Pre-reading 9, Part 3

These last few questions are about putting likelihoods and priors together to do inference. Remember, we are going to be modeling language as a process of inferring a grammar based on data and a prior probability distribution over possible language types - so this dice example might seem a bit weird and irrelevant, but the idea is it's underlyingly exactly the same kind of inference that's involved.

1. Now let's put all the bits together. Imagine you have a large bag containing 50 dice. 49 of those dice are 'fair' dice, equally likely to roll 1, 2, 3, 4, 5 or 6. However, the 50th dice is loaded, and *always* rolls a 6. You friend is going to dip their hand into the bag, pull out a dice, roll it a few times, and you have to use Bayes' Rule to decide if it is a normal dice or the loaded dice.

Let's build up the various component parts. First, what is the prior probability that your friend will pull a normal, 'fair' dice out of the bag? In the notation used in the reading, we would represent this probability as something like p(fair-dice).

49/50, because 49 of the 50 marbles are fair.

What is the prior probability of them pulling out the loaded dice [i.e. p(loaded-dice)]?

1/50, because only 1 dice in the bag of 50 is loaded.

If they pull out a fair dice, what is the probability of getting a 6 when they roll that dice? In the notation used in the reading, we could write this down as p(6 | fair-dice), i.e. the probability of rolling a 6 given that we are rolling a fair dice.

1/6 - as discussed in earlier questions, a fair dice has 6 possible things it can do (roll a 1, roll a 2, etc), the probability mass is divided up evenly across those 6 possible outcomes, so each has a probability of 1/6.

If they pull out a fair dice, what is the probability of getting two 6s in a row when they roll that dice, i.e. p(6,6 | fair-dice)?

1/36: we covered this already, so you can figure out by enumeration that there are 36 possible outcomes of rolling a dice twice, and only 1 of them features a double-6; or you can multiply the probabilities to find out the probability of two independent events both occurring, 1/6 x 1/6 = 1/36.

If they pull out the loaded dice, what is the probability of getting a 6 when they roll that dice, i.e. p(6 | loaded-dice)?

1 - the loaded dice *always* rolls a 6.

If they pull out the loaded dice, what is the likelihood of getting two 6s in a row when they roll that dice, i.e. p(6,6 | loaded-dice)?

Also 1 - it always rolls a 6, so no matter how many times you roll it, it always produces all 6s. Or if you want to multiply the probabilities, the probability of a double-6 is 1 x 1 = 1.

Now, you friend reaches into the bag, pulls out a dice, and rolls it - it's a 6! Then they roll it again - another 6! Using Bayes' Rule: is it more probable that they are rolling a fair dice or the loaded dice?

Hint: Bayes Rule states that the posterior probability of a hypothesis given some data is *proportional to* the prior probability of that hypothesis times the likelihood of the data given that hypothesis. To get actual posterior probabilities, you have to divide by the probability of the data, but all that does is normalise everything, so you don't need to do it to answer this question - you can figure out the answer by just considering the priors and likelihoods that you worked out above. The hypotheses are that the dice is fair or it is loaded; the data is the dice rolls.

It's more probable that they are rolling a fair dice.

Here are the prior probabilities of a loaded and fair dice, which we worked out above:

p(fair) = 49/50
p(loaded) = 1/50

Here are the likelihoods we just worked out, for a single roll:
p(6 | fair) = 1/6
p(6 | loaded) = 1

And for a roll of a double 6
p(6,6 | fair) = 1/36
p(6,6 | loaded) = 1

So now we can work out the posterior probability. As advised in the hint, I am going to ignore the denominator of Bayes' Rule (p(d), the probability of the data - in this case the probability of getting a 6 with a dice pulled out of our bag, regardless of which dice it is). So Bayes Rule says that the posterior probability of a hypothesis given some data is proportional to the prior probability of the hypothesis, times the likelihood of the data given that hypothesis. I am going to use ~ to mean "is proportional to", so we can write that as:

p(hypothesis | data) ~ p(data | hypothesis) * p(hypothesis)

So considering just the first roll, for the fair dice:

p(fair | 6) ~ p (6 | fair) * p(fair) = 1/6 * 49/50 = 49/300, which is approximately 1/6

And for the loaded dice:

p(loaded | 6) ~ p (6 | loaded) * p(loaded) = 1 * 1/50 = 1/50

So after a single roll of a 6, it's about 8 times more likely that the dice is a fair dice than the loaded dice (the ratio of the posterior probabilities is roughly 8 to 1, or p(fair | 6) is about 8 times larger than p(loaded | 6))- the loaded dice is way more likely to roll 6, but we know that it was very unlikely a priori that our friend happened to grab that dice out of all the dice they could have randomly selected, so even though they rolled a 6 we should still guess that they are probably rolling a fair dice.

But then they roll a second 6. So now we have more data, we can work through the numbers again:

p(fair I 6,6) ~ p (6,6 I fair) * p (6 I fair) = 1/36 * 49/50 = 49/1800, which is roughly 1/36

p(loaded I 6,6) ~ p (6,6 I loaded) * p(loaded) = 1 * 1/50 = 1/50

So after the second 6, we are much less confident that they are rolling a fair dice - that's still our best guess, but the posterior probabilities for the two hypotheses are now much closer. Now we are starting to get suspicious - something weird is going on, either our friend pulled a fair dice (very probable) then rolled a double 6 with it (quite improbable), or they pulled a loaded dice (very improbable) and rolled a double 6 with it (which always happens if they got the loaded dice).

Now your friend rolls the dice a third time. Another 6! Based on this new data, is it more probable that they are rolling a fair dice or the loaded dice?

It's more probable that they are rolling the loaded dice.

After the 3rd 6, the Bayesian individual changes their mind - the dice must be loaded after all. To see why, we need to figure out the likelihoods of rolling 3 6s given the two hypotheses about the dice:

p(6,6,6 I fair) = p(6 I fair) * p(6 I fair) * p(6 I fair) = 1/6 * 1/6 * 1/6 = 1/216

p(6,6,6 I loaded) = p(6 I loaded) * p(6 I loaded) * p(6 I loaded) = 1 * 1 * 1 = 1

Then we can drop them in to our calculation of the posterior:

p(fair I 6,6,6) = p (6,6,6 I fair) * p(fair) = 1/216 * 49/50 = 49/10800, which is roughly 1/216

p(loaded I 6,6,6) = p(6,6,6 I loaded) * p(loaded) = 1 * 1/50 = 1/50

So now it's over 4 times more likely that the dice is loaded than fair - it was really unlikely that they pulled the loaded dice out of the bag, but it's even less likely that they pulled a fair dice and then got it to roll 3 sixes (in fact, 4 and a bit times less likely), so the rational individual will conclude that the dice is probably loaded. Note how data and prior probability trade off in this model - even if something has low prior probability, we can be forced to accept that it is true given enough data; however, given little data, we tend to go with our prior expectations.