Bayes Theorem: A Brief Introduction

In its root form Bayes Theorem is stated as follows:

$$p(H|E) = \frac{p(E|H)*p(H)}{p(E|H)*p(H) + p(E|{\sim}H)*p({\sim}H)}$$

Pretty simple looking, isn't it? The p ( ) operator indicates the probability of the phenomenon or hypothesis (H, ~H) or the relationships (H|E, E|H, E|~H) inside the parentheses being true. The relationship delineated by the vertical bar "|" is read as the probability of the left hand variable being true *given* or *presuming* that the right hand variable *is* true. On the left hand side of the full equation, the term p (H|E) stands for the probability of hypothesis H being true *given* the existence of evidence E. This term is what we're looking for, the results of the equation or the *posterior probability*. In the case of the sunrise this would be the likelihood of the Sun rising tomorrow. This quantity is derived from the rest of the factors on the right hand side of the equation, collectively called the *prior probabilities* or the "priors." The priors are described as follows:

p (E|H) = The probability of finding evidence E assuming that hypothesis H is true. This is called the *inverse conditional probability* of E. It is generally perceived as a supporting factor for the posterior probability (the answer we're after) in that it addresses the prior relevance of the evidence to the phenomenon we are trying to assess. In the case of the Sun rising it would be the probability of the Sun rising today, given that it is a certainty that it will rise tomorrow, which would be all but certain or a value of close to 1.

p (H) = The prior probability of hypothesis H being true *independent* of the evidence E, or *the prior probability of* H. This factor is the baseline assessment of the validity of the phenomenon in question (either from prior calculations or educated guess) prior to any consideration of the current evidence E. In terms of our sunrise scenario, this value would be assigned subjectively or with values from previous calculations of sunrises, with the current evidence E, (today's sunrise) being ignored.

We could assign it a factor of 1 as being true, but that would be the same as saying that the Sun would always rise, rendering moot the probability assessment in the first place. If we have a situation that we anticipate a lot of continuing evidence by which to exhaustively repeat our probability calculations, we could give p (H) a very unlikely value of a thousand to one against the Sun rising (0.001) just to be on the conservative side and thus have a great deal of confidence if the equation eventually reaffirmed the likelihood of the Sun continuing to rise. But if we have only one or just a few data points by which to determine our probability (say we were just visiting earth for a week), we might assign the prior probability of the Sun rising a 50/50 chance (0.5) for fairness sake just to get a trend on the increase or decrease in the near-term prospects for the sunrise.

p (E|~H) = The probability of evidence E being true presuming that phenomenon H is *not* true. This is the *normalizing conditional probability*. This is usually perceived as a mitigating factor in that it addresses the prior *irrelevance* of the evidence to the phenomenon in question. In the sunrise scenario this would be the probability of the Sun rising tomorrow if the Sun never rose. This presumes you know something about the relationship of the evidence to the phenomenon. In the case of the sunrise the relationship is clearly direct and the value would be either extremely low or 0 (false). In other cases the relationship might be more enigmatic or indeterminate and the value would have to be subjectively assigned.

p (~H) = The prior probability that H is *not* true, independent of E, or the *normalizing alternative probability* (sometimes consisting of a summation of the probabilities of all mutually exclusive alternative phenomena). This factor is definitely a mitigating factor to the phenomenon in question in that it is a direct assessment of the weight of all conceivable alternative explanations for the phenomenon in question. Of course, H and ~H, representing all possible alternatives must add up to 1 or "true." If we conservatively set the prior probability of the sunrise (p (H)) at one in a thousand (0.001), we must necessarily set the prior probability of the Sun not rising tomorrow (p (~H)) at a thousand to one, or 0.999 (1-0.001=0.999)

Unambiguous Bayesian Prediction

For a pragmatic look at how Bayes theorem works on its simplest level, we will select a problem where the priors are unambiguously determined by the data. From E. S. Yudkowsky's exceptionally lucid explanation of Bayes theorem, we examine the mammography test results scenario http://yudkowsky.net/bayes/bayes.html . Imagine there is a screening test (mammography) for breast cancer which exhibits the following overall statistics:

> 80% of women at the age of 40 who have breast cancer will have a positive test results on their mammography. However, 9.6% of 40 year-old women who do *not* have breast cancer will show a false positive test results on their mammography. Given that the established breast cancer rate for 40 year-old women is 1%, what is the likelihood that a woman who has a positive test results on her mammography, actually has cancer?

Try and make an off-the-cuff guess at the answer. It turns out that even most doctors (17 out of 20) will guess too high on the probability of a positive test's indication of cancer.[1] Most people look at the large percentage (80%) of women who will test positive and compare it to the small percentage (9.6%) of false positives without much consideration to the comparative sizes of the pools of women to which these two percentages apply. Bayes theorem disallows this lack of consideration. Take another crack at the answer, given this new information.

Did you come up with something around 8%? It's not surprising if you didn't. Here is the informed calculation. The women who have a positive mammography results *and* actually have breast cancer amount to 80% of 1%, or 0.8%, of all the women who were tested. The women who do *not* have breast cancer, but still show a (false) positive mammography results, amount to 9.6% of 99%, or 9.504%, of all the women who were tested. Therefore, total positive mammography results amount to 0.8% + 9.504%, or 10.304%, of all the women who were tested. Of this percentage, the women who actually have cancer (0.8%) are in the distinct *minority*. The accurate figure for the probability of a woman who has a positive test results actually having cancer is 0.8 divided by 10.304 = 0.0776 or **7.76%.**

Of course, this begs the question of why anyone should have a mammography at all if so few positives actually have the disease. It should be noted for clarity's sake (not to mention public safety's sake) that such a test's utility is derived from its ability to deliver critical treatment to 80% of those few women who actually have breast cancer more than it is to reassure the 99% who do not have the disease. Mammography is just an inexpensive screening process. Further testing sorts out the false positives.

Had we started our assessment of mammography results by applying Bayes Theorem, we would have made the following unavoidably obvious selections for the priors: With C (the phenomenon) representing the actual presence of cancer and T_P (the evidence) representing a positive mammography test results we have, p $(C|T_P)$ = the *posterior probability* of a woman actually having cancer given a positive test results (the overall probability we are seeking). The priors are unambiguously indicated as follows:

p $(T_P |C)$ = The *inverse conditional probability* or the probability of a positive mammography test results (the evidence), given that a woman actually has cancer (the phenomenon) = 80% or 0.80.

p (C) = The *prior probability of* C or the probability of a woman actually having cancer prior to the mammography = 1% or 0.01 (the phenomenon).

p $(T_P |{\sim}C)$ = The *normalizing conditional probability* or the probability of a woman having a positive mammography test results (the evidence), given that she does *not* have cancer (the alternative phenomenon) = 9.6% or 0.96.

p (~C) = The *prior alternative probability* of C or the probability of a woman not actually having cancer = 99% or 0.99 (the alternative phenomenon).

Thus, we have:

$$p(C|\ T_P) = \frac{0.80 * 0.01}{0.80 * 0.01 + 0.09.6 * 0.99}$$

$$p\ (C\ |\ T_P) = 0.0776 = \mathbf{7.76\%}$$

As you can see, the priors were unambiguously clear cut and statistical in nature. In fact, the priors were so clear-cut that after a single application of the theory, the probability established is unambiguous and objectively acceptable by all reasonable criteria. In practice, however, Bayes theorem would more usefully applied early in the research of mammogram efficacy to derive, as soon as possible, our initial statistics we started with in our above example. For example, p (C) and p (~C) would be known (1% and 99%) and p $(C|T_P)$, p $(T_P |C)$ and p $(T_P |{\sim}C)$ would all be assigned initial subjective values and independently solved after each new patient's outcome was established (the new evidence), and cross-substituted in place of the subjective assignments to begin to converge on the relevant statistics.

Any corrections or improvements will be greatly appreciated.

Send all inquiries to: author@thegodofreason.com

[1] Casscells, Schoenberger, and Grayboys 1978; Eddy 1982; Gigerenzer and Hoffrage 1995