | C) \): if knowing \( C \) tells me everything about \( A \), I don’t gain anything by knowing \( B \).

What You Should Know

• How to do inference in joint probability distributions
• How to use Bayes Rule
• Why independence and conditional independence is useful