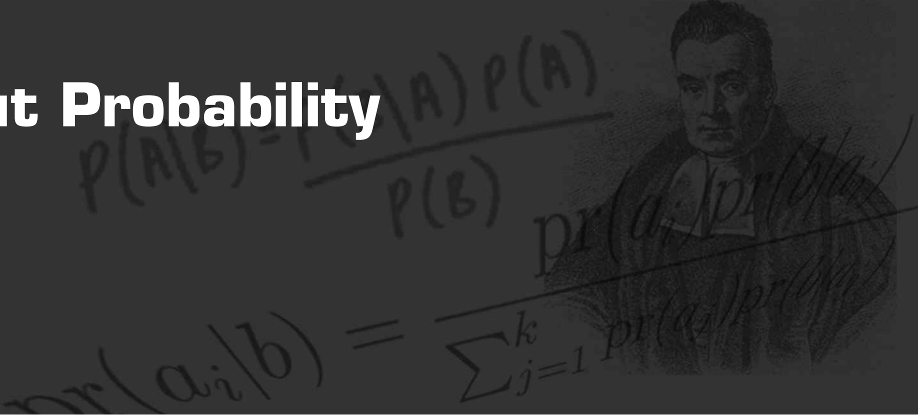
Bayes Rule Without Probability



Yap Von Bing

Look up Bayes rule or Bayes theorem, and you will find that it is almost always stated in terms of probability. Here, I describe a deterministic analogue which is easier to understand and is perhaps more suited for appreciating the applications. Starting with a realistic example from medical diagnostics, the reader is invited to work through several exercises leading to the derivation.

Problem. In a town of 200 people, 10% have a disease. A diagnostic test is very accurate: among the healthy, 70% test negative; among the sick, 80% test positive. Among the people who test positive, what percentage are sick?

A good approach is to use the 2×2 table shown below. The symbols are h: healthy, s: sick, $-$: tests negative, $+$: tests positive. The four cells defined by these symbols should contain the number of people in each category. The numbers sum to 200, which can be checked through the row or column sums. These sums appear at the margins of the table. Try filling up Table 1.

	$-$	$+$	*Row sum*
h			
s			
Column sum			200

TABLE 1. Counts.

There are many ways to proceed. The most efficient one is to figure the row sums, then the row counts, and finally the column sums. There are 70 people who test positive, of whom 16 are sick. So the answer is $16/70 \approx 23\%$.

The answer is surprising. If a person in the town tests positive, he is not confident that the test is correct, since less than 1 in 4 who test positive are sick. This happens whenever the diseases is rare, even if the test is very accurate (but not perfect). The practical lesson: for a rare disease, such as HIV infection, multiple positive tests are needed before diagnosis. However, since only 1 in 10 town residents is sick, the test helps: the proportion of sick people is higher in the positive group than in the negative group. Just not as much as we expect.

Clearly, the table method works for any test on any population. To get a formula, we convert the counts into proportions (Table 2). The proportion of people who test positive and are sick is 0.08, the proportion of people who test positive is 0.35, and the answer is their ratio. This is nothing new: the proportions are obtained from the counts by division by 200.

	$-$	$+$	*Row sum*
h	0.63	0.27	0.90
s	0.02	0.08	0.10
Column sum	0.65	0.35	1.00

Table 2. Proportions.

Now we introduce some shorthands. Let $\text{pr}(c)$ denote the proportion of people in category c. Here we have $\text{pr}(s) = 0.10$ and $\text{pr}(+) = 0.35$. The category can be more complicated: $\text{pr}(h, -) = 0.63$ is the (joint) proportion of people who are healthy and test negative. That h appears before $-$ is arbitrary: we can also write the same quantity as $\text{pr}(-, h)$. These symbols are all represented in Table 2. Next, the proportion of people who are sick among those who test positive, which we worked out to be about 0.23, is a conditional proportion. It is written $\text{pr}(s|+)$, read "conditional proportion of s given $+$". While a proportion refers to the whole population, a conditional proportion refers to a subpopulation specified by the category appearing after the $|$ symbol. Thus, $\text{pr}(s|+)$ is the proportion of sick people in the subpopulation who test positive, but $\text{pr}(+|s)$ is the proportion of people who test positive in the subpopulation who are sick, which is given as 0.8. Unlike the joint proportion, switching categories in a conditional proportion matters.

The notations allow us to restate the problem, as you should verify. *Given* $pr(s) = 0.1$, $pr(-|h) = 0.7$ *and* $pr(+|s) = 0.8$, *find the value of* $pr(s|+)$. Table 2 shows that given the three proportions, we can find 8 proportions (excluding the 1.00): there are general rules.

Addition Rule. In the example, it is easy to see that $\text{pr}(h) + \text{pr}(s) = 1$. There are $200 \times 0.1 = 20$ sick people, so there are $200 - 20 = 180$ healthy people, and $\text{pr}(h) = 180/200 = 0.9$. Similarly, $\text{pr}(-) + \text{pr}(+) = 1$. The general argument, or proof, has almsot the same structure. Let the town have n people, of whom $n(h)$ are healthy and $n(s)$ are sick. Since a person is either healthy or sick, but not both, $n(h)+n(s) = n$. Divide this equation by n. By definition, $\text{pr}(h) = n(h)/n$ and $\text{pr}(s) = n(s)/n$, so the proof is complete.

Exercise 1. Prove an addition rule for joint proportions: $\text{pr}(h, -) + \text{pr}(s, -) = \text{pr}(-)$. Similarly, $\text{pr}(h, +) + \text{pr}(s, +) = \text{pr}(+)$.

Multiplication Rule. $\text{pr}(h, -) = \text{pr}(h) \times \text{pr}(-|h)$. This rule implies that in the example, $\text{pr}(h, -) = 0.9 \times 0.7 = 0.63$. Similarly, $\text{pr}(s, +) = \text{pr}(s) \times \text{pr}(+|s) = 0.1 \times 0.8 = 0.08$.

Exercise 2. Prove the multiplication rule. To start, let n, $n(h)$ be as before, and let $n(h,-)$ be the number of people who are healthy and test negative.

Exercise 3. Prove an addition rule for conditional proportions: $\mathrm{pr}(h|-)+\mathrm{pr}(s|-)=1$ in two ways: (a) with the multiplication rule, (b) without.

Table 2 can now be filled up without going through Table 1. By addition rule, the second row sum is $\mathrm{pr}(s,-)+\mathrm{pr}(s,+)=\mathrm{pr}(s)$, given as 0.1, hence by the same rule, the first row sum is $\mathrm{pr}(h)=1-\mathrm{pr}(s)=0.9$. We have seen above that multiplication rule gives $\mathrm{pr}(h,-)=0.63$, the first number in the first row. Then addition rule implies $\mathrm{pr}(h,+)=\mathrm{pr}(h)-\mathrm{pr}(h,-)=0.90-0.63=0.27$, the second number. The second row is found similarly. Finally, by addition rule, the first column sum is $\mathrm{pr}(-)=\mathrm{pr}(h,-)+\mathrm{pr}(s,-)=0.63+0.02=0.65$. The second column sum is $\mathrm{pr}(+)=1-\mathrm{pr}(-)=1-0.65=0.35$, by addition rule.

The previous paragraph seems superfluous. But it is a good warm-up for deriving $\mathrm{pr}(s|+)$ from $\mathrm{pr}(s),\mathrm{pr}(-|h)$ and $\mathrm{pr}(+|s)$.

$$\begin{aligned}
\mathrm{pr}(s|+) &= \frac{\mathrm{pr}(s,+)}{\mathrm{pr}(+)} && \text{(multiplication rule)}\\
&= \frac{\mathrm{pr}(s,+)}{\mathrm{pr}(h,+)+\mathrm{pr}(s,+)} && \text{(addition rule)}\\
&= \frac{\mathrm{pr}(s)\mathrm{pr}(+|s)}{\mathrm{pr}(h)\mathrm{pr}(+|h)+\mathrm{pr}(s)\mathrm{pr}(+|s)} && \text{(multiplication rule)}\\
&= \frac{\mathrm{pr}(s)\mathrm{pr}(+|s)}{[1-\mathrm{pr}(s)][1-\mathrm{pr}(-|h)]+\mathrm{pr}(s)\mathrm{pr}(+|s)} && \text{(addition rule)}
\end{aligned}$$

Usually, Bayes rule looks like the third line. The fourth line shows the relationship more explicitly. It is worthwhile to work the derivation backwards, paying close attention to the justifications.

Bayes rule is efficient, but not the only way to find $\mathrm{pr}(s|+)$. By definition, $\mathrm{pr}(s|+)$ is a ratio of two integers, and can be expressed in other ways (first two lines of the derivation). If the given information fits, any of these three should be used instead. Bayes rule is like the Circle Line route from Kent Ridge to Serangoon. If you start from HarbourFront instead, it is fine to take the Circle Line via Kent Ridge (In mathematics speak, we *reduce* the new problem to an old one.), but a smarter way is the North East Line. The formula is a snapshot of the complete picture embodied in Table 2, which is sometimes useful to compute. For instance, to find both $\mathrm{pr}(s|+)$ and $\mathrm{pr}(h|-)$, you just need to compute two ratios. If you are a computer, you will happily store many instances of Bayes rule. But a human, as I think you are, should appreciate the pros and cons of both approaches, and their connection.

Bayes rule is a powerful tool for answering important practical questions. For which values of $\mathrm{pr}(s),\mathrm{pr}(-|h),\mathrm{pr}(+|s)$ is the test sensible, i.e., $\mathrm{pr}(s|+)>\mathrm{pr}(s)$, such as in the example here? If a test is sensible, is it true that $\mathrm{pr}(h|-)>\mathrm{pr}(h)$? More interestingly, it facilitates a quantitative exploration of the example. What do we mean by "surprise"? Under what conditions does it occur? Approaching all such questions via the table method is laborious.

Although Bayes rule was derived for a specific example, stripped to the logical bare bones, it applies for any variable with only two categories, which in this case is health status: healthy or sick. This naturally generalises to more categories. Suppose a variable has k categories $a_1, \ldots, a_k$, and let b be any category.

Exercise 4. Show that for $1 \leq i \leq k$, the conditional proportion of a_i given b is

$$\text{pr}(a_i|b) = \frac{\text{pr}(a_i)\text{pr}(b|a_i)}{\sum_{j=1}^{k} \text{pr}(a_j)\text{pr}(b|a_j)}$$

Finally, some words about probability. A probability is a proportion, but a proportion need not be a probability. In our town, $\text{pr}(s) = 0.1$ is the proportion who are sick, pure and simple. If we select a person at random, then the probability that this person is sick is 0.1. In contrast, proportions, even the conditional ones, are elementary descriptions of a population. Hence, I think the deterministic Bayes rule should be learnt first. As shown in our example, proportions are sufficient for understanding the paradoxical issue in medical diagnostics. Invoking probability here, as is done in many textbooks, muddles the issue, and can land the instructor in some trouble: What does "The probability that a person is sick is 0.1." mean, exactly? Elementary textbooks rarely offer a careful discussion of this innocuous-looking question.

More technically, the deterministic addition and multiplication rules can be used to prove the analogous rules in the frequency theory of probability. In particular, the conditional probability $P(A|B)$ has a frequency definition, and $P(A, B) = P(B)P(A|B)$ is a theorem. Many textbooks, as well as the H2 Mathematics Syllabus, has $P(A|B) = P(A, B)/P(B)$, which looks like a definition for conditional probability. This is necessary in the very general framework of Kolmogorov, but the frequency theory of probability is so important for applications, it makes sense to adopt the stronger set of axioms, at least for the beginner.

The author is an Associate Professor in the Department of Statistics and Applied Probability, National University of Singapore. He is also currently the assistant editor of this magazine.