Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

2.4 Bayesian updating

> 'When the facts change, I change my mind. What do you do, sir?'
> *John Maynard Keynes.*

Bayesian updating is the act of updating your (probabilistic) beliefs in light of new information. Formally named after Thomas Bayes (1701–61), for two events A and B, the simplest form of **Bayes' theorem** is:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}.$$

Example

Consider the probability distribution of the score on a fair die:

$X = x$	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Suppose we define the event A to be 'roll a 6'. Unconditionally, i.e. *a priori* (before we receive any additional information), we have:

$$P(A) = P(X = 6) = \frac{1}{6}.$$

Now let us suppose we are told that the event:

$$B = \text{even score} = \{2, 4, 6\}$$

has occurred (where $P(B) = 1/2$), which means we can effectively revise our sample space, S^*, by eliminating 1, 3 and 5 (the odd scores), such that:

$$S^* = \{\cancel{1},\ 2,\ \cancel{3},\ 4,\ \cancel{5},\ 6\} = \{2, 4, 6\}.$$

So now the revised sample space contains three equally likely outcomes (instead of the original six), so the **Bayesian updated probability** (known as a **conditional probability** or *a posteriori* probability) is:

$$P(A \mid B) = \frac{1}{3}$$

where '$|$' can be read as 'given', hence $A \mid B$ means 'A given B'.

Deriving this result formally using Bayes' theorem, we already have $P(A) = 1/6$ and also $P(B) = 1/2$, so we just need $P(B \mid A)$, which is the probability of an even score given a score of 6. Since 6 is an even score, $P(B \mid A) = 1$. Hence:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} = \frac{1 \times 1/6}{1/2} = \frac{2}{6} = \frac{1}{3}.$$

Suppose instead we consider the case where we are told that an odd score was obtained. Since even scores and odd scores are **mutually exclusive** (they cannot occur simultaneously) and **collectively exhaustive** (a die score must be even or odd), then we can view this as the *complementary* event, denoted B^c, such that:

$$B^c = \text{odd score} = \{1, 3, 5\} \quad \text{and} \quad P(B^c) = 1 - P(B) = 1/2.$$

So, given an odd score, what is the conditional probability of obtaining a 6? Intuitively, this is zero (an impossible event), and we can verify this with Bayes' theorem:

$$P(A \mid B^c) = \frac{P(B^c \mid A)\, P(A)}{P(B^c)} = \frac{0 \times 1/6}{1/2} = 0$$

where, clearly, we have $P(B^c \mid A) = 0$ (since 6 is an even, not odd, score, so it is impossible to obtain an odd score given the score is 6).

Example

Suppose that 1 in 10,000 people (0.01%) has a particular disease. A diagnostic test for the disease has 99% **sensitivity** (if a person has the disease, the test will give a positive result with a probability of 0.99). The test has 99% **specificity** (if a person does not have the disease, the test will give a negative result with a probability of 0.99).

Let B denote the presence of the disease, and B^c denote no disease. Let A denote a positive test result. We want to calculate $P(A)$.

The probabilities we need are $P(B) = 0.0001$, $P(B^c) = 0.9999$, $P(A \mid B) = 0.99$ and also $P(A \mid B^c) = 0.01$, and hence:

$$\begin{aligned} P(A) &= P(A \mid B)\, P(B) + P(A \mid B^c)\, P(B^c) \\ &= 0.99 \times 0.0001 + 0.01 \times 0.9999 \\ &= 0.010098. \end{aligned}$$

We want to calculate $P(B \mid A)$, i.e. the probability that a person has the disease, given that the person has received a positive test result.

The probabilities we need are:

$$P(B) = 0.0001 \qquad\qquad P(B^c) = 0.9999$$
$$P(A \mid B) = 0.99 \qquad \text{and} \qquad P(A \mid B^c) = 0.01$$

and so:

$$P(B \mid A) = \frac{P(A \mid B)\, P(B)}{P(A \mid B)\, P(B) + P(A \mid B^c)\, P(B^c)} = \frac{0.99 \times 0.0001}{0.010098} \approx 0.0098.$$

Why is this so small? The reason is because most people do not have the disease and the test has a small, but non-zero, **false positive rate** of $P(A \mid B^c)$. Therefore, most positive test results are actually **false positives**.

In order to revisit the 'Monty Hall' problem, we require a more general form of Bayes' theorem, which we note as follows. For a general **partition**[1] of the sample space S into $B_1, B_2, \ldots, B_n$, and for some event A, then:

$$\boldsymbol{P(B_k \mid A) = \frac{P(A \mid B_k)\, P(B_k)}{\sum\limits_{i=1}^{n} P(A \mid B_i)\, P(B_i)}}.$$

Example

You are taking part in a gameshow. The host of the show, who is known as Monty, shows you three outwardly identical doors. Behind one of them is a prize (a sports car), and behind the other two are goats.

You are asked to select, but not open, one of the doors. After you have done so, Monty, who knows where the prize is, opens one of the two remaining doors.

He always opens a door he knows will reveal a goat, and randomly chooses which door to open when he has more than one option (which happens when your initial choice contains the prize).

After revealing a goat, Monty gives you the choice of either switching to the other unopened door or sticking with your original choice. You then receive whatever is behind the door you choose. What should you do, assuming you want to win the prize?

Suppose the three doors are labelled A, B and C. Let us define the following events.

- D_A, D_B, D_C: the prize is behind Door A, B and C, respectively.
- M_A, M_B, M_C: Monty opens Door A, B and C, respectively.

[1] Technically, this is the division of the sample space into mutually exclusive and collectively exhaustive events.

Suppose you choose Door A first, and then Monty opens Door B (the answer works the same way for all combinations of these). So Doors A and C remain unopened.

What we want to know now are the conditional probabilities $P(D_A \mid M_B)$ and $P(D_C \mid M_B)$.

You should switch doors if $P(D_C \mid M_B) > P(D_A \mid M_B)$, and stick with your original choice otherwise. (You would be indifferent about switching if it was the case that $P(D_A \mid M_B) = P(D_C \mid M_B)$.)

Suppose that you first choose Door A, and then Monty opens Door B. Bayes' theorem tells us that:

$$P(D_C \mid M_B) = \frac{P(M_B \mid D_C)\, P(D_C)}{P(M_B \mid D_A)\, P(D_A) + P(M_B \mid D_B)\, P(D_B) + P(M_B \mid D_C)\, P(D_C)}.$$

We can assign values to each of these.

- The prize is initially equally likely to be behind any of the doors. Therefore, we have $P(D_A) = P(D_B) = P(D_C) = 1/3$.
- If the prize is behind Door A (which you choose), Monty chooses at random between the two remaining doors, i.e. Doors B and C. Hence $P(M_B \mid D_A) = 1/2$.
- If the prize is behind one of the two doors you did *not* choose, Monty cannot open that door, and *must* open the other one. Hence $P(M_B \mid D_C) = 1$ and $P(M_B \mid D_B) = 0$.

Putting these probabilities into the formula gives:

$$P(D_C \mid M_B) = \frac{1 \times 1/3}{1/2 \times 1/3 + 0 \times 1/3 + 1 \times 1/3} = \frac{2}{3}$$

and hence $P(D_A \mid M_B) = 1 - P(D_C \mid M_B) = 1/3$ (because also $P(M_B \mid D_B) = 0$ and so $P(D_B \mid M_B) = 0$).

The same calculation applies to every combination of your first choice and Monty's choice. Therefore, you will *always* double your probability of winning the prize if you switch from your original choice to the door that Monty did not open.

The *Monty Hall problem* has been called a **'cognitive illusion'**, because something about it seems to mislead most people's intuition. In experiments, around 85% of people tend to get the answer wrong at first. The most common incorrect response is that the probabilities of the remaining doors after Monty's choice are both 1/2, so that you should not (or rather need not) switch.

This is typically based on **'no new information'** reasoning. Since we know in advance that Monty *will* open one door with a goat behind it, the fact that he does so appears to tell us nothing new and should not cause us to favour either of the two remaining doors – hence a probability of 1/2 for each (people see only two possible doors after Monty's action and implicitly apply classical probability by assuming each door is equally likely to reveal the prize).

It is true that Monty's choice tells you nothing new about the probability of your *original* choice, which remains at 1/3. However, it tells us a lot about the other two doors. First, it tells us everything about the door he chose, namely that it does not contain the prize. Second, all of the probability of that door gets 'inherited' by the door neither you nor Monty chose, which now has the probability 2/3.

So, the moral of the story is to switch! Note here we are using updated probabilities to form a **strategy** – it is sensible to 'play to the probabilities' and choose as your course of action that which gives you the greatest chance of success (in this case you double your chance of winning by switching door). Of course, just because you pursue a course of action with the most likely chance of success *does not guarantee you success*!

If you play the Monty Hall problem (and let us assume you switch to the unopened door), you can *expect* to win with a probability of 2/3, i.e. you would win 2/3 of the time *on average*. In any single play of the game, you are either lucky or unlucky in winning the prize. So you may switch and end up losing (and then think you applied the wrong strategy – hindsight is a wonderful thing!) but in the *long run* you can expect to win twice as often as you lose, such that in the long run you are better off by switching!

If you feel like playing the Monty Hall game again, I recommend visiting:

`http://www.math.ucsd.edu/~crypto/Monty/monty.html`.

In particular, note how at the end of the game it shows the percentage of winners based on multiple participants' results. Taking the view that in the long run you should win approximately 2/3 of the time from switching door, and approximately 1/3 of the time by not switching, observe how the percentages of winners tend to 66.7% and 33.3%, respectively, based on a large sample size. Indeed, when we touch on statistical inference later in the course, it is emphasised that as the sample size increases we tend to get a more representative (random) sample of the population. Here, this equates to the **sample proportions** of wins converging to their **theoretical probabilities**.

Note also the site has an alternative version of the game where Monty does not know where the sports car is! Good luck!