ELECTRONOTES
1016 Hanshaw Road
Ithaca, NY 14850

APPLICATION NOTE NO. 393

December 10, 2012

THE BAYES EQUATION DENOMINATOR

In the last AN we gave the simple derivation of the Bayes formula (theorem, law, rule, equation) based on conditional probability. We will write down the standard form of the Bayes equation – here using H=Hypothesis and E=Evidence as these have a specific meaning which we can keep in mind as a convenient reference, unlike for example, A and B which are just letters. It makes no difference what letters we actually use, but here we have in mind we want to know what the probability is that a Hypothesis H is true given the Evidence E.

$$\mathrm{P(H|E)} = \frac{\mathrm{P(H)P(E|H)}}{\mathrm{P(E)}} \qquad (1a)$$

For example, the Hypothesis might be that a stranger (S) walked through our back yard given the Evidence that the dog barked (B). We seek the probability P(S|B), the probability of S given B.

$$\mathrm{P(S|B)} = \frac{\mathrm{P(S)P(B|S)}}{\mathrm{P(B)}} \qquad (1b)$$

Thus we can compute P(H|E) if we have P(H), P(E|H) and P(E). This is our start, but we also remember that P(E) was tricky in that it was in general (always?) a summation of at least two terms, since the evidence E could be the result of many different things; alternative Hypotheses. The general form becomes, instead of (1a):

$$\mathrm{P}(\mathrm{H_j}|\mathrm{E}) = \frac{\mathrm{P}(\mathrm{H_j})\mathrm{P}(\mathrm{E}|\mathrm{H_j})}{\mathrm{P}(\mathrm{H_1})\mathrm{P}(\mathrm{E}|\mathrm{H_1}) + \cdots + \mathrm{P}(\mathrm{H_j})\mathrm{P}(\mathrm{E}|\mathrm{H_j}) + \mathrm{P}(\mathrm{H_k})\mathrm{P}(\mathrm{E}|\mathrm{H_k}) + \cdots + \mathrm{P}(\mathrm{H_n})\mathrm{P}(\mathrm{E}|\mathrm{H_n})} \qquad (2a)$$

where we have n alternative hypotheses <u>for the same evidence E</u>, H_1 through H_n. It is important that the denominator includes all possible hypotheses, including H_j, and that the <u>sum of all probabilities of these hypothesis sum to 1</u>. This is a requirement distinct from saying that the denominator itself should sum to 1, which it does not do in general.

In our “dog that did bark” example, perhaps a Cat (C) went through the yard, or the dog was enthusiastic about it being time for his Food (F). So while the Hypothesis S relates to the stranger, P(B) gets contributions from the other two hypotheses C and F. Thus, writing the hypothesis as H=S (Stranger) and E=B (Bark):

$$\mathrm{P(S|B)} = \frac{\mathrm{P(S)P(B|S)}}{\mathrm{P(S)P(B|S)} + \mathrm{P(C)P(B|C)} + \mathrm{P(F)P(B|F)} + \mathrm{P(X)P(B|X)}} \qquad (2b)$$

Here, since there may be other reasons why the dog barked, we have included an X for other causes.

<u>NOTES RIGHT HERE</u>

(1) We also note that if we really were looking to find out if a stranger had walked through the yard, there might well be other elements of evidence, perhaps strange shoeprints in a child’s sandbox, or perhaps a news item in the paper about recent trespassers in the area. Each of these additional items may upgrade any value of P(S) we originally choose, or as upgraded by E=B.

(2) We note that we expect two terms or more in the denominator. If there is only one term (the same one as in the numerator), P(H/E)=1. The evidence and the hypothesis always occur together. This is not interesting except as it is approximated by multiple terms all but one of which are weak.

(3) It is quite common to have just two terms. For example, given that a student likes math, we might form a hypothesis that the student is a boy. The only alternative hypothesis seems to be that the student is a girl.

(4) Note well the replacement of P(E) in equation (1a) with a <u>sum</u> of terms, each of which is the <u>product</u> of an ordinary (marginal) probability and a conditional probability. It is possible in some or perhaps many cases that we know P(E) directly. For example, if we are running a test for drug use, we may know the number of positive test without (yet) determining how many were true and how many false.

PROBABILITIES BETWEEN 0 AND 1?

In our equations we have the symbol P for Probability everywhere and it is by definition that the probability is between zero (impossible) and one (certain). [We may speak of a probability in percent, like 23% likely, but we always put in numbers as decimals, or as ratios of percents.] The product of two probabilities (both marginal, both conditional, or as we have here, one of each in most cases) must accordingly also be between 0 and 1.

Note that equations (2a) or (2b) must give, by their very form, a result that is between 0 and 1. Less obvious, <u>but because the sum of all $P(H_i)$ in the denominator must be equal to one</u>, the total probability in the denominator must be less than one (and we should avoid 0). In originally denoting the denominator P(E), this restriction was implied. In fact, the total probability could be very small, and it does not "amplify" the ratio, as the numerator is always even smaller. The ratio of very improbable things could easily approach 1.

VOLTAGE DIVIDER !

<u>Note that equation (2b) looks very familiar to electrical engineers</u>. If we substitute variables we can make it look exactly like:

$$T = \frac{R_1}{R_1 + R_2 + R_3 + R_4} \tag{3}$$

which corresponds exactly to the resistive voltage divider (Fig. 1 - four resistors here - but it can be generalized) and we know that the attenuation factor or "transfer function" of this divider has exactly the property of being between 0 and 1.

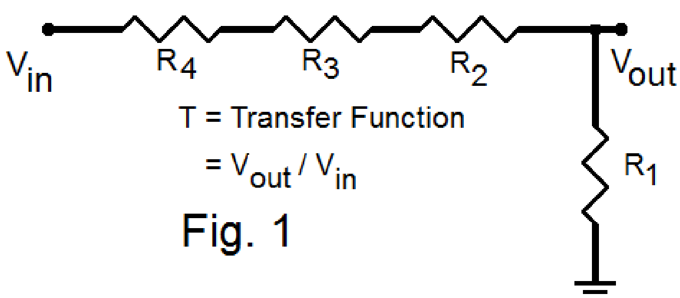

Fig. 1

Equations (2b) and (3) have the same general form, and we can easily relate to equation (3). It corresponds to a dimensionless voltage transfer ratio T that varies from 0 to 1. If R_2, R_3, and R_4 are all zero, T=1. If R_1=0, then T=0. Likewise, equation (2b) has both sides that are dimensionless and vary from 0 to 1. So this is helpful, but there are differences if we try to draw too tight an analogy.

LOOKING AGAIN AT THE BARKING DOG:

Above in equation (2) we worried about a barking dog as evidence that a stranger walked into our yard. We might suppose we could easily estimate (or at least make guesses) about the probabilities. The thing we can't do without care is assign the marginal probabilities. It would be, perhaps, natural to say that P(S)=0.3 (trespassers not unusual), P(C)=0.8 (lots of cats), P(F)=0.7 (dog is enthusiastic about suppertime) and P(X) = 0.2 (for good measure). We might make justifications for these, in general, but here we need to assign the probabilities in specific relationship to their representing the (total) causes of the dog barking. Here the probabilities add to 2.0 ! This needs to be fixed – we will just divide them by 2 in this case. It is not enough to contend, as may well be true, that the associated conditional probabilities might have contributed to reduce the denominator to less than 1. Further, note that unless all the conditional probabilities on the right side are 1, the denominator will add to a value less than 1.

What are our numbers now? Dividing our first guesses by two (in this case) we get P(S)=0.15, P(C)=0.4, P(F)=0.35, P(X)=0.1. Note that now P(S)+P(C)+P(F)+P(X)=1. Now for the conditionals. What are the probabilities that the evidence (Barking) would occur given the four hypotheses? Take P(B|S)=0.9 (the dog is a reliable watchdog), P(B|C)=0.9 (he like to bark at cats), P(B|F)=0.5 (half the time he just jumps wildly about, not barking, as his food is prepared), and P(B|X)=0.5 (who knows what X is anyway).

Plugging in the numbers into equation (2b) we find that:

$$\mathrm{P(S|B)} = \frac{(0.15)(0.9)}{(0.15)(0.9) + (0.4)(0.9) + (0.35)(0.5) + (0.1)(0.5)}$$

$$= \frac{0.135}{0.135 + 0.36 + 0.175 + 0.05} = \frac{0.135}{0.72} = 0.1875 \qquad (4)$$

Thus our notion that perhaps a stranger is trespassing about is slightly strengthened (from 0.15 to 0.1875) by the barking of the dog. We can also easily calculate P(C|B) by replacing the 0.135 in the numerator with 0.36, and we get P(C|B) = 1/2. The cat is the most likely suspect. We can try lots of different cases. For example, if we determine that the dog is not really that concerned about cats, so that P(B|C)=0.1 instead of 0.9, then P(S|B) goes from the just computed 0.1875 to 0.3375.

Note that the denominator, which is P(E) = P(B) only summed to 0.72. So the dog does not always bark – no surprise there. Further, the sum of the four conditional probabilities was 2.8. We are not at all concerned that the conditional probabilities are so large. Indeed, a conditional probability could easily be one under multiple hypotheses. For example, the evidence of a totally demolished house could be totally expected if we hypothesize a tornado, a gas explosion, or the crash of a UFO into the house. For these three hypotheses, $P(E,H_x)$ is probably 1. The contributions to the denominator, $P(H_x)(E|H_x)$, are not only not 1, but likely not all equal. Perhaps the probability of the tornado is 10 times larger than that of the explosion, and the chance of a UFO crash, well, substantially less! Further, it is $\sum P(H_x)$ that must equal 1. Thus we need to make sure the we identify and include all H_x. For example, perhaps the local Zoning Board ordered the house demolished. And so on.

We close here by thinking back to the math, which we saw juxtaposed to a problem of a resistor voltage divider. It is clear that the resistor can get bigger or smaller with corresponding effects on the transfer ratio T. Likewise, our “components” in the Bayes equation actually are the product of two probabilities, one marginal and one conditional. These can vary and cause the products to vary in obvious ways, with significant changes in the updated output probability.

<u>SO – WHICH IS RIGHT?</u>

Equation (1a) is often described as the “simple form” of equation (2a), but P(E) in the denominator of (1a) always means the same thing as the form in (2a). In equation (1a) we are concerned with only the probability of one selected hypothesis H, selected from among possible H_i, and must understand that P(E) is <u>still to be calculated</u> from all the additional $H_i = H_1 \ldots H_n$, or has been independently obtained (like counted or tabulated).