COMP304: Knowledge Representation and Reasoning

Lecture 18: Probability Theory (2)

Today

- Prior probability and posterior probability
  - Examples
  - Bayes' rule
  - Jeffrey's rule
- Events that are unrelated
  - Examples and definitions

Bayes' Rule

- So far we have considered the probabilities of outcomes *before* the experiment takes place. Now we will consider probabilities *after* the experiment has taken place.

- Consider a student who needs to retake one course in order to graduate. The choices are mathematics (M), chemistry (C) or computer science (S).

- Based on his interest in the subjects, the student gives probabilities of 0.1, 0.6 and 0.3 to the event of choosing each of these.

- Based on his past performance, his tutor estimates the probability of his passing (P) as being 0.8 if he takes maths, 0.7 if he takes chemistry, and 0.75 if he takes computer science.

Bayes' Rule

- What is the probability of the student passing?

$$\Pr(P) = \Pr(P \cap M \text{ or } P \cap C \text{ or } P \cap S)$$

Since the three "or"ed events are mutually exclusive, we can use the addition law and the multiplication rule to give:

$$\begin{aligned}\Pr(P) = {} & \Pr(P \mid M)\Pr(M) \\ & + \Pr(P \mid C)\Pr(C) \\ & + \Pr(P \mid S)\Pr(S) \\ = {} & (0.8)(0.1) + (0.7)(0.6) + (0.75)(0.3) \\ = {} & 0.725\end{aligned}$$

- This gives the probability the student passes before he takes the course.

Bayes' Rule

- If we know the student passed, what is the probability he took maths?

- This is equivalent to asking "what is $\Pr(M \mid P)$?".

- From Definition 6 (conditional probability) we know that:

$$\Pr(M \mid P) = \frac{\Pr(M \cap P)}{\Pr(P)}$$

- We already know that $\Pr(P)$ is 0.725, we also know that:

$$\Pr(M \cap P) = \Pr(P \mid M)\Pr(M)$$

Bayes' Rule

- Thus:

$$\Pr(M \mid P) = \frac{\Pr(P \mid M)\,\Pr(M)}{\Pr(P)}$$

$$= \frac{(0.8)(0.1)}{0.725}$$

$$= 0.1103$$

- This is the ratio of the probability of passing by taking maths, to the probability of all possible ways of passing.

Exercise

- How would we calculate the probability that the student took maths given that he is known to fail to graduate?

Answer

$$\Pr(F) = 1 - 0.725$$
$$= 0.275$$

$$\Pr(M \mid F) = \frac{\Pr(M \cap F)}{\Pr(F)}$$

$$\Pr(M \mid F) = \frac{\Pr(F \mid M)\Pr(M)}{\Pr(F)}$$

$$= \frac{(0.2)(0.1)}{0.275}$$

$$= 0.0727$$

Posterior Probability

- Two important ideas can be drawn from this example.

- The first is the idea of *posterior* probability – after we know the outcome of the experiment, we know more and so can update the probabilities of events about which we are still not certain:

Event	Prior	Posterior given P	Posterior given F
M	0.1	0.1103	0.0727
C	0.6	0.5793	0.6545
S	0.3	0.3103	0.2727

- This is a situation we are frequently in – we have a probability model, learn something about its overall outcome, and then want to update the probability of some of the events in the model.

(Note: F is the probability of the student failing)

Bayes’ Rule

- The second important idea is Bayes’ rule, which is the name often given to the relation:

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)}$$

  This idea has particular resonance in artificial intelligence.

- This is because, as in the last example, it makes it possible to compute the posterior probability of B given that A occurs, from the conditional probability that A happens given that B occurs.

- This is useful because one is usually something that is easy to measure and the other is something we want to know.

Bayes' Rule

- If we consider *B* to be the event "the patient has disease *X*" and *A* to be the event "the patient displays symptom *Y*", then Bayes' rule lets us take:
  - The probability that a patient with the disease also has the symptom, Pr(*A* | *B*), which we can measure;
  - The prior probability the patient has the disease, Pr(*B*), which we can measure; and
  - The prior probability the patient has the symptom, Pr(*A*), which we can measure.

and from these calculate that a patient with the symptom has the disease, Pr(*B* | *A*), which is something we would like to know.

Bayes' Rule

- We can generalise Bayes' rule.

- Consider an arbitrary event F, and recall that we can decompose this using a set of mutually exclusive events $A_1, \dots, A_n$:

  $$F = (A_1 \cap F) \cup \dots \cup (A_n \cap F)$$

  where $A_1, \dots, A_n$ need not be a partition of S. Applying the addition law:

  $$\Pr(F) = \sum_{i=1 \dots n} \Pr(A_i \cap F)$$

  and this can be transformed by applying the multiplication law:

  $$\Pr(F) = \sum_{i=1 \dots n} \Pr(F \mid A_i) \Pr(A_i)$$

- We can then add this expression for $\Pr(F)$ into our previous version of Bayes' rule to find the probability of a particular A_j taking place, given that F is known to have happened.

Bayes' Rule

- This gives the general version of Bayes' rule:

- **Theorem 5:** *Given an arbitrary event F in S which is a subset of the union of the mutually exclusive events* $A_1, \ldots, A_n$*, such that* $\Pr(F) \neq 0$*, then:*

$$\Pr(A_j \mid F) = \frac{\Pr(F \mid A_j)\Pr(A_j)}{\sum_{i=1 \ldots n} \Pr(F \mid A_i)\Pr(A_i)}$$

for any $j = 1, \ldots, n$.

- The expression:

$$\Pr(F) = \sum_{i=1 \ldots n} \Pr(F \mid A_i)\Pr(A_i)$$

is sometimes known as Jeffrey's rule. It provides a way of calculating the probability of F when we don't know which of the A_i have taken place.

Independence

- In everyday language we refer to events that "have nothing to do with each other" as being *independent*.

- A similar notion of independence is useful in probability theory as it helps to structure probabilistic knowledge.

- To develop this, consider the following scenario: six trees planted in a straight line, two of which we know are diseased.

  (a) If all trees are equally likely to be diseased, what is the probability that the diseased trees are next to each other?

  (b) If we know that tree 3 is diseased, what is the probability that the diseased trees are next to each other?

  (c) Now, if the trees are planted in a circle with 1 and 6 adjacent, and 3 is known to be diseased, what is the probability that the diseased trees are next to each other?

Independence - Example

- The sample space consists of all subsets of size 2:

$$S = \{ (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 3), (2, 4), (2, 5), (2, 6), (3, 4), (3, 5), (3, 6), (4, 5), (4, 6), (5, 6)\}$$

- Let these represent the possible pairs of diseased trees.
- Five of these pairs represent adjacent trees:

$$\{(1, 2), (2, 3), (3, 4), (4, 5), (5, 6)\}$$

- Let A denote the event that diseased trees are adjacent,

$$\Pr(A) = \frac{5}{15} = \frac{1}{3}$$

and we have the solution to (a).

Independence - Example

- Now we condition on event B, that tree 3 is diseased. Recall:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- Looking at S, event B contains five sample points:

$$\{(1, 3), (2, 3), (3, 4), (3, 5), (3, 6)\}$$

- The joint event $\Pr(A \cap B)$ corresponds to the set of sample points:

$$\{(2, 3), (3, 4)\}$$

Thus:

$$\Pr(A \mid B) = \frac{\frac{2}{15}}{\frac{5}{15}} = \frac{2}{5}$$

and we have the solution to (b).

Independence - Example

- How does planting the trees in a circle alter the problem?

- Consider whether *S* changes in this new situation, what the sample points of A are and what the probabilities are…

Exercise

- How does planting the trees in a circle alter the problem?

Answer

- How does planting the trees in a circle alter the problem?
- Well, S does not change in this new situation, but now A has an extra sample point, (1, 6).

As a result:

$$\Pr(A) = \frac{6}{15} = \frac{2}{5}$$

$$\Pr(A \mid B) = \frac{\frac{2}{15}}{\frac{5}{15}} = \frac{2}{5}$$

and we have the solution to (c).

Independence - Example

- When the trees were planted in a line, learning that tree 3 was diseased (learning that *B* was the case) changed the probability that two adjacent trees were diseased because:

$$\Pr(A) \neq \Pr(A \mid B)$$

- When the trees were planted in a circle, knowing that tree 3 is diseased gives us no new information:

$$\Pr(A) = \Pr(A \mid B)$$

- It is this sense of "unrelatedness" that we use to define independence in probability theory.

Independence

- **Definition 7:** *Two events E and F in a sample space S are* statistically independent *if:*

$$\Pr(E \mid F) = \Pr(E)$$

- We will refer to statistically independent events as being "independent".

- Independence leads to a special case of the multiplication law:

  Theorem 6: *Two events F and E are independent if and only if:*

$$\Pr(E \cap F) = \Pr(E)\,\Pr(F)$$

- Thus, we can establish the independence of *E* and *F* by showing that either of the above hold.

Independence for More Than Two Events

- Generalising for n different events; $A_1, \ldots, A_n$ are independent if:

$$\Pr(A_1 \cap ,\ldots, \cap A_n) = \Pr(A_1)\Pr(A_2) \ldots \Pr(A_n)$$

- However, for $n > 2$ this is not sufficient to guarantee that all the probability laws hold, so instead we have:

  Definition 8: *The events A_1 ,…, A_n are* mutually independent *if the joint probability of every combination of events is equal to the product of their individual probabilities.*

- For the events A_1, A_2, A_3 this means that:

$$\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1)\Pr(A_2)\Pr(A_3)$$
$$\Pr(A_1 \cap A_2) = \Pr(A_1)\Pr(A_2)$$
$$\Pr(A_1 \cap A_3) = \Pr(A_1)\Pr(A_3)$$
$$\Pr(A_2 \cap A_3) = \Pr(A_2)\Pr(A_3)$$

Pairwise Independence

- All four conditions *must* hold.
- When only the last three hold, the events are said to be *pairwise independent*.

- Consider the nickel and dime example again:

  A_1 : Head on nickel, A_2 : Head on dime, A_3 : Coins match

- From the sample space and the probabilities of the sample points:

$$\Pr(A_1) = \Pr(A_2) = \Pr(A_3) = ½$$

  and:

$$\Pr(A_1 \cap A_2) = \Pr(A_2 \cap A_3) = \Pr(A_1 \cap A_3) = ¼$$

- But:

$$\Pr(A_1 \cap A_2 \cap A_3) = ¼ \neq \Pr(A_1) \Pr(A_2) \Pr(A_3)$$

- The events are thus pairwise independent but not mutually independent.

Independence

- It is also possible to have situations in which events are neither mutually independent nor pairwise independent despite the fact that our generalisation of independence for *n* different events holds.

- For example, the situation in which we have events M, E and W, with:

$$\Pr(W) = 0.6$$
$$\Pr(E) = 0.8$$
$$\Pr(M) = 0.5$$
$$\Pr(W \cap E) = 0.54$$
$$\Pr(W \cap M) = 0.24$$
$$\Pr(E \cap M) = 0.42$$
$$\Pr(W \cap E \cap M) = 0.24$$

- While it is true that: $\Pr(E \cap M \cap W) = \Pr(E)\Pr(M)\Pr(W)$
the events are neither pairwise independent nor mutually independent.

Multiplication Law Revisited

- Mutual independence is a useful concept as we can easily extend the multiplication law for mutually independent events.

- The extended law is:

  Theorem 7: *If we have a set of mutually independent events* $A_1 , \ldots , A_n$*, then:*

$$\Pr(A_1 \cap, \ldots ,\cap A_n) = \Pr(A_1) \Pr(A_2) \ldots \Pr(A_n)$$

- This follows directly from the definition of mutual independence.

- This allows us to establish the joint probability of a set of events from their individual probabilities.

Summary

- The difference between prior probability (before an experiment takes place) and posterior probability (after an experiment has taken place)
- Bayes' rule
- Independence
  - Establishing when events are independent
  - Mutual independence
  - Pairwise independence