



© 1997–2009, Millennium Mathematics Project, University of Cambridge.

Permission is granted to print and copy this page on paper for non-commercial use. For other uses, including electronic redistribution, please contact us.

January 1998

Regulars

Solution to the taxi problem revisited



Let's write down the information that we are given using probability statements. First, we need to introduce some notation. When we write $P(A|B)$ we mean: the probability of event A *given that* (denoted by "|") the event B has occurred.

For the taxi problem we have

$$\begin{aligned} P(\text{witness says that the taxi is blue}|\text{taxi is blue}) \\ = P(\text{witness is correct}) = 0.8; \end{aligned}$$

this probability statement tells the police how *likely* the witness is to be correct about the taxi being blue. We are also told that $P(\text{taxi is blue}) = 0.85$; this probability statement describes the strength of belief that the police have in the hypothesis that the taxi is blue *prior* to the witness coming forward.

What really interests the police is

$$P(\text{taxi is blue}|\text{witness says that the taxi is blue}).$$

In other words, they want to know the probability that a blue taxi is involved in the crime given the data that they have from the witness. Bayes' theorem provides us with a way of finding this probability from the two known probabilities:

$$P(\text{witness says that the taxi is blue}|\text{taxi is blue}) \text{ and } P(\text{taxi is blue}).$$

In its simplest form Bayes' theorem can be written as:

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

for two events A and B , provided $P(B) > 0$.

The denominator $P(B)$ can be calculated from the formula:

$$P(B) = P(B|A)P(A) + P(B|\text{not } A) P(\text{not } A)$$

Solution to the taxi problem revisited

If you haven't met these formulae before, please don't worry. Just take them on trust. If you want to learn more about them, have a look at a book such as the one mentioned at the bottom of the page.

It is worth noting that Bayes' theorem allows us to reverse conditional probabilities: if we know $P(A)$ and $P(B)$, we can find $P(A|B)$ from $P(B|A)$ (and $P(B|A)$ from $P(A|B)$ as well).

To apply Bayes' theorem to the taxi problem let A be the event that the taxi is blue and B be the event that the witness says that the taxi is blue. We already know that $P(B|A) = 0.8$ and $P(A) = 0.85$. It is easy to work out $P(B|\text{not } A)$ and $P(\text{not } A)$:

$$\begin{aligned}P(B|\text{not } A) &= P(\text{witness says that the taxi is blue}|\text{taxi is not blue}) \\ &= P(\text{witness is not correct}) \\ &= 1 - P(\text{witness is correct}) \\ &= 1 - 0.8 \\ &= 0.2\end{aligned}$$

$$\begin{aligned}P(\text{not } A) &= P(\text{taxi is not blue}) \\ &= 1 - P(\text{taxi is blue}) \\ &= 1 - 0.85 \\ &= 0.15\end{aligned}$$

(It happens in this case that $P(B|\text{not } A) = 1 - P(B|A)$; please note, however, that this result is *not* generally true.)

We can now calculate $P(B)$:

$$\begin{aligned}P(B) &= P(B|A)P(A) + P(B|\text{not } A)P(\text{not } A) \\ &= (0.8 \times 0.85) + (0.2 \times 0.15) \\ &= 0.71\end{aligned}$$

Hence, $P(\text{blue taxi is involved}|\text{witness says that the taxi is blue})$

$$= P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.8 \times 0.85}{0.71} = \frac{0.68}{0.71} = 0.96$$

which is exactly the result given in the solution to the taxi problem by the contingency table approach used in Issue No. 2 (see "[The solution to the taxi problem](#)").

There are two key probabilities in the above formulation. The first is $P(\text{witness says that the taxi is blue}|\text{taxi is blue})$. Since the only *data* available to the police is the account given by the witness, we may think of this probability as $P(\text{data}|\text{taxi is blue})$. This probability is known as the *likelihood* of the data given the hypothesis that the taxi is blue, and represents how likely the data are if the taxi is blue.

The second key probability is $P(\text{taxi is blue})$. This probability is known as the *prior* probability that the taxi is blue, and represents the strength of belief that the police give to the taxi being blue before they learn of the

Solution to the taxi problem revisited

data. We have seen that Bayes' theorem enables us to find $P(\text{taxi is blue}|\text{data})$ from the likelihood and the prior probability. The probability $P(\text{taxi is blue}|\text{data})$ is known as the *posterior* probability that the taxi is blue because it is the probability that the taxi is blue *after* the data have been taken into account.

It is now simple to work out $P(\text{taxi is not blue}|\text{witness says the taxi is blue})$

$P(\text{taxi is not blue}|\text{witness says that the taxi is blue})$

$$\begin{aligned} &= P(\text{not } A|B) \\ &= 1 - P(A|B) \\ &= 1 - 0.96 \\ &= 0.04 \end{aligned}$$

The probability $P(\text{taxi is blue}|\text{witness says that the taxi is blue})$ is very much larger than $P(\text{taxi is not blue}|\text{witness says that the taxi is blue})$. Accordingly, if the witness says that the taxi is blue, the police should conclude that it is indeed blue.

Up to now we have considered the case when the data available is the statement of the witness that the taxi is blue. In the solution to the taxi problem the other case when the witness says that the taxi is not blue is also considered. It was found that:

$$P(\text{taxi is blue}|\text{witness says that the taxi is not blue}) = 0.59$$

and that

$$P(\text{taxi is not blue}|\text{witness says that the taxi is not blue}) = 0.41$$

The probability $P(\text{taxi is not blue}|\text{witness says that the taxi is not blue})$ turns out to be smaller than $P(\text{taxi is blue}|\text{witness says that the taxi is not blue})$. This means that, even though the witness says that the taxi is not blue, the police should conclude that the taxi is blue. Here the prior belief of the police about the colour of the taxi has swamped the data supplied by the witness. In both the above cases the police estimate the colour of the taxi as the one (between blue and not blue) that maximises the posterior probability.

Acknowledgements

This solution was written by Julian Stander. You may also be interested in his article "[Image analysis – a modern application of mathematics](#)" elsewhere in this issue.

Further reading:

Francis, A. (1988). *Advanced Level Statistics: An Integrated Course*. 2nd ed. Cheltenham: Stanley Thornes (Publishers) Ltd.



Solution to the taxi problem revisited

Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the NRICH and MOTIVATE sites.