

20.2: Bayes' Theorem and Inverse Inference

The reason that Bayesian statistics has its name is because it takes advantage of Bayes' theorem to make inferences from data about the underlying process that generated the data. Let's say that we want to know whether a coin is fair. To test this, we flip the coin 10 times and come up with 7 heads. Before this test we were pretty sure that the $P_{heads} = 0.5$, but finding 7 heads out of 10 flips would certainly give us pause if we believed that $P_{heads} = 0.5$. We already know how to compute the conditional probability that we would flip 7 or more heads out of 10 if the coin is really fair ($P(n \geq 7 | p_{heads} = 0.5)$), using the binomial distribution.

TBD: MOTIVATE SWITCH FROM 7 To 7 OR MORE

The resulting probability is 0.055. That is a fairly small number, but this number doesn't really answer the question that we are asking – it is telling us about the likelihood of 7 or more heads given some particular probability of heads, whereas what we really want to know is the probability of heads. This should sound familiar, as it's exactly the situation that we were in with null hypothesis testing, which told us about the likelihood of data rather than the likelihood of hypotheses.

Remember that Bayes' theorem provides us with the tool that we need to invert a conditional probability:

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)}$$

We can think of this theorem as having four parts:

- *prior* ($P(Hypothesis)$): Our degree of belief about hypothesis H before seeing the data D
- *likelihood* ($P(Data|Hypothesis)$): How likely are the observed data D under hypothesis H?
- *marginal likelihood* ($P(Data)$): How likely are the observed data, combining over all possible hypotheses?
- *posterior* ($P(Hypothesis|Data)$): Our updated belief about hypothesis H, given the data D

In the case of our coin-flipping example: - *prior* (P_{heads}): Our degree of belief the likelihood of flipping heads, which was $P_{heads} = 0.5$ - *likelihood* ($P(7 \text{ or more heads out of } 10 \text{ flips} | P_{heads} = 0.5)$): How likely are 7 or more heads out of 10 flips if $P_{heads} = 0.5$? - *marginal likelihood* ($P(7 \text{ or more heads out of } 10 \text{ flips})$): How likely are we to observe 7 heads out of 10 coin flips, in general? - *posterior* ($P_{heads} | 7 \text{ or more heads out of } 10 \text{ coin flips}$): Our updated belief about P_{heads} given the observed coin flips

Here we see one of the primary differences between frequentist and Bayesian statistics. Frequentists do not believe in the idea of a probability of a hypothesis (i.e., our degree of belief about a hypothesis) – for them, a hypothesis is either true or it isn't. Another way to say this is that for the frequentist, the hypothesis is fixed and the data are random, which is why frequentist inference focuses on describing the probability of data given a hypothesis (i.e. the p-value). Bayesians, on the other hand, are comfortable making probability statements about both data and hypotheses.

This page titled 20.2: Bayes' Theorem and Inverse Inference is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.