Log Probabilities

A log probability $\log P(E)$ is simply the log function applied to a probability. For example if $P(E) = 0.00001$ then $\log P(E) = \log(0.00001) \approx -11.51$. Note that in this book, the default base is the natural base e. There are many reasons why log probabilities are an essential tool for digital probability: (a) computers can be rather limited when representing very small numbers and (b) logs have the wonderful ability to turn multiplication into addition, and computers are much faster at addition.

You may have noticed that the log in the above example produced a negative number. Recall that $\log b = c$, with the implied natural base e is the same as the statement $e^c = b$. It says that c is the exponent of e that produces b. If b is a number between 0 and 1, what power should you raise e to in order to produce b? If you raise e^0 it produces 1. To produce a number less than 1, you must raise e to a power less than 0. That is a long way of saying: if you take the log of a probability, the result will be a negative number.

$$0 \leq P(E) \leq 1 \qquad \text{Axiom 1 of probability}$$
$$-\infty \leq \log P(E) \leq 0 \qquad \text{Rule for log probabilities}$$

Products become Addition

The product of probabilities $P(E)$ and $P(F)$ becomes addition in logarithmic space:

$$\log(P(E) \cdot P(F)) = \log P(E) + \log P(F)$$

This is especially convenient because computers are much more efficient when adding than when multiplying. It can also make derivations easier to write. This is especially true when you need to multiply many probabilities together:

$$\log \prod_i P(E_i) = \sum_i \log P(E_i)$$

Representing Very Small Probabilities

Computers have the power to process many events and consider the probability of very unlikely situations. While computers are capable of doing all the computation, the floating point representation means that computers can not represent decimals to perfect precision. In fact, python is unable to represent any probability smaller than 2.225e-308. On the other hand the log of that same number is -307.652 is very easy for a computer to store.

Why would you care? Often in the digital world, computers are asked to reason about the probability of data, or a whole dataset. For example, perhaps your data is words and you want to reason about the probability that a given author would write these specific words. While this probability is very small (we are talking about an exact document) it might be larger than the probability that a different author would write a specific document with specific words. For these sort of small probabilities, if you use computers, you would need to use log probabilities.