Adding a Bayes Leaf to the Law

Ian Hunt
Chartered Statistician (Royal Statistical Society)

In *A Local Authority v B and C* [2014] EWHC 121 (Fam) Mr Justice Mostyn uses the laws of probability to avoid an absurdity he claims would result from the approach to fact-finding he was invited to adopt by counsel for the local authority. This note seeks to clarify and explore some issues related to probability that are raised in the case. It is intended to assist anyone who might be unfamiliar with the laws of probability and the application of Bayes' formula in particular. I consider that Mr Justice Mostyn's approach to assessing and combining probabilities accords with the laws of probability, as normally understood. However, his methods are not the only way of arriving at the probability of a disputed fact.

1. Overview

The key scenario to which Mr Justice Mostyn must assign a probability in this case (in order to find a fact on the balance of probabilities) is whether "the oxygen supply [a tap connected to an infant child] was deliberately turned off by the mother" .[1] I will call this an event and refer it as X.[2]

The probability of X could be estimated directly: a judge could assess the available evidence and background information, and reasonably form a probabilistic belief about whether or not event X happened. Mr Justice Mostyn does not take this approach in this case.

I assume that X is logically equivalent to the conjunction of two events, A and B, where A is "the tap was turned off" and B is the "the mother turned the tap off deliberately". Splitting X in this way is not trivial or without purpose, as the approach of Mr Justice Mostyn demonstrates.

In a hypothetical example at paragraph 37 of his judgment, Mr Justice Mostyn uses a piecemeal approach to find the probability of X. In the example, he calculates the probability that both A and B

---

[1] This is the phrase that is used by Mr Justice Mostyn in scenario iii, at paragraph 33.

[2] In this note I use the following shorthand expressions for different events:

X = "the oxygen supply [tap] was deliberately turned off by the mother";

A = "the tap was turned off";

B = "the mother turned the tap off deliberately"; and

C = "Nurse J turned the tap off accidentally" occurred together, as a way of determining whether X occurred. The formula he use equates the joint probability of events A and B occurring with the probability of A occurring, multiplied by the probability of B occurring when it is assumed that A is true. This relationship was first posited by Thomas Bayes[3] (I will call this Bayes' formula).

2. The laws of probability

I assume that the laws of probability are best defined by Kolmogorov's axioms.[4] These axioms have several practical implications. First, all probabilities must be between 0 and 1. Secondly, the total probability over a set of mutually exclusive events must sum to one.

Bayes' formula can be derived directly from the laws of probability.[5] So Mr Justice Mostyn's approach is, in principle, consistent with the laws of probability. In his hypothetical example at paragraph 37 of the judgment, Mr Justice Mostyn demonstrates how following Bayes' formula and using appropriate numerical values can ensure adherence to the laws of probability.

3. Example calculations

In this case all parties were agreed that there were three mutually exclusive scenarios that could explain what happened.[6] I will assume, for the sake of clarifying the calculations involved, that the following sentences are logically equivalent to those three scenarios:

i. the tap was NOT turned off; or
ii. the tap was turned off AND Nurse J turned off the tap by accident; or
iii. the tap was turned off AND the mother turned off the tap deliberately.

In terms of algebra (using straightforward notation and denotation[7]) these scenarios can be re-expressed as:

---

[3] "The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2d on supposition the 1st happens." Bayes T. (1763) An Essay towards solving a Problem in the Doctrine of Chances, Proposition 3 (Philosophical Transactions of the Royal Society of London, 53 (1763), 370 - 418).

[4] Kolmogorov A. (1933) Foundations of the theory of probability, New York: Chelsea.

[5] Calculating probabilities with Bayes' formula is very different to "Bayesianism", which is a nebulous doctrine combining probability assessments with decision theory.

[6] Paragraph 33 describes the three scenarios as follows: "i) The oxygen supply was not in fact turned off, and Nurse G is mistaken in believing that it was; or ii) The oxygen supply was accidentally turned off by J; or iii) The oxygen supply was deliberately turned off by the mother."

[7] Where A is "the tap was turned off", B is "the mother turned off the tap deliberately" and C is " Nurse J turned off the tap by accident".

i. not A; or
ii. A and C; or
iii. A and B.

Finally, I think that the following three formulas explain Mr Justice Mostyn's example calculations more clearly than the text at paragraph 37. Using the same values as the example, the laws of probability and Bayes' formula, the following relationships hold:

i. prob(not A) = 1 – prob(A) = 1 - .55 = .45
ii. prob(A and C) = prob(A) x prob(C given A) = .55 x .6 = .33
iii. prob (A and B) = prob(A) x prob(B given A) = .55 x .4 = .22

4. Alternative forms of Bayes' formula

As I construe it, Bayes' formula (used in ii. and iii. above) can take a variety of forms. For any A and B, these forms include:

a. prob(A and B) = prob(A) x prob(B given A)
b. prob(A and B) = prob(B) x prob(A given B)
c. prob(B given A) = prob (A and B) / prob(A)
d. prob(A given B) = prob (A and B) / prob(B)

Any form can be used, supposing that you have the required components i.e. you can calculate the left-hand-side by plugging in values on the right-hand-side. I will discuss forms a. and b. directly below, and talk briefly about forms c. and d. in section 8.

In this case, Mr Justice Mostyn finds it convenient to use form a. of Bayes' formula as an exemplar: he gives values for prob(A) and prob(B given A), and calculates prob(A and B). As noted above, X is logically equivalent to the conjunction of A and B. In other words, the probability of X is calculated on a piecemeal basis.

Form b. of Bayes' formula offers an alternative piecemeal approach. There are several steps. First, note that prob(A given B) would equal one (if the mother turned off the tap, then the tap is indeed turned off). Secondly, I assume that prob(B) can be estimated directly[8]. In other words, using form b. is actually the same as estimating the probability of X directly (which, if you analyse what we have called X, A and B, is what you would expect).

[8] I assume that Mr Justice Mostyn estimates prob(A) directly in his example at paragraph 37.

My practical point in this section, and I why I bothered with the perhaps painstakingly obvious algebra, is that the probabilities you might be interested in can be found by following a variety of different paths – you can choose whichever form of Bayes' formula that you want, provided that you have the requisite probabilities for the right-hand-side of the equation.[9]

5. Splitting an event into conjunctions of other events

The laws of probability and Bayes' formula are silent about whether you should assess probabilities directly, or use a piecemeal approach where events are split into what I call conjunctions of smaller events (or logic atoms, if you like).

In assessing probabilities for the scenarios in this case Mr Justice Mostyn appears to split "The oxygen supply was deliberately turned off by the mother" and "The oxygen supply was accidentally turned off by J" (ii. and iii. in paragraph 33). On the other hand, at least for the purposes of the example, he appears to directly assess a probability for "The oxygen supply was not in fact turned off, and Nurse G is mistaken in believing that it was" (i. in paragraph 33). Given the details of the case, and what was in dispute, I think this approach is reasonable.

6. Decomposing the prior and likelihood

The probability formulas in scenarios ii. and iii. above each have two parts. I will call these different parts the "prior" and the "likelihood". The priors are both denoted as "prob(A)". The likelihoods are the probabilities of B and C separately, where A is assumed to be given (or true). Finding appropriate values for these probabilities requires care, especially for the following two reasons.

First, the estimate of a prior probability should be made without regard to the particular facts inherent in the likelihoods. In this case, this entails assessing the probability that the tap was turned off, irrespective of the fact that the student nurse and the mother were present and had the opportunity to turn off the tap. I will discuss this example in more detail in the next section.

Secondly, assessing the likelihood probabilities requires hypothetical reasoning. In this case, probabilities must be assigned to events in which different people turned off the tap, assuming that the tap was in fact turned off (something that is not known for sure).

---

[9] You might uncover an inconsistency in your probabilistic beliefs, with respect to the laws of probability. You may then feel compelled to revise your beliefs, or "massage your probabilities".

7. Example of a prior

The appropriate prior probability associated with "the tap was turned off", as required in the probability formulas for scenario ii. and iii., is very different in form (and most likely value) to the appropriate probability for "the tap was turned off, given all the details in this case".

In reckoning the prior for "the tap was turned off", it seems reasonable to take account, as Mr Justice Mostyn did, of information concerning how the tap mechanism works, the reliability of such taps in general and how good Nurse G is at recognising which way taps are turned. This information is not inherent in the likelihood functions, where the tap is simply assumed to have been turned off.

Reckoning the probability for "the tap was turned off, given all the details in this case" would incorporate additional information. In particular, I assume that this probability would be affected by the facts that there had been a student nurse manipulating the tap mechanism, and that the mother, who was allegedly capable of filicide, had access to the tap.

An example will help illustrate the difference. Consider a hypothetical case in which the facts are the same as the real case, except that the only possible causes of the tap being turned off were a hospital cat passing by and a pot-plant swishing in the breeze. In the hypothetical case, the appropriate probability for "the tap was turned off", for use as a prior, should be exactly the same as in the real case; but, at the risk of stating the obvious, the probability for "the tap was turned off, given all the details in this case" would presumably be different.

I have laboured this point because in some cases both types of probability - one a prior and the other conditional on all the facts - could be of interest. They should never be confused.

8. An alternative approach

There is a different form of Bayes' formula which I think can adequately express the key issue in this case (it may also be used to express the conditional probability in the section directly above). The probabilities would relate to a hypothesis and the evidence for and against it: let the hypothesis, H, be "the mother turned off the tap" and the evidence, E, be "all the details in this case". The key detail, or piece of evidence, is that Nurse G thought the tap had been turned off. The appropriate form of Bayes' formula here[10] is:

prob(H given E) = prob(H) x prob(E given H) / prob(E)

---

[10] The formula shown is analogous in form to c. and d. in section 4 above. And it is, I believe, what most forensic scientists would recognise as Bayes' formula.

In this formula, the prior is prob(H), the likelihood is prob(E given H) and I would call prob(E) the "total evidence probability" (the probability of observing the evidence, no matter what happened).

This formula would work along the lines of the others: the left-hand-side can be calculated by plugging in appropriate numerical values on the right-hand side. I will not comment further on this version of Bayes' formula in this note, as it does not feature in Mr Justice Mostyn's judgment.

9. Binary transformations

I do not know the law well enough to comment properly on the transformation of probabilities into binary values according to "Lord Hoffman's binary method" (as referred to by Mr Justice Mostyn at paragraph 35). But I would like to express an opinion on a related point: it is not necessarily illogical, or contrary to the laws of probability, to adjust your probabilistic beliefs according to principles. Many meaningful principles that affect probabilities are conceivable, and they could be derived from sources such as legal rules or even pragmatic concerns.

For example, assume that I am (subjectively) reckoning a set of probabilities and that I am told that I must, according to some meaningful rule, fix a particular probability in this set equal to one (or zero). I can still always adhere to the laws of probability if I appropriately massage, change or revise some of the other probabilities within the set. The only conceptual difference these adjustments would make is that my probabilities, and the inferences that I base upon them, would be in a sense "conditional on following the rule".

Using Mr Justice Mostyn's example at paragraph 37 is perhaps more illustrative. Assume, for the sake of argument, that there *is* a principled reason[11] to set the probability for "The oxygen supply was not turned off" to zero. Only two acts are required to satisfy the laws of probability. First, using the notation from the example, since P1 is set to zero then Q1 must be set to one. Secondly, the inferences and probabilities subsequently derived must be calculated with these new values.[12]

However, requiring more than one probability to be set to zero or one is potentially problematic. For example, consider the probabilities of heads or tails coming up on a coin toss. I cannot, without violating the laws of probability, set both my probability of heads and my probability of tails to one (or to zero).

---

[11] I am not suggesting that there is any such principle.

[12] This perhaps demonstrates that the laws of probability give us very limited advice about what our probability assessments should be in any particular case.

A deeper discussion of the issues in this section would require consideration of legal precedents and the subtle differences between probability interpretations. This is beyond the scope of this note.

10. Conclusion

In this case Mr Justice Mostyn handles his probability calculations with adroitness. And his vim for probabilistic argumentation is refreshing – especially compared to attitudes such as the one expressed by Lord Justice Henry, who waved-off expert statisticians during the case in *R v Sally Clark* [2000] EWCA Crim 54 by saying "We don't need to hear them – it would only be argumentative. After all, it is hardly rocket science".[13]

I believe that probability assessment is never straightforward, but both statisticians and legal professionals have much to say about it. More dialogue between practitioners from these distinct disciplines would be interesting and worthwhile for all.

Ian Hunt CStat (Royal Statistical Society), MSc (LSE, London), MSc (City, London), PGDipStat (Otago)
ihunt@bunhill.co.uk

1 March, 2014

[13] Bell, Swenson-Wright and Tybjerg (2008) Evidence. CUP, page 125.