

10.11: Learning from Data

Another way to think of Bayes' rule is as a way to update our beliefs on the basis of data – that is, learning about the world using data. Let's look at Bayes' rule again:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$$

The different parts of Bayes' rule have specific names, that relate to their role in using Bayes rule to update our beliefs. We start out with an initial guess about the probability of B ($P(B)$), which we refer to as the *prior* probability. In the PSA example we used the base rate for the prior, since it was our best guess as to the individual's chance of cancer before we knew the test result. We then collect some data, which in our example was the test result. The degree to which the data A are consistent with outcome B is given by $P(A|B)$, which we refer to as the *likelihood*. You can think of this as how likely the data are, given the particular hypothesis being tested. In our example, the hypothesis being tested was whether the individual had cancer, and the likelihood was based on our knowledge about the sensitivity of the test (that is, the probability of cancer given a positive test outcome). The denominator ($P(A)$) is referred to as the *marginal likelihood*, because it expresses the overall likelihood of the data, averaged across all of the possible values of A (which in our example were the positive and negative test results). The outcome to the left ($P(B|A)$) is referred to as the *posterior* - because it's what comes out the back end of the computation.

There is a another way of writing Bayes rule that makes this a bit clearer:

$$P(B|A) = \frac{P(A|B)}{P(A)} * P(B)$$

The part on the left ($\frac{P(A|B)}{P(A)}$) tells us how much more or less likely the data A are given B, relative to the overall (marginal) likelihood of the data, while the part on the right side ($P(B)$) tells us how likely we thought B was before we knew anything about the data. This makes it clearer that the role of Bayes theorem is to update our prior knowledge based on the degree to which the data are more likely given B than they would be overall. If the hypothesis is more likely given the data than it would be in general, then we increase our belief in the hypothesis; if it's less likely given the data, then we decrease our belief.

This page titled [10.11: Learning from Data](#) 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.