Weight of Evidence, Evidential Completeness and Accuracy

Rafal Urbaniak and Marcello Di Bello

1 Introduction

Suppose we want to represent our uncertainty about a proposition in terms of a single probability that we assign to it. It is not too difficult to inspire the intuition that this representation does not capture an important dimension of how our uncertainty connects with the evidence we have or have not obtained. In another paper we have already argued that to properly represent this aspect, we need to turn to second-order probability. Here, we are pushing these ideas further. Suppose the mean or median probability of an estimate somehow crudely represents the balance of probabilities. How do you we conceptualize measuring the other dimension, the amount of evidence obtained?

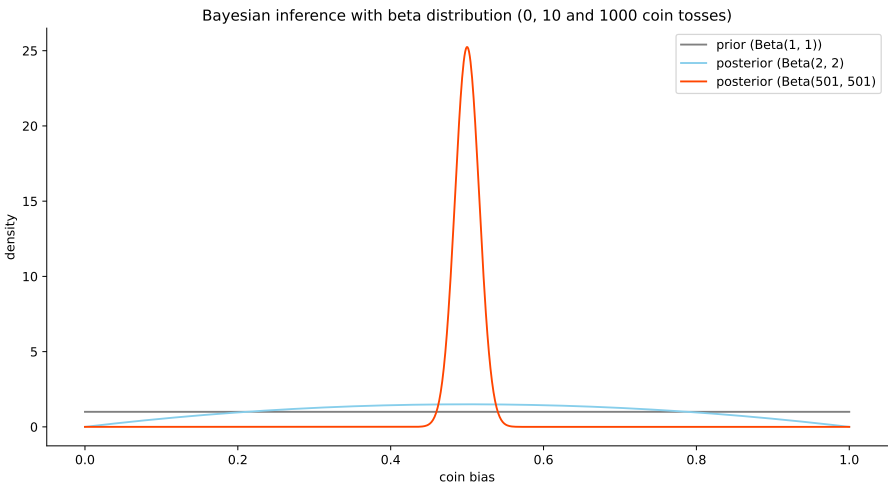
Chapter?

Add a structure descr etc.

```
2.5359400011538504
```

In a 1872 manuscript of The Fixation of Belief (W3 295) C. S. Peirce gives the following example.

" When we have drawn a thousand times, if about half have been white, we have great confidence in this result. We now feel pretty sure that, if we were to make a large number of bets upon the color of single beans drawn from the bag, we could approximately insure ourselves in the long run by betting each time upon the white, a confidence which would be entirely wanting if, instead of sampling the bag by 1000 drawings, we had done so by only two. "



How much information a distribution contains will be measured in its entropy. `H = -sum(p * log(p))`. We want to normalize it to make insensitive to our grid size choice as long as it is kept fixed, and we want it to be scaled between 0 and 1 for interpretability. So we use $1 - \frac{H(p)}{H(uniform)}$

```
0.0 0.01825506826172696 0.39529270774644865
```

Now, in application we want to also compare the prior to the posterior, answering the question of how much **more** confident we are after obtaining the evidence as compared to our point of departure. Notably, the prior doesn't have to be uniform. Here, one conceptualization (which has the feature of being on the same scale) is simply the difference in weights of the distributions.

```
# imagie our prior is in fact obtained by having seen 1 head and 1 tail already.

def weight_delta(prior: List[str], posterior: List[str]) -> float:
    return weight(posterior) - weight(prior)

print(weight_delta(posterior_2, posterior_1000))
```

```
0.3770376394847217
```

Another would be to take the proportionate difference between them, except now it's a bit harder to understand the scale.

```
21.653860839031083
```