Predictions Derby – Candidates, Nominees, and Next President (Part1)

George Rebane

Ever since Nate Silver became a phenom prognosticator of some renown (*The Signal and the Noise*), I've considered starting a prognostication feature on *RR* that involves its readership. The methodology will be based on the Bayesian inference techniques on which I have reported here before (search *RR* for 'Bayes'). While Silver's book is primarily a self-promotion that claims to explain his methodology but doesn't, *RR* readers following my explication of Bayesian inference will be able to develop and use the methodology for countless purposes in their own private and commercial affairs. Look at it as part of the service ;-)

What really tipped my decision to give this a try on *RR* was twofold – Silver's disastrous performance at the FIFA World Cup predicting Brazil's defeat of Germany (at around 0.68), and some ideas for commercial applications. But the already ferocious political maneuvering by both parties preparing candidates for the 2016 elections was the real clincher to motivate the development of a participative predictions process (think of it also as 'crowd sourcing' and integration of evidence). So I started to push some squigglies with Bayes theorem and came up with an easily understood formula that should provide some entertainment, and that also has a rigorous decision theoretic basis – that is, it's not just idle fun.

The idea here is to predict who will run, and then who will be nominated as each party's presidential candidate. If there is enough interest, I'll extend the methodology to predict our next president. And all of this will be done by computing and updating our 'collective' beliefs about the future. The measure of belief is the probability that something yet unknown will come true. Belief calculations are based on a combination of hard probabilistic data, and subjective assessments all brought together under the celebrated formalism discovered long ago by the good Reverend Bayes (and rediscovered about 50 years ago when the methodology really took off).

So how will it work? Let's discuss the 'hat into the ring' predictions first. We can all come up with a list of potential candidates from each party and add to the list an 'Other' candidate. The fundamental Bayes approach takes a prior chance, odds, or probability, and combines it with recent evidence to calculate an updated or posterior chance, odds, or probability that now reflects the latest evidence. The prior probability summarizes all the knowledge we had about each candidate's propensity to throw their hat in the ring. And incorporating the latest evidence (news report, stump statement, lurid revelation, etc) will yield the updated posterior probability that again incorporates and summarizes all the previous knowledge we have about each candidate in the context of the hypothesis 's/he will throw his/her hat in the ring'. You can visualize the result as being a histogram with each bar labeled with a potential candidate's name and having a height from zero to one. As time goes on, the bars will get taller and shorter, and some may appear while others are removed.

Everyone can bring evidence to the table and give his assessment of that evidence in how it impacts each candidate being tracked. We may debate the assessment as others refute/modify it, but I will be the final arbiter because, well dammit, it's my blog (so there). But so as not to unreasonably piss off anyone, I will always do my best to found my adjudication. In any case, as opposed to Silver's close-to-the-chest methodology, here you can refuse to accept the *RR* prediction and run a parallel one yourself, and show us all up.

Let's get specific. Suppose today we have a pretty good idea that Abe, Betty, and Chuck are eyeing becoming candidates for their party. Of course, there may be others who will emerge, so we'll add the 'Other' category. And let's say that we really don't know whether A, B, C. or someone from Other will actually throw their hats in the ring and declare their candidacy. To express such ignorance in our beliefs we assign 0.5 (50-50 if you will) to the prior probabilities for each potential candidate. We remind ourselves that these probabilities do not have to sum to one or any other number since each person can independently make up their own mind, and they may even all declare to run or all decline to run.

Since it is certain that someone will wind up being a candidate if none of the identified ones choose to run, the only category that must be constrained then is the Other category. The probability that one or more candidates will emerge from that category will be at least the probability that all known potential candidates – here A, B, C – will decline to run. That probability can be computed from the three probabilities that each will run, and is simply the product of the complements of these three probabilities. We'll return to this point later. For now let's use the shorthand $P(HA)$ as the probability of *HA*, the hypothesis that A will run, and similarly for the others.

Let's set up some more shorthand so that the discussion can be compact and precise. Additionally let's let *E1, E2, …* and so on be pieces of evidence bearing on the hypotheses. For instance, *E4* may be the news that Betty was discovered to have been indicted for fraud ten years ago, but acquitted due to a legal technicality during her trial.

An Illustrative Example

With these labels under our belt we can now express chances or probabilities that are conditioned on something being true, like what is the chance of having a cloudy and misty morning on days during which it later rains. If you say, 'Oh, I'd give it 80%.' Then you're stating that in your experience and/or from your knowledge that about 4 out of 5 of subsequently rainy days started out cloudy and misty. And you might also reflect on it a bit more and conclude that about 1 out of 4 days when it didn't rain, the day started out similarly – cloudy and misty. This kind of assessment summarizes your experience when observing (the evidence of) a cloudy and misty January morning in Seattle.

In this little exercise we see that 'It will rain today' is the hypothesis (*H*) of interest since that may cause you to wear different clothes or reschedule an event planned for later in the day. And we'll use ¬H to denote the negation of the hypothesis (note the minus sign with the little hangy down part), i.e. that it will NOT rain today. So you have come up with two 'conditional probabilities' to encode or quantitatively express how the evidence *E* – i.e. observing a cloudy and misty morning – relates to a rainy and NOT rainy days. These probabilities are written as $P(E|H) = 0.8$, and $P(E|\neg H) = 0.25$ (read them as 'probability of observing *E* given that *H* is true' and 'probability of observing *E* given that *H* is NOT true' respectively).

So how are you going to use the *E* to calculate what the chances for rain are today? Well, you need a starting point for that rain chance before observing the evidence. A little online surfing of a weather site reveals the chance that it will rain on any given January day in Seattle is 73%. That makes the prior (to observing evidence *E*) probability of rain today to be 0.73. (Note that this also says that the complementary probability that it won't rain today is $P(\neg H) = 1 - 0.73 = 0.27$, since it is certain that it either will or won't rain today). From *Wikipedia* or *RR* you discover the Bayes formula, for incorporating new evidence *E* with prior knowledge *P(H)* that gives the updated probability P(H|E), is

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)} = \frac{0.80*0.73}{0.80*0.73 + 0.25*0.27} = 0.90 \ .$$

Where now *P(H|E)* is called the posterior probability (the new or updated knowledge) of rain given that you are witnessing a cloudy and misty morning. Putting in the numbers reveals that the chance that it will rain, given that we have a cloudy and misty morning, is now 90% from a pre-evidence chance of 73%. This illustrates how evidence is used to update prior knowledge *P(H)* to posterior knowledge *P(H|E)* that is now conditioned on having the evidence (*E*) at hand.

Just to make sure we understand that evidence doesn't only ratchet up prior probabilities, consider the alternative scenario where the morning dawned with bright sunshine and nary a cloud in the sky. In this case you may use such alternative evidence *E* to noodle out $P(E|H) = 0.20$ and $P(E|\neg H) = 0.75$. This simply says that in your experience/knowledge only about one in seven rainy days dawned bright and beautiful, and 3 out of 4 rainless days had a similar dawn. Now if we recalculate the probability that it will rain today using these data to represent the new evidence E, we have

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)} = \frac{0.20*0.73}{0.20*0.73 + 0.75*0.27} = 0.42$$

And here we see that a bright beautiful morning decreased our chances of rain from 73% to 42%, or about two chances in five of its raining later in the day. Conversely, the prior chance of no rain was 27%, and that chance has now increased to 58%. This illustrates how new evidence can alter or condition our prior (predictive) beliefs, expressed as the probability *P(H)* of a yet to be realized hypothesis (*H* = it will rain today) being true. And depending on the new evidence, it can either lend further support to the hypothesis – by increasing the chance of rain – or diminish support, thereby decreasing the probability that the hypothesis will bear out – here, decreasing the chance of rain.

What you have read here is an extremely powerful method of using uncertain/unreliable evidence (i.e. actual realworld evidence) to modify prior knowledge. Scientists tell us that this very paradigm is what appears to drive reasoning with uncertain knowledge in intelligent critters; in other words, our brains approximate the above calculations in the daily round of decision making. Of course, in the longer term, such 'Bayesian updating' is what has driven evolution over the eons. And now you know how to use it in your own decision making process.

With this under our belts, we can start talking about predicting which candidates will run for president in 2016. But before we do, let me introduce an even more remarkable aspect of the Bayes theorem, and that is the concept of likelihood. Likelihood in probabilistics is a ratio or a fraction that answers the question of 'how much more likely is it for this to happen when something is true than for it not to happen when the same thing is false?' We will apply it to the likelihood of a specific piece of evidence *E* happening. Formally, this likelihood is expressed as *L(E|H)* – 'likelihood of evidence E given that H will be true' - and is calculated from the fraction

$$L(E|H)=\frac{P(E|H)}{P(E|\neg H)} \ .$$

L(E|H) is simply the ratio of the probability of receiving such evidence E given that the hypothesis is true, divided by the probability of the evidence occurring when the hypothesis is false. In our example for *E* = {cloudy, misty morning} we have

$$L(E|H)=\frac{P(E|H)}{P(E|\neg H)}=\frac{0.80}{0.25}=3.2 \ .$$

Likelihoods come in really handy when we reason about our beliefs concerning specific future events coming to pass, since using likelihood allows us to bypass knowing or estimating the probabilities of receiving a given piece of evidence. All we need to do is substitute our estimate of how much more likely was the chance of the evidence occurring, with the future event coming true, vs its occurrence in the absence of the future event. So where does *L* fit into Bayes? Well, going back to 8th grade algebra, let's divide both the numerator and denominator of the Bayes formula above by *P(E|¬H).*

$$P(H|E)=\frac{\dfrac{P(E|H)P(H)}{P(E|\neg H)}}{\dfrac{P(E|H)P(H)+P(E|\neg H)P(\neg H)}{P(E|\neg H)}}=\frac{L(E|H)P(H)}{L(E|H)P(H)+P(\neg H)}, \quad \text{or}$$

$$P(H|E)=\frac{L(E|H)P(H)}{L(E|H)P(H)+1-P(H)}$$

In the second line we recognize that *P(¬H)* is just the complement of *P(H)* given by *P(¬H)* = 1 - *P(H)*. So there you have it, all you need to update your belief in *H* occurring is the likelihood of the evidence *E*, and the (prior) chance that *H* would occur before you became aware of *E*. Plug these values into the above formula, and you have reasonably (and correctly) incorporated new evidence to update your belief in *H*. Check out that you get the same answers for rain in Seattle.

Note that evidence which doesn't bear much on *H* drives its likelihood value close to unity. When *E* does not relate to *H* in one way or the other, then *L(E|H)* = 1. Put such evidence into the above equation and you quickly discover that *P(H|E) = P(H)*, in other words, the evidence did not impact *H* and your belief in *H* being true remains unchanged in the face of irrelevant evidence. We can see this in our rainy day example by considering something like *E* = {announcement that unemployment rate remained unchanged last month}. We don't have to estimate what the probabilities of that happening whether or not it will rain later in the day, because in any event $P(E|H)$ will be equal to $P(E|\neg H)$, and therefore $L(E|H) = 1$. You can play with the numbers yourself to demonstrate that the likelihoods of weakly supportive evidences have values near unity.

Now Back to Abe, Betty, and Chuck deciding whether to run

So let's return to the candidacies of Abe, Betty, and Chuck. Let the event that Abe will run be *HA*, and similarly for *HB* and *HC*. Suppose you start by having no idea whether any of them will run. We represent this by assigning $P(HA) = P(HB) = P(HC) = 0.5$, in other words you believe it's a 50-50 chance or a coin toss whether any of them will run. But on your way home you hear on the radio that Betty attended a political rally last night during which she received a prominent politician's endorsement. In his endorsement the politician cited his support for Betty's position on abortion which you know is opposed by Chuck, but Abe has yet to declare his position on the issue. That report on the radio is now your first piece of evidence E bearing on the candidacies. How should you incorporate E to update your belief about the candidacies?

You noodle a bit and conclude that $L(E|HB) = 4$, that with such an endorsement (especially regarding the abortion issue) consists of evidence that is connected four times more frequently (likely) with a candidate who will run, than with one who will not. Since Abe has yet to declare on abortion and has yet to get a major endorsement, you conclude that the same evidence is somewhat less frequently connected with a candidate who will ultimately run, meaning that $P(E|\neg HA) > P(E|HA)$ that drives the likelihood below unity, so you estimate $L(E|HA) = 0.9$. Finally, you know from talk going around that Chuck was really trying to get that plumb endorsement, and since he not only didn't get it, but also got dunned on his abortion stance, you feel that this *E* has pushed Chuck's $L(E|HC)$ down to about 0.5 – i.e. that such disappointing evidence is much more frequently connected with potential candidates who will ultimately give the election a pass. When you get home, you grab your calculator and quickly update your beliefs in the three candidates vying in the coming election.

$$P(HA|E) = \frac{L(E|HA)P(HA)}{L(E|HA)P(HA)+1-P(HA)} = \frac{0.90*0.50}{0.90*0.50+1-0.50} = 0.47,$$

$$P(HB|E) = \frac{L(E|HB)P(HB)}{L(E|HB)P(HB)+1-P(HB)} = \frac{4.00*0.50}{4.00*0.50+1-0.50} = 0.80,$$

$$P(HC|E) = \frac{L(E|HC)P(HC)}{L(E|HC)P(HC)+1-P(HC)} = \frac{0.50*0.50}{0.50*0.50+1-0.50} = 0.33,$$

You have now correctly integrated your previous knowledge about the candidates' intentions to run with the latest evidence you received on your way home, and you are ready to use these updated beliefs (as prior probabilities) with the next piece of news that will undoubtedly come your way as the election season draws near.