More on conditioning and Mr. Bayes

Saad Mneimneh

1 Multiplication rule for conditioning

We can generalize the formula $P(A,B) = P(A|B)P(B)$ to more than two events. For instance, $P(A,B,C) = P(A)P(B|A)P(C|A,B)$. In general,

$$P(A_1, A_2, \ldots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot \ldots \cdot P(A_n|A_1, A_2, \ldots A_{n-1})$$

This is easy to verify by replacing $P(A_i|A_1, \ldots A_{i-1})$ with its equivalent expression $P(A_1, \ldots, A_i)/P(A_1, \ldots A_{i-1})$, as follows:

$$P(A_1) \cdot \frac{P(A_1, A_2)}{P(A_1)} \cdot \frac{P(A_1, A_2, A_3)}{P(A_1, A_2)} \cdot \ldots \cdot \frac{P(A_1, \ldots, A_{n-1})}{P(A_1, \ldots, A_{n-2})} \cdot \frac{P(A_1, \ldots, A_n)}{P(A_1, \ldots, A_{n-1})}$$

In the above expression, the numerators and denominators cancel each other and what remains is $P(A_1, \ldots A_n)$.

2 Conditioning on multiple events

Given an event B, consider a partition of the sample space into events $A_1, A_2, A_3, \ldots$.

```
---------------------
| A1  | A2   | A3   |
|     |      |      |
|   -------------   |
|   | B         |   |
|   |           |   |
|   -------------   |
|     |      |      |
---------------------
```

Now, $B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \ldots$. These events are exclusive and, therefore,

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \ldots \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \ldots \end{aligned}$$

.

Therefore, we conclude that:

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

3 Bayes' rule

From the definition of conditional probability and above, we have:

$$P(A_i|B) = \frac{P(B, A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)}$$

Bayes' rule is particularly useful when A_i is not observable but B is. For instance, assume we know the probability of A_i but we cannot observe the occurrence of such event. However, we can observe an event B, the occurrence of which depends on A_i, and assume we can easily compute $P(B|A_i)$. Then we can say something about the occurrence of A_i. $P(A_i)$ is called the prior, B the observation, and $P(A_i|B)$ the posterior.

Let's revisit the king's sibling paradox. We can imagine the following scenario.

```
                |
         --------------
        |              |
A1: same      A2: different    non-observable
        |              |
       1/2            1/2
        |              |
    -------        -------
    |     |        |     |
   1/2   1/2      1/2   1/2
    |     |        |     |
   MM    FF       MF    FM
   B              B     B      observable
```

Events A_1 and A_2 correspond to whether both children have the same gender or not, respectively. These events are not observable, but influence event B that at least one of the siblings is a male (observable). Note: physically, the process of having two children is not necessarily carried out this way, i.e. by deciding first whether they will have the same gender or not, but we can model it this way. Now,

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}$$

and from the diagram above,

$$P(A_1|B) = \frac{1/2 \cdot 1/2}{1/2 \cdot 1/2 + 1/2 \cdot (1/2 + 1/2)} = 1/3$$

Next, we will consider the examples of medical tests, colored balls and bins, evidence in the court room, biology of twins, and the famous Monty Hall.

4 Medical test

Consider a medical test that is positive with probability 0.99 if the patient has the disease. Moreover, if the patient does not have the disease, the test is positive with probability 0.05.

A patient is tested positive. What is the probability that he has the disease? (note that this is not 0.99)

Let A be the event that the patient has the disease, and A^c be its complement (not having the disease). Let B be the event that the test is positive. Using Bayes' rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

$$= \frac{0.99P(A)}{0.99P(A) + 0.05(1 - P(A))} = \frac{0.99P(A)}{0.05 + 0.94P(A)}$$

If the disease is rare, say $P(A) = 0.001$, then $P(A|B) \approx 0.02$. Therefore, we have a 98% chance of false positive. For such a rare disease, the test is not good enough.

5 Colored balls and bins

Consider a bin with 30 red balls and 10 black balls, and another bin with 20 balls of each color. A ball is picked from a bin with equal probability; therefore, if B_i is the event that bin i was chosen, $P(B_1) = P(B_2) = 1/2$. The ball is red, what is the probability that bin 1 was chosen?

Let R be the event that the ball is red, then:

$$P(B_1|R) = \frac{P(R|B_1)P(B_1)}{P(R|B_1)P(B_1) + P(R|B_2)P(B_2)}$$

$$= \frac{30/40 \cdot 1/2}{30/40 \cdot 1/2 + 20/40 \cdot 1/2} = 3/5 = 0.6$$

The result is expected since bin 1 contains more red balls. But which is really important, the number, or the fraction of red balls? (think about it).

6 Evidence in the court room

Consider using a DNA match as evidence for guilt. If G is the event that the person is guilty, and M is the event of a DNA match, then $P(M|G)$ is usually high (close to 1) and $P(M|G^c)$ is usually low (close to 0, a match between two random DNAs). This is why a DNA match is considered to be a strong evidence. We are interested in reaching a verdict of guilty or not guilty, given that evidence. Therefore, we need to compute $P(G|M)$ and $P(G^c|M)$. One possibility is to compute the log-odd ratio:

$$\log \frac{P(G|M)}{P(G^c|M)}$$

and decide based on whether the result is positive (guilty) or non-positive (not guilty). This is not really a new concept, if the logarithm is positive, then $P(G|M)/P(G^c|M) > 1$ and, hence, $P(G|M) > P(G^c|M)$. So essentially we are deciding based on which is more likely. One advantage of this strategy is that we do not have to compute $P(M)$ (the denominator in Bayes' rule) because it is a common factor in $P(G|M)$ and $P(G^c|M)$.

$$\frac{P(G|M)}{P(G^c|M)} = \frac{P(M|G)P(G)}{P(M|G^c)P(G^c)} = \frac{(1-\epsilon)P(G)}{\delta(1-P(G))} \approx \frac{1}{\delta}\frac{P(G)}{1-P(G)}$$

$P(G)$ is of course a controversial measure since one could argue that it is not feasible to determine the probability of crime in a given society (we may however have an estimate of it). Let's say $\delta = 1/10^6$, then $1/\delta = 1000000$.

$$\frac{P(G|M)}{P(G^c|M)} = 1000000\frac{P(G)}{1-P(G)}$$

Therefore, it takes a really clean society to overcome these odds. Most likely, $1/\delta$ will determine the verdict (guilty). That's the classical approach. However, if the crime rate is really low, there is a chance that the verdict will be not guilty. This is an example where the classical approach and the Bayesian approach lead to different results (later we will see Lindley's paradox).

7 Biology of twins

Twins can be either monozygotic (developed from a single egg) or dizygotic. It is always the case that monozygotic twins are of the same sex, whereas dizygotic twins can be of opposite sex. Denote monozygotic by M and dizygotic by D, and let B stand for boy and G stand for girl, then:

$$P(GG|M) = P(BB|M) = \frac{1}{2},\ P(GB|M) = P(BG|M) = 0$$

$$P(GG|D) = P(BB|D) = P(GB|D) = P(BG|D) = \frac{1}{4}$$

Now, assume that we conduct the following experiment: we sample many twins of the same sex, and discover (by looking for different features) that the probability they are dyzygotic is p. What is the proportion of dizygotic twins in the population.

$$P(D|\{GG, BB\}) = p = \frac{P(\{GG, BB\}|D)P(D)}{P(\{GG, BB\})} = \frac{1/2 \cdot P(D)}{1/2 \cdot P(D) + 1 \cdot (1 - P(D))}$$

Therefore, $P(D) = 2p/(1 + p)$.

8 Monty Hall

There are three boxes: A, B, and C. Only one of them contains money, but all three are equally likely to contain the money. We pick one randomly (with equal probability). Then, one of the remaining two boxes is opened to reveal that it is empty. Finally we are given the choice to stick to the box we have or switch to the remaining one. What is the best strategy? One argument is that all boxes are equally likely to contain the money, so it makes no difference whether we switch or not. Using Bayesian analysis, however, we can show that there is always a benefit in switching.

Let (x, y, z) represent the outcome of this game:

- x: the box containing to money
- y: the box we pick
- z: the box that is opened

For instance, we are interested in computing:

$$P(x = A|y = A, z = B) = P(x = A, y = A, z = B)/P(y = A, z = B)$$

Note that:

$$P(x, y, z) = P(x)P(y|x)P(z|x, y) = 1/3 \cdot 1/3 \cdot P(z|x, y) = 1/9 \cdot P(z|x, y)$$

$$\begin{aligned} P(y, z) &= \sum_x P(y, z|x)P(x) = \sum_x P(y|x)P(z|x, y)P(x) = \\ &= \sum_x 1/3 \cdot 1/3 \cdot P(z|x, y)] = 1/9 \sum_x P(z|x, y) \end{aligned}$$

Therefore,

$$P(x|y, z) = \frac{P(z|x, y)}{\sum_x P(z|x, y)}$$

$$P(x = A|y = A, z = B) =$$
$$\frac{P(z = B|x = A, y = A)}{P(z = B|x = A, y = A) + P(z = B|x = B, y = A) + P(z = B|x = C, y = A)}$$
$$= \frac{1/2}{1/2 + 0 + 1} = 1/3$$

Similarly, we find that:

$$P(x = A|y = A, z = B) = 1/3 \text{ (from above)}$$
$$P(x = B|y = A, z = B) = 0$$
$$P(x = C|y = A, z = B) = 2/3$$

The same result holds if we consider other (valid) permutations of x, y, and z. Therefore, there is always a benefit in switching. Another way to see this is by listing all the outcomes with their probabilities:

(A, A, B)	1/18	(B, B, A)	1/18	(C, C, B)	1/18
(A, A, C)	1/18	(B, B, C)	1/18	(C, C, A)	1/18
(A, B, C)	1/9	(B, A, C)	1/9	(C, B, A)	1/9
(A, C, B)	1/9	(B, C, A)	1/9	(C, A, B)	1/9

Clearly, $P(x = A|y = A, z = B) = P(A, A, B)/(P(A, A, B) + P(C, A, B)) = 1/3$.

It is also possible to view the Monty Hall paradox as a tree with three phases, the first of which is non-observable, while the last two are observable (in a similar way to the king's sibling paradox).

```
                               |
          -------------------------------------------
         |                     |                     |
        1/3                   1/3                   1/3
         |                     |                     |
         A                     B                     C
         |                     |                     |
    -----------           -----------           -----------
   |     |     |         |     |     |         |     |     |
  1/3   1/3   1/3       1/3   1/3   1/3       1/3   1/3   1/3
   |     |     |         |     |     |         |     |     |
   A     B     C         A     B     C         A     B     C
   |     |     |         |     |     |         |     |     |
  ---    |     |         |    ---    |         |     |    ---
 |   |   |     |         |   |   |   |         |     |   |   |
1/2 1/2  1     1         1  1/2 1/2  1         1     1  1/2 1/2
 |   |   |     |         |   |   |   |         |     |   |   |
 B   C   C     B         C   A   C   A         B     A   A   B
```