Appendix for "Updating Bayesian(s): A Critical Evaluation of Bayesian Process Tracing"

Sherry Zaks

A Iteration in Practice

The odds-ratio form of Bayes' rule provides a convenient heuristic for thinking about how a piece of evidence affects our confidence in one hypothesis relative to another, but it has clear shortcomings in the context of facilitating iterative research. I demonstrate the problem with an example. Putting aside other possible issues with the Bayesian approach, let's assume optimal research conditions: we have two mutually exclusive and exhaustive hypotheses, H_M and H_R, and evidence E_1 bears on both hypotheses in opposite ways.

The explicit Bayesian approach instructs us to assign probabilities to represent the prior on each hypothesis and likelihood of finding E_1 conditional on H_M and then H_R. For the sake of simplicity I assign equal prior probabilities of 0.5 to each hypothesis. Since E_1 strongly points to—and is thus "unsurprising under"—H_M, I assign a likelihood of 0.7 to $P(E_1|H_M)$. On the flip side, since E_1 contradicts H_R, there is a low probability of finding E_1 in a world where H_R is true, so I assign $P(E_1|H_R) = 0.1$. These quantities give us the following equation:

$$\underbrace{\frac{P(H_M|E_1 I)}{P(H_R|E_1 I)}}_{\text{Posterior}} = \underbrace{\frac{P(H_M|I) = 0.5}{P(H_R|I) = 0.5}}_{\text{Prior}} \times \underbrace{\frac{P(E_1|H_M I) = 0.7}{P(E_1|H_R I) = 0.1}}_{\text{Likelihood}} \tag{1}$$

$$\frac{P(H_M|E_1 I)}{P(H_R|E_1 I)} = \frac{0.5}{0.5} \times \frac{0.7}{0.1} = \frac{0.35}{0.05} \tag{2}$$

Accordingly, this piece of evidence increases our confidence in H_M relative to H_R seven-fold. The question that follows is what we do next. Since what matters is the ratio—rather than the absolute probabilities in the numerator and denominator—the logical next step in an iterative process is to use the posterior odds to inform analysis on the next piece of evidence (i.e. rather than using equal priors, use priors that reflect the 7:1 odds in favor of H_M).

However, the scholars advocating the odds-ratio form of Bayesian analysis do not take this step, which calls into question how systematic the iteration process really is. In applying the Bayesian approach to Kurtz's state-building work, Fairfield and Charman begin with equal probability on two hypotheses—H_{RC} (resource curse) and H_W (welfare)—and a single piece of evidence, E_1. Since E_1 "strongly favors the resource-curse hypothesis" the authors conclude that "the likelihood ratio is large and it significantly boosts our confidence in the resource curse hypothesis."[1] Yet, as they move to analyze the next piece of evidence, they "keep equal odds on H_R versus H_W," thereby contradicting their updated probabilities on the two hypotheses.[2]

[1] Fairfield and Charman (2019, 159)

[2] (2019, 162)

The decision to keep equal odds on the two "competing" hypotheses after finding evidence that only supports one raises many questions about how to implement this method in practice. When should researchers update their priors and when should they not? What factors should drive this decision? How does the decision to not update affect the final results? If the goal is to formalize the process of updating our confidence in hypotheses as we move through the evidence, advocates of the method must not only provide guidelines for the process of updating (or not), but they must also practice what they preach.

Iteration with 3+ Hypotheses

The process of iteration is further complicated when a third hypothesis is added into the mix. By restricting our analyses to the odds-ratio form of Bayes' rule, the probabilities in the numerator and denominator of the posterior cannot be isolated for each individual hypothesis. For example, the analysis in Equation 2 is interpreted as follows: the evidence E_1 increases our confidence in H_M seven-fold relative to H_R. Even if updating with respect to H_M and H_R were consistent, the existing method leaves researchers in the dark about how to handle testing a third hypothesis, H_C.

Researchers cannot merely pluck the 0.35 and use that as the new prior on H_M, because that quantity does not reflect our updated confidence in H_M alone. The intuition underlying the problem is that since we're always multiplying fractions, even the most supporting evidence is going to result in a numerator in the posterior that is smaller than the numerator in the prior. In this example, our prior confidence in H_M is 0.5, then we found strong support for H_M with E_1, but the numerator in the posterior is 0.35. If we just used the updated numerator, the analysis would proceed as though we were less confident in H_M.

To be sure, proponents are not advocating that researchers take this step; the problem is that they remain silent on how to proceed. If we then want to analyze H_M relative to H_C, we have three options:

1. One option is to analyze how E_1 affects H_M relative to H_C by beginning that analysis with equal odds placed on the two hypotheses. The choice to revert to equal odds seems strange because we already know E_1 supports H_M. Though, it could help avoid bias from double-counting evidence.

2. A second option is to analyze how E_1 affects H_M relative to H_C by beginning that analysis with a slightly higher prior on H_M. This choice seems consonant with respect to updating our confidence in H_M, but problematic in that it double-counts the effect of E_1.

3. A third option is to analyze H_M relative to H_C by examining a different piece of evidence entirely, E_2. This choice is also difficult to justify because then researchers must make the case for why they chose a given piece of evidence for one analysis, but not another.

Moreover, if the right way forward is to choose either the first or second option (i.e. to continue analyzing how E_1 bears on all hypotheses), putting aside the issue of choosing the "correct" prior, the number of analyses researchers must conduct increases by an order of magnitude for each additional hypothesis. Specifically, for a given project, the number of analyses the researcher must conduct is a combinatorics problem. Thus, if she is only testing 2 hypotheses, she must conduct as many analyses as there are pieces of evidence, E_n. If she is testing 3 hypotheses, she must conduct $C_{(3,2)} = 3$ analyses per piece of evidence. If she is testing 4 hypotheses, she must conduct $C_{(4,2)} = 12$ analyses per piece of evidence. Even if not all pieces of evidence demand explicit testing, analyses will likely become intractable very quickly.

Currently, the Bayesian literature lacks instructions on how to proceed iteratively in the context of the odds-ratio analyses, how to deal with more than two hypotheses, and how to interpret and integrate results of these analyses. Ultimately, the computations quickly become intractable and the complexity and ambiguity raise questions of whether and to what extent this procedure is overshadowing good qualitative interpretation.