

Tabular Approach to Bayesian Inference

A *hypothesis* is a proposed explanation for a phenomenon.

Sometimes there are several competing hypotheses that can explain the very same phenomenon. The purpose of *Statistical Inference*, and more specifically *Bayesian Inference* is to form rational beliefs about the relative likelihood between several competing hypotheses given the relevant information (or data) on a phenomenon.

One way to interpret *probability* is the subjective degree of belief in each hypothesis, which sums to 1. The following concepts are central to Bayesian Inference:

Prior: $P(H_i)$: The probability (as a pre-existing belief) of a hypothesis before seeing data

Posterior: $P(H_i|D)$: The updated probability of a hypothesis after seeing the data

Likelihood: $P(D|H_i)$: The probability of observing the data assuming a given hypothesis is true

If you have a finite number of hypotheses (let's say N), then Bayes Rule can be written as:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_{i=1}^N P(D|H_i)P(H_i)}$$

In other words, the *posterior is proportional to the prior times the likelihood*.

Bayes rule is a method to update a set of probabilistic beliefs about hypotheses after new data is taken into account. If you start with a prior and gather data, Bayes rule can be used to calculate the posterior probability for each hypothesis.

This calculation can be organized using a table like the one on the following page, which can be used for practice and also as a template to enter into a spreadsheet.

Step 1: Write each hypothesis down in column A as labels

Step 2: Come up with a prior probability for each hypothesis, and record these in column B.

They can be relative (in other words, they don't need to add to 1 or be normalized)

Step 3: In Column C, calculate the likelihood of the data under each hypothesis. In each case, assume that the hypothesis for that row is true, and calculate the probability of seeing the data collected.

Step 4: Multiply the values in columns B and C to fill in Column D

Step 5: Calculate the Sum of Column D

Step 6: Divide each item in column D by the Sum to get the posterior probability for each hypothesis in column E.

