Sets, Relations and Probability. Part IA Formal Methods.

Lecture VI, *Bayes Theorem*, 5th March.

Christopher J. Masterman (cm789@cam.ac.uk, christophermasterman.com)

Last week, we introduced some basic elements of probability: outcome spaces V, fields F_V, probability functions, the Kolmogorov Axioms, the notion of a conditional probability, and the notions of probabilistically independent and dependent events. This week, we continue with probability, focusing on Bayes' Theorem.

1. Probabilistic independence and the intersection of two events.

1.1. Recall that an event X is probabilistically independent of Y if and only if $Pr(X \mid Y) = Pr(X)$. An immediate consequence of this is that the probability that two probabilistically independent events X and Y both occur, i.e., $Pr(X \cap Y)$, is simply the probability of X multiplied by the probability of Y.

(1) **PROBABILITY OF INDEPENDENT X and Y:** If X and Y are independent, then $Pr(X \cap Y) = Pr(X) \times Pr(Y)$

Quick Proof. If X and Y are independent, then $Pr(X|Y) = Pr(X)$. $Pr(X|Y) = \dfrac{Pr(X \cap Y)}{Pr(Y)}$.
Thus, if $Pr(X|Y) = Pr(X)$, then $Pr(X \cap Y) = Pr(X) \times Pr(Y)$.

1.2. How do we calculate the probability of two events X and Y both occurring if they are probabilistically dependent? Well, we can't use (1). But we can use the definition of conditional probability. That is, given the definition of conditional probability, if we know $Pr(X|Y)$ and $Pr(Y)$, then we can use it to calculate the probability of $Pr(X \cap Y)$ using the fact that $Pr(X \cap Y) = Pr(X \mid Y) \times Pr(Y)$.

1.3. Here's a simple application of this. Suppose that $1/2$ of all the cats are male and that $3/10$ of all male cats are fat. How would we calculate the probability that a randomly selected cat is male *and* fat? Let M be the event of selecting a male and F be the event of selecting a fat cat. These two events are not independent. So, we calculate $Pr(M \cap F)$ by appealing to $Pr(M) = 1/2$ and $Pr(F \mid M) = 3/10$. We know that $Pr(M \cap F) = Pr(F \mid M) \times Pr(M)$. So, $Pr(M \cap F) = 3/10 \times 1/2 =$ **3/20**.

2. The first version of Bayes' Theorem

2.1. There are two important results which allow us to calculate the conditional probability of certain events, given other conditional and non-conditional probabilities. But first, why would we want to do that? Consider:

> Polly is interested in the reasons why people voted to leave the European Union. **52%** of the electorate **voted to leave** and **48% voted to remain**. Polly conducts a highly accurate survey and comes to two conclusions. First, **59%** of the electorate cited reasons pertaining to **constitutional sovereignty** as their main motivation **for voting the way they did**. Second, **21 out of 26** people **who voted to leave** cited reasons pertaining to **constitutional sovereignty** as their main motivation **for voting the way they did**. Polly asks:
>
> **Polly's Question:** Suppose I randomly select a member of the electorate. What's the probability that, given the person cites reasons pertaining to constitutional sovereignty as their main motivation for the way they did, they voted to leave? (Griffiths & Roberts, *Sets, Relations and Probability*)

What is Polly asking here? Let C be the event of randomly selecting someone from the electorate whose main motivation for voting was constitutional sovereignty and let L be the event of randomly selecting someone from the electorate who voted to leave. Polly wants to know $Pr(L \mid C)$. What does Polly already know? Polly already knows that $Pr(L) = 52/100$, and that $Pr(C) = 59/100$. Moreover, Polly knows from her survey that $Pr(C \mid L) = 21/26$. So, we need a way of calculating $Pr(L \mid C)$ using $Pr(L)$, $Pr(C)$, and $Pr(C \mid L)$.

2.2. Here, we can use the first-version of **Bayes' Theorem** to calculate $Pr(L|C)$:

(2) **BAYES' THEOREM FIRST VERSION:** $Pr(X|Y) = \dfrac{Pr(Y|X) \times Pr(X)}{Pr(Y)}$

Quick Proof. $Pr(X|Y) = \dfrac{Pr(X \cap Y)}{Pr(Y)}$. $Pr(X \cap Y) = Pr(Y \cap X)$. Thus, $Pr(X|Y) = \dfrac{Pr(Y \cap X)}{Pr(Y)}$. $Pr(Y|X) = \dfrac{Pr(Y \cap X)}{Pr(X)}$, so $Pr(Y \cap X) = Pr(Y|X) \times Pr(X)$. So, replacing $Pr(Y \cap X)$:

$$Pr(X|Y) = \frac{Pr(Y \cap X)}{Pr(Y)} = \frac{Pr(Y|X) \times Pr(X)}{Pr(Y)}$$

2.3. Let's apply this to Polly's question. Polly wants to know $Pr(L|C)$, knowing $Pr(L)$, $Pr(C)$, and $Pr(C|L)$. We know that $Pr(L) = 52/100$, $Pr(C) = 59/100$, and $Pr(C|L) = 21/26$. So, plugging into Bayes' Theorem:

$$Pr(L|C) = \frac{Pr(C|L) \times Pr(L)}{Pr(C)} = \frac{(21/26) \times (52/100)}{59/100} = \frac{42}{59}$$

3. The second version of Bayes' Theorem

3.1. There is a second version of Bayes' Theorem. This second version is useful for when you don't directly know $Pr(Y)$ but you do know $Pr(Y|X)$, $Pr(Y|\overline{X})$, and $Pr(X)$. The second version of Bayes' Theorem follows from the first version of Bayes' Theorem, given the following fact about probability theory.

(3) $Pr(Y) = (Pr(Y|X) \times Pr(X)) + (Pr(Y|\overline{X}) \times Pr(\overline{X}))$

Quick Proof. Left as an exercise for you!

Given (3), we can straightforwardly obtain the second version of Bayes' Theorem by replacing $Pr(Y)$ in the first version for $(Pr(Y|X) \times Pr(X)) + (Pr(Y|\overline{X}) \times Pr(\overline{X}))$. This gets us:

(4) **BAYES' THEOREM SECOND VERSION:** $Pr(X|Y) = \dfrac{Pr(Y|X) \times Pr(X)}{(Pr(Y|X) \times Pr(X)) + (Pr(Y|\overline{X}) \times Pr(\overline{X}))}$

3.2. Here's an example of the second version of Bayes' Theorem in action.

Question. There're two boxes, A and B. A contains 2 red balls and 8 yellow balls. B contains 7 red balls and 3 yellow balls. You randomly select a box (you don't know which) with equal probability and randomly select a ball from the box – it's red. What is the probability that you have box A? (Griffiths & Roberts, *Sets, Relations and Probability*)

We are being asked for the probability that you chose box A *given* that the ball randomly selected from the box is red. Let $A :=$ the event of selecting box A and $R :=$ the event of selecting a red from the box chosen. We want to calculate $Pr(A|R)$. The second version of Bayes' Theorem tells us:

$$Pr(A|R) = \frac{Pr(R|A) \times Pr(A)}{(Pr(R|A) \times Pr(A)) + (Pr(R|\overline{A}) \times Pr(\overline{A}))}$$

From the question, we know the following probabilities:

- $Pr(R|A) = \frac{2}{10}$ (there are 2 red balls in A out of a total of 10 balls)
- $Pr(R|\bar{A}) = \frac{7}{10}$ (if we don't select A, we select B; B contains 7 red balls out of a total of 10 balls)
- $Pr(A) = \frac{1}{2}$ (we randomly select a box with 'equal probability'; there are two boxes)
- $Pr(\bar{A}) = \frac{1}{2}$ ($Pr(\bar{A}) = 1 - Pr(A)$)

So, plugging all this into the second-version of Bayes' Theorem we get:

$$Pr(A|R) = \frac{Pr(R|A) \times Pr(A)}{(Pr(R|A) \times Pr(A)) + (Pr(R|\bar{A}) \times Pr(\bar{A}))} = \frac{\frac{2}{10} \times \frac{1}{2}}{(\frac{2}{10} \times \frac{1}{2}) + (\frac{7}{10} \times \frac{1}{2})} = \frac{2}{9}$$

4. Bayes' Theorem and Bayesian Epistemology

4.1. Suppose we take the probability of a certain event to measure our **degree of confidence**, or our **degree of belief** that such an event takes place. This is often known as a subjectivist interpretation of probability. A core claim of **Bayesianism** is that once you come to learn some new evidence, E, your degree of confidence in any hypothesis H should be the conditional probability of H *given* E, i.e., $Pr(H|E)$. Examples:

- For Bayesians, it follows from the example in 1.3 that, if you come to learn that a particular cat is male M, then your degree of confidence that the cat is fat F should be 0.3 because $Pr(F|M) = 0.3$.
- For Bayesians, it follows from the example in 3.2. that, if you come to learn that the ball drawn is red, then your degree of confidence that you have selected box A should be $2/9$ because $Pr(A|R) = 2/9$.

4.2. With this idea in mind, we can apply Bayes' Theorem to get some interesting philosophical results. Most famously, we can better understand what Hume had to say about miracles using Bayes' Theorem. Suppose you are told by a reliable witness that a man walked on water. Call this event E—the testimonial evidence. Call the event that the man did walk on water H—the hypothesis. The witness is reliable, so $Pr(E|H) = 0.9$ and $Pr(E|\overline{H}) = 0.01$; but the hypothesis is unlikely, say $Pr(H) = 1/1000000$. We can calculate the probability that the man did walk on water (H) given the testimonial evidence (E), i.e., $Pr(H|E)$:

$$Pr(H|E) = \frac{Pr(E|H) \times Pr(H)}{(Pr(E|H) \times Pr(H)) + (Pr(E|\overline{H}) \times Pr(\overline{H}))} = \frac{0.9 \times (1/1000000))}{(0.9 \times (1/1000000)) + (0.01 \times 999999/1000000}$$

So, $Pr(H|E) \approx 0.0001$. In other words, the probability that the man walked on water given the testimonial evidence of even a reliable witness is very small. Indeed, we can conclude a general point here. $Pr(H|E)$ will only be greater than 0.5, if $Pr(E|H) \times Pr(H) > Pr(E|\overline{H}) \times Pr(\overline{H})$. Unpacking this inequality: $Pr(H|E)$ will only be greater than 0.5 if $Pr(E \cap H) > Pr(E \cap \overline{H})$. This means that $Pr(H|E)$ will only be likely, if the probability of false testimony is less likely than the miracle itself. Thus Hume (1748) wrote:

> That no testimony is sufficient to establish a miracle, unless the testimony be of such a kind, that its falsehood would be more miraculous, than the fact, which it endeavours to establish

References

Hume, David (1748). *Enquiry Concerning Human Understanding; Section X: Of Miracles*.