15 October 2014

INTRODUCTION TO MATHEMATICAL MODELING. BAYESIAN MODEL COMPARISON

David A. Meyer

Department of Mathematics
University of California/San Diego
La Jolla, CA 92093-0112
http://math.ucsd.edu/~dmeyer/; dmeyer@math.ucsd.edu

Background

A recent news article by Yglesias complained that because of the "hazy metaphysics of probability", even after the November election "we're never going to know which model is correct", referring to the many 'models'* giving a probability for a Republican takeover of the US Senate [1]. Taking two of them for the purposes of example, FiveThirtyEight's model, M_1 [2] gives a probability $p_1 = 0.58$ for this event, while Princeton Election Consortium's model, M_2 [3], gives a probability $p_2 = 0.51$.† While Yglesias drops hints that he knows better, he asserts, "in an epistemological sense, the way we check probabilistic statements is to run the experiment over and over again. Flipping a coin twice doesn't really prove anything.", and since there's only going to be one election, we can't determine which model is correct.

Bayes' Theorem

It's true, of course, that we can't be sure about which model is correct (other than to say, of course, that neither of them is!), but Bayes tells us how to think about this. Recall Bayes' Theorem:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)},$$

for any events A and B. We can use this to compute $\Pr(M_i \mid R)$, namely the posterior probability of M_i being correct given that the Republicans gain control of the Senate,

* These are not really models in the sense of this course, being primarily aggregations of poll results. But this discussion is useful to understand how the results of probabilistic models should be presented, and how they can be misunderstood.

† These models actually give probabilities for each Senate race; these numbers are the consequences for the Senate as a whole.

using $\Pr(R \mid M_1) = p_1$ and $\Pr(R \mid M_2) = p_2$. *Assuming we have no reason to prefer one model over the other initially, and that exactly one of the models is correct*, Bayes' Theorem gives:

$$\Pr(M_i \mid R) = \frac{\Pr(R \mid M_i)\Pr(M_i)}{\sum_j \Pr(R \mid M_j)\Pr(M_j)} = \frac{p_i}{p_1 + p_2},$$

since we have assumed $\Pr(M_1) = \Pr(M_2)$. Thus if the Republicans win control of the Senate we should increase the probability we assign to the correctness of FiveThirtyEight's model from 0.5 to about 0.53, while lowering the probability of the Princeton Election Consortium's model to 0.47. Conversely, if the Democrats retain control of the Senate, these probabilities would be reversed.

What if we assign different prior probabilities to the correctness of these two models, say m_1 and $1 - m_1$, respectively? Then

$$\Pr(M_1 \mid R) = \frac{p_1 m_1}{p_1 m_1 + p_2(1 - m_1)} \quad \text{and} \quad \Pr(M_2 \mid R) = \frac{p_2(1 - m_1)}{p_1 m_1 + p_2(1 - m_1)},$$

so the posterior probability of M_1, given that the Republicans take control of the Senate, is greater than that of M_2, provided

$$p_1 m_1 > p_2(1 - m_1) \quad \Longrightarrow \quad m_1 > \frac{p_1}{p_1 + p_2} \approx 0.47.$$

So as long as our prior belief in the Princeton Election Consortium's model is not too great, a Republican win will lead us to assign a larger posterior probability of correctness to FiveThirtyEight's model.

References

[1] M. Yglesias, "The real problem with Nate Silver's model is the hazy metaphysics of probability", *Vox* (11 October 2014); http://www.vox.com/2014/10/11/6949965/nate-silver-sam-wang-election-models.

[2] "FiveThirtyEight's Senate Forecast", http://fivethirtyeight.com/interactives/senate-forecast/.

[3] S. Wang, "PEC switching (as planned) to short-term forecast", http://election.princeton.edu/2014/09/30/pec-switching-as-planned-to-short-term-forecast/.

