Blood Evidence

Bayes' Rule Practice Problem

December 12, 2019

Problem

Someone murdered a man earlier this year. The only evidence is some fresh blood at the scene that is not from the victim. The police interviewed several suspects. The police arrested the one suspect who has the same rare blood type found at the scene. The prosecution bases their case on the blood evidence and will use Bayes rule to calculate the probability that the accused did the crime.

Most judges and almost all people outside the statistics community do not understand conditional probabilities.

Though not the actual calculations, the next two illustrations are intended to acquaint the jury (and perhaps the judge) with Bayesian analysis.

Suppose any of 10,001 suspects could have been in contact with the victim at the time of his murder. A rare blood type occurs in about 1 in 1000 in the population. Figure 1 illustrates the odds.

Figure 1. Population and subset with the rare blood type. 10,000 people (10 sheets x 1,000 per sheet) plus one for the arrested suspect. One person in 1,000 has the rare blood type. So, on average, 10 people (upper-left group on the top sheet) of 10,000 would have the same blood type. The accused has the rare blood type and is the added person at the lower-right.[1]

Figure 2 illustrates a simple calculation without using Bayes' formula. The probability that the accused is guilty given that he has the rare blood type is 1/11 or 9.1%.

[1] Figure inspired by Norman, Fenton and Martin Neil, 2011, "Avoiding Legal Fallacies in Practice Using Bayesian Networks," *Australian Journal of Legal Philosophy*, v. 36, pp. 114-151, ISSN 1440-4982.

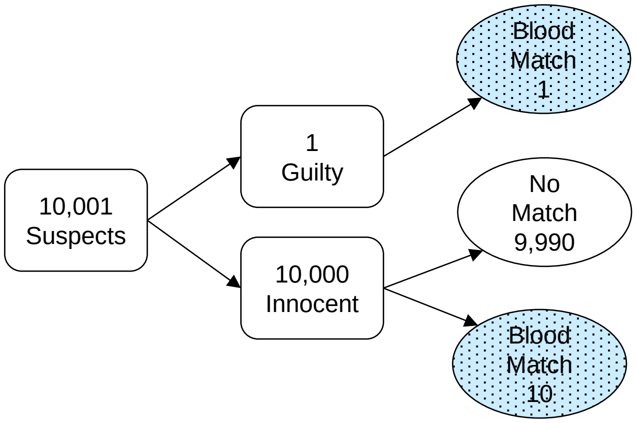


Figure 2. Classifying the possibilities. The accused and ten of the remaining 10,000 people have the rare blood type.[2]

We will change the blood type frequency in the exercise.

Exercise

Assume, as above, that 10,000 people could have been in contact with the victim.

The accused has rarest blood type, B-negative (B-), that occurs in only 1.5% of the population.

Using the common Bayes’ rule formula, calculate the probability that the accused is guilty.

Symbols and parameters:

N = 10,000 number of people in the area at the time

G = Guilty as accused

“B” = Blood match to the accused and 1.5% of the population

The quotation marks help distinguish the evidence information, “B” or “not B”.

Method 1. Bayes’ formula

The simplest form of Bayes’ rule, using this problem’s notation, is:

$$P(G\,|\,"B") = \frac{P("B" \bullet G)}{P("B")}$$

where the numerator on the right is $P("B" \bullet G) = P("B"|G)\,P(G)$

and, expanding the denominator, $P("B") = P("B"|G)\,P(G) + P("B"|\overline{G})\,P(\overline{G})$

[2] Figure inspired by Fenton, Norman, 2011, “Improve Statistics in Court,” *Nature*, v. 279, pp 36-37.

Method 2. Joint probability table (JPT)

Use the 3-step process as in class:

1. Set up a probability tree to represent data or subject matter experts (SME) judgments: G or not G as the leading node, then “B” or “not B” for the successor nodes.
2. Calculate the joint probabilities for each path in (1) and put these values into a JPT.
3. Redraw the tree in inverted sequence. Obtain the marginal and condition probabilities for the tree from the JPT. Circle the P(G|“B”) answer.

Workspace (resist peeking ahead at the solutions)

Solution 1 with Bayes' formula

$$P(G\,|\,\text{"B"}) = \frac{P(\text{"B"} \bullet G)}{P(\text{"B"})} = \frac{P(\text{"B"}|G)\,P(G)}{P(\text{"B"}|G)\,P(G) + P(\text{"B"}|\overline{G})\,P(\overline{G})}$$

$P(G) = \dfrac{1}{10{,}000} = 0.0001$ with no evidence.

$P(\overline{G}) = 1 - P(G) = 0.9999$

$P(\text{"B"}\,|\,G) = 1$ We know the killer had B- blood type.

$P(\text{"B"}\,|\,\overline{G}) = .015$ Fraction of the population with B- blood.

Substituting and solving the equation:

$$P(G\,|\,\text{"B"}) = \frac{P(\text{"B"}|G)\,P(G)}{P(\text{"B"}|G)\,P(G) + P(\text{"B"}|\overline{G})\,P(\overline{G})}$$

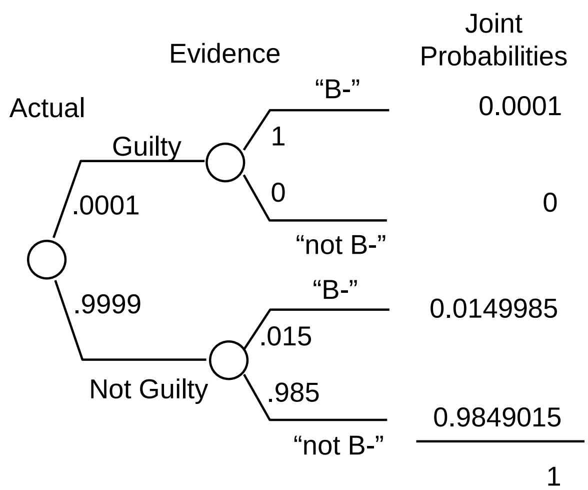
$$= \frac{1 \times 0.0001}{1 \times 0.0001 + 0.015 \times 0.9999} = \frac{0.0001}{0.0001 + .0149985} = 0.006623$$

This is 66 times greater than P(G) = 0.0001 with no evidence.

Recall, from the introductory example, if the blood type 1/1000 in the population, even more rare, then P(G|"B") = .9091. This likelihood is 9,091 times more likely than P(G) = 0.0001 with no evidence.

Solving with a JPT

Step 1. Judgments as captured from data and/or SME judgments



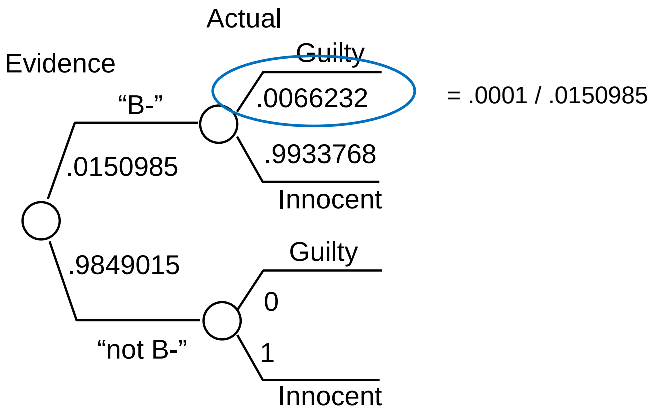
Try to finish if you have not already done so.

Step 2. Joint and marginal probabilities in the JPT

	"B-"	"not B-"	
Guilty	.0001	0	.0001
Not Guilty	.0149985	.9849015	.9999
	0.0150985	.9849015	

Next, invert the tree and obtain the probabilities using the table.

Step 3. Inverted tree



Obtain the probabilities at the upper-right node by normalizing the joint probabilities in the JPT's "B-" column. For example, P(G|"B-") = 0.0001 / 0.0150985 = 0.0066232. Again, the probability of guilt is 66 times greater than P(G) without the blood evidence.

These days, with modern forensic science, a drop of blood would provide more information than just blood type. Where possible, genome sequencing (DNA) and analysis of blood proteins would be much more conclusive.

Multiple pieces of evidence, unless conflicting, would further improve the confidence that the accused is the right guy. Examples of additional evidence (symptoms of guilt) include these items not of the victim:

- Shoeprint
- A hair
- Partial fingerprint
- Clothing fibers at the scene matching the accused's clothes
- Video or photo of a parked car that appears the same as the accused's
- Testimony of witnesses

Combining probabilities of multiple pieces of evidence, some correlated, usually requires software to solve the representation as a *Bayesian belief network*. This is a topic for another time.

■