

More Items on Conditional Probabilities

Math 394

1 Quick Note on “Total Probabilities”

Recall this important formula: if $\{E_1, E_2, \dots, E_n\}$ is a *partition* of the sample space (that is they have no points in common and $\Omega = \bigcup_{k=1}^n E_k$), then, for any event A

$$P[A] = \sum_{k=1}^n P[A|E_k] P[E_k]$$

You should note that this extends to the case of a *countable partition*, that is a countable sequence of disjoint events, whose union equals the sample space (that is, $\Omega = \lim_{n \rightarrow \infty} \bigcup_{k=1}^n E_k$, where the sequence of events on the right hand side is monotone increasing – the limit means that for any $\omega \in \Omega$, there is a $k = 1, 2, \dots$ such that $\omega \in E_k$). That follows from the *continuity axiom* for probabilities. Indeed, for such a partition,

$$A = A \cap \Omega = \lim_{n \rightarrow \infty} A \cap \left(\bigcup_{k=1}^n E_k \right) = \lim_{n \rightarrow \infty} \bigcup_{k=1}^n (A \cap E_k)$$

where the limit is monotone increasing. By continuity then

$$P[A] = \lim_{n \rightarrow \infty} P \left[\bigcup_{k=1}^n (A \cap E_k) \right] = \lim_{n \rightarrow \infty} \sum_{k=1}^n P[A|E_k] P[E_k]$$

where the right hand side is the definition of $\sum_{k=1}^{\infty} P[A|E_k] P[E_k]$.

2 Two Examples of Use of Bayes’ Rule

2.1 Noisy Transmission Channels

Assume you are at the receiving end of a channel transmitting a stream of bits (a potentially infinite sequence of 0s and 1s). Assume the channel is *noisy*, that is the bit you receive may be different from the one that was sent. Calling S_i the event that the value of the bit sent is i ($i = 0, 1$), and R_j the event that the received bit has value j ($j = 0, 1$), the noise can be formalized by specifying

$$P[R_1|S_0] = p, \quad P[R_0|S_1] = q$$

where p and q are the probability of a mistake, when sending a 0 or a 1 respectively.

The question now is: assuming you received a 0 or a 1, how confident are you that the transmission was correct? It is tailor made for Bayes' formula, however, as always in its applications, you need *a priori information* on the probabilities of 0 and 1 in your stream. Looking at the formula in our case, for the case you observed a 1,

$$P[S_1 | R_1] = \frac{P[R_1 | S_1] P[S_1]}{P[R_1]} \quad (1)$$

it is clear that the useful a priori information would be the composition of the stream at the source, that is $P[S_0]$ (let's call it s), and $P[S_1] = 1 - P[S_0]$. Since $P[R_1] = P[R_1 | S_0] P[S_0] + P[R_1 | S_1] P[S_1]$, (1) can be easily worked out:

$$P[S_1 | R_1] = \frac{(1 - q)(1 - s)}{ps + (1 - q)(1 - s)}$$

As an example, if $s = \frac{1}{2}$, this becomes

$$\frac{1 - q}{p + 1 - q}$$

and for, say, $p = 0.05, q = 0.01$, it equals $\frac{.99}{.05 + .99} = \frac{.99}{1.04} \approx .95$. You can work out a similar expression for $P[S_0 | R_0]$.

If you don't have access to the source, that information is not easy to come by. Thus, let's assume that, from an analysis of your stream, you have concluded that $P[R_0] = r$ (hence, $P[R_1] = 1 - r$). We need $P[S_1]$ to insert in (1), and, since we don't know its value we do what we always do in math in these circumstances: we give the unknown quantity a name (say, x), and play with it with the appropriate rules (in this case, basic algebra):

$$r = P[R_0] = P[R_0 | S_0] P[S_0] + P[R_0 | S_1] P[S_1] = (1 - p)(1 - x) + qx$$

Solving for x ,

$$\begin{aligned} r &= 1 - p + (q - 1 + p)x \\ x &= \frac{r - 1 + p}{q - 1 + p} \end{aligned}$$

Hence,

$$P[S_1 | R_1] = \frac{P[R_1 | S_1] P[S_1]}{P[R_1]} = \frac{(1 - q)(p + r - 1)}{(1 - r)(p + q - 1)}$$

You are encouraged to play with this formula, plugging in several choices for p, q, r . As an example, if $p = 0.1, q = 0.05, r = 0.8$,

$$P[S_1 | R_1] = \frac{0.95 \cdot (-0.1)}{0.2 \cdot (-0.85)} \approx 0.5588$$

which does not look like a very reliable channel. Repeating the very same steps for $P[S_0 | R_0]$, with the same numbers, will, instead, result in a probability of correct transmission of approximately 0.9926.

2.2 Testing For Rare Illnesses

This is a somewhat simplified, but still realistic, model for the screening of illnesses (or, for that matter, for screening for banned substances in sports). Suppose you are screening for an illness with a less than perfect test (and all real medical tests are less than perfect), which has a probability of resulting in a *false positive* (the test suggests wrongly that the patient has the illness) of p , and a probability of a *false negative* (the test fails to spot an existing illness) of n . If a patient takes the test and results positive, what is the probability that she or he is in fact ill?

This is a typical Bayes situation: calling S the event that the patient is ill, and A the event that the test is positive (indicates the patient is ill), we know

$$P[A|S^c] = n, \quad P[A^c|S] = p$$

and hence

$$P[A^c|S^c] = 1 - n, \quad P[A|S] = 1 - p$$

but we are interested in $P[S|A]$. To apply Bayes we need an a priori evaluation of, for example, the probability that the patient is ill, before we take a test – this could be a statistical estimate of the incidence of the illness in the population. One interesting result of our calculation will be how relevant this number turns out to be!

Let's say that we assume $P[S] = q$. Then,

$$P[S|A] = \frac{P[A|S] P[S]}{P[A]} = \frac{(1-n)q}{p(1-q) + (1-n)q} = \frac{(1-n)q}{p + (1-p-n)q} \quad (2)$$

(using the “total probability” formula). Again, you should play around with this formula, with various choices for q, p, n . For example, choosing

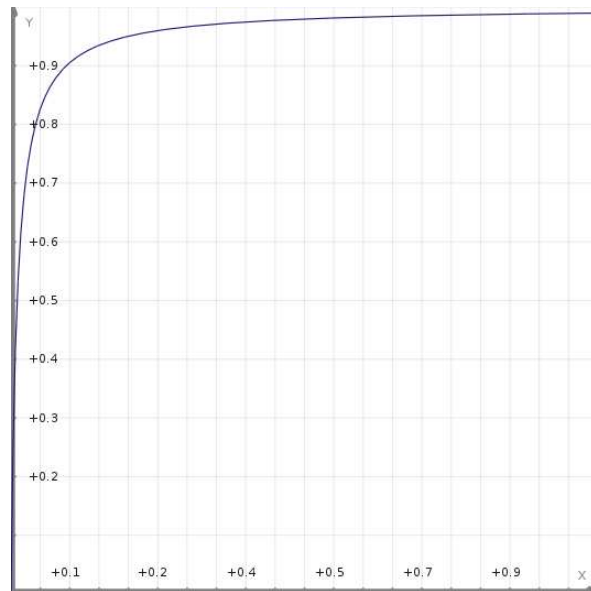
$$q = 0.005, \quad p = 0.01, \quad n = 0.05$$

we find that our probability is

$$\frac{0.95 \cdot 0.005}{0.95 \cdot 0.005 + 0.01 \cdot 0.995} \approx 0.32$$

While this number, is much larger than the 0.005 (0.5%) we started from, it is still small – it is still more likely that the patient is not ill! Most people, including prominent physicians, tend to overestimate this number wildly.

Here is a graph of (2), as a function of q , with $p = 0.01$, and $n = 0.05$, $\frac{95q}{0.995+95q}$ (the horizontal grid is spaced by 0.05):



As you can see, things become less dramatic as soon as the incidence of the illness rises above, say, 5%, when the probability is about 70%.

The Effect of A Second Test

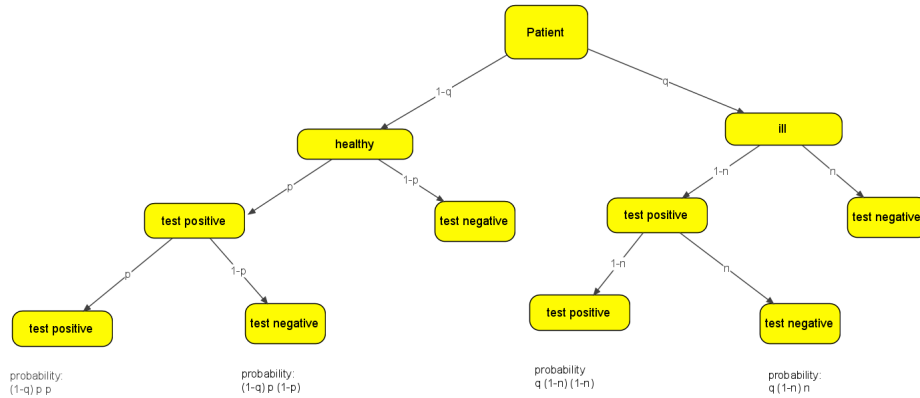
Suppose a patient tests positive. Since it is still highly unlikely that the patient is really ill, we administer a second similar test, with similar false positives and negatives (feel free to play around with variations – for example, the second test could be more precise, but more expensive, hence reserved as a second opinion). If the second test turns out positive too, how likely is it now that the patient is ill?

Please, note that for this second test to be useful, it must add genuine information. This is an important issue, which we are going to address now – so-called *independence*. Suppose, for example, that we use the very same test we used the first time. Whether this is a reasonable idea or not depends on the source of the randomness in the test's results. Specifically, if the false results are due to unpredictable behavior of the chemistry and physics involved in the test, the second time around, these behavior will produce a similarly random effect as the first time, but not the same. On the other hand, if the randomness stems essentially from the individual response of patients to the procedure, repeating the same test will most likely produce the same result, which makes the second test irrelevant. Technically, if we use the same notations above, and call A' the event that the second test is positive, in order for this to be meaningful, we need

$$P[A' | S \cap A] = 1 - p$$

and similar conditions for A'^c , A^c , and S^c . That is, the probability of false or correct result is unaffected by the outcome of the first test.

One way we represent these problems is with *graphs* – usually *trees*. Each alternative is represented by a branch of the tree. In our case,



Each branch of the graph is labeled with the conditional probability of its bottom node, given its upper node. Now, if both tests were positive, the probability of the patient being ill is given by

$$P[S|A \cap A'] = \frac{P[A \cap A' | S] P[S]}{P[A \cap A']}$$

From the graph we have that the numerator is equal to $(1-n)(1-n) \times q$, while the denominator is given by $(1-q)p \cdot p + q \cdot (1-n) \cdot (1-n)$. All in all,

$$P[S|A \cap A'] = \frac{q(1-n)^2}{q(1-n)^2 + p^2(1-q)}$$

With our original numbers, this is equal to

$$\frac{0.005 \cdot 0.95^2}{0.005 \cdot 0.95^2 + 0.01^2 \cdot 0.995} \approx 0.98$$

with a spectacular increase in the confidence we can have in our test.