BAYESIAN PARAMETER ESTIMATION

Have you ever tried to explain to someone what a 95% confidence interval is? What do you usually say? Is it easy to explain? How about explaining what a *p*-value is? How many words do you have to use?

Here are the actual definitions:
95% confidence interval: A range of numbers obtained with a method that, if repeated over and over again, would contain the true value of a parameter 95% of the time.

p-value: The probability of unobserved data that are more "extreme" than the observed data, if the null-hypothesis were true. (in other words, it is the probability of data that didn't occur!).

Kinda a mouthful, isn't it? Wouldn't it be nice to say that a range of numbers has a 95% chance of containing the actual parameter value? Or to be able to state the probability of a given hypothesis? Frequentist statistics doesn't allow you to calculate these things.

To do this, we need to use Bayes Theorem:

$$\mathrm{P}(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)}$$

In words, this says the probability of *x*, given *y*, is equal to the probability of *y* given *x*, times the probability of *x*, divided by the probability of *y*.

How does this help us in parameter estimation? Recall that the likelihoods are equal to P(data|hypothesis). What we really want to know is the P(hypothesis|data). Thus, we can use Bayes theorem convert likelihoods into the probability that we're interested in:

$$\mathrm{P(Hypothesis \mid data)} = \frac{P(\mathrm{data \mid Hypothesis})P(\mathrm{Hypothesis})}{P(\mathrm{data})}$$

P(Hypothesis) is called the "Prior Probability" because it reflects prior beliefs about the probability of a given hypothesis. P(Hypothesis|data) is usually called the posterior probability, meaning that it reflects your belief about the hypothesis after collecting the data.

Advantages of Bayesian Estimation:
(1) The prior probability allows you to incorporate knowledge about a particular hypothesis. Suppose you are conducting a study on lemur foraging, and a similar study was conducted on lemur foraging 20 years ago. Should you pretend that the early study doesn't exist, or should you incorporate that study into your scientific decision-making?

A frequentist says that your conclusions should be independent of prior work, while a Bayesian says that science works by building on established knowledge.

(2) The name of the game is to compare the relative probability of competing hypotheses. The Bayesian framework allows you to make that comparison (although, as we will see, likelihoods also give you that information but with no prior knowledge).

Disadvantages of Bayesian Estimation

(1) Frequentist statisticians reject the notion of the prior-probability because it is "subjective". You can get very different posterior distributions by changing what parameters have uninformative priors. In other words, there are some tricky mechanical issues.

(2) The frequentist-based framework is ideal for the Popperian view of science because it allows you to falsify hypothesis. Under Bayesian statistics, there is no such thing as falsification, just relative degrees of belief. Many feel that the frequentist argument falls flat. First of all, the Popperian view of science is not the end-all of scientific philosophies (incidentally, it is no coincidence that R.A. Fisher, the founder of frequentist statistics, was heavily influenced by his contemporary, Karl Popper). Many have argued that the Popperian view of science is overly narrow and is ill-suited for complex, multi-causal systems. Even more importantly, the frequentist approach still has the same problem of translating probabilities into "accept" vs. "reject" rules, but tries to avoid dealing with the problem by relying on arbitrarily-selected threshold for falsification.

(3) Frequentist statistics is "easy" and has accepted conventions of method and notation. The same can’t be said of Bayesian statistics. Some have argued that the primary reason for the idea that frequentist statistics is “easy” is that most people have been trained in it during introductory classes, and today’s software packages can do most of the work for you. In other words, you can do frequentist statistics without knowing what you're doing, but to do Bayesian statistics you need to understand probability and likelihood.

<u>The Bayesian Squirrel</u>

Hilborn and Mangel describe the Bayesian squirrel. This squirrel has stashed away her fall loot of acorns in one of two patches. The only problem is, she can't remember which one it is. There are two hypotheses: Hypothesis 1 is that the food is in patch 1 and Hypothesis 2 is that the food is in patch 2. This squirrel is pretty sure that she left the food in patch 1. In fact, she's willing to say that there's an 80% chance that the food is in patch 1. She also knows that she's really good at hiding her food. Consequently, there's only a 20% chance of finding the food per day when she's looking in the right patch (and, of course, a 0% probability if she's looking in the wrong patch).

Before she even starts searching, she has a prior probability P(patch 1)=0.8. Fortunately, this squirrel has been trained in Bayes theorem, and can therefore calculate posterior probabilities. Suppose she looks in patch 1 and doesn't find any food? What's the probability that the food is in patch 1, given that she didn't find anything? In terms of Bayes Theorem, we see that:

$$P(\text{Food is in patch 1} \mid \text{Find no food}) = \frac{P(\text{Find no food} \mid \text{Food is in patch 1})P(\text{Food is in patch 1})}{P(\text{Find no food})}$$

The probability of not finding food when searching in the right patch is 0.8., so:

$$P(\text{Find no food} \mid \text{Food is in patch 1}) = 1 - 0.2 = 0.8$$

The second term is our prior probability that the food is in patch 1, which equals 0.8:

$$P(\text{Food is in patch 1}) = 0.8$$

What's the denominator? There are two ways that she could not find food. She could have been looking in the right patch but didn't come across the food, or she's looking in the wrong patch (e.g., the food is in patch 2). Whenever we see an "or" statement, we know that we should be adding probabilities, so, the probability of her not finding her food is P(food is in patch 2)+0.8 P(food is in patch 1). Again, P(food is in patch 1) is our prior probability, and P(food is in patch 2) is simply 1-the prior probability:

$$P(\text{Find no food}) = P(\text{Food is in patch 2}) + \text{P(Food is in Patch 1)P(Find no Food} \mid \text{Food is in Patch 1)}$$
=0.2 + (0.8)(0.8)

Open the excel spreadsheet called Bayesian squirrel. This spreadsheet will calculate the posterior probability that the food is in patch 1, given that she fails to find food there. She knows that she should search in patch 1 until the probability that the food is in patch 1 drops below 0.5. At that point, she should switch and look in patch 2.

If this squirrel looks once in patch 1 and does not find her food, what is the probability that the food is in patch 1? Should she continue to look in patch 1?

Copy the posterior probability in the table below.

Suppose the looks again in patch 1? Before she starts looking, she should use her old posterior probability as her new prior probability, thereby updating her belief about where the food is. Click on the squirrel to update the prior, and then see the new calculated posterior probability if she **again** fails to find food. Fill it out on the table below.

Keep repeating this procedure until the probability that the food is in Patch 1 drops below 50% How many times does she have to fail to find food in patch 1 before she begins to think that the food is patch 2?

Time Step	P(Food is in patch 1)

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Now, suppose that she was much better at locating food within a patch, so that the she locates food 30% of the time that she searches in the right place. Repeat the procedure you did before (*remember to reset the prior probability to 0.8*). Now how many time steps does she have to fail to find food in patch 1 before she begins to think that the food is in patch 2? Explain why the results were different when the squirrel is better at locating her acorns.

From Squirrels to Unicorns:

Recall our unicorn example (this is the last time, I promise). A survey spotted 4 unicorns in 20 km^2, and we saw that we could use likelihood to gauge the probability of seeing 4 unicorns given different unicorn densities. We then calculated the probability of the data, given different values of unicorn density (r; $\#/km^2$). In other words we calculated:

P(saw 4 unicorns | *r*)

But, what we really want to know is what is the probability of a given unicorn density given that we saw 4 unicorns. In probability notation, this is:

P(*r* | saw 4 unicorns)

Of course, we can use Bayes theorem to calculate this:

$$P(r \mid \text{saw 4 unicorns}) = \frac{P(r)P(\text{saw 4 unicorns} \mid r)}{P(\text{saw 4 unicorns})}$$

P(*r*) is the prior probability of *r*, P(saw 4 unicorns | *r*) is the likelihood, and the bottom is basically a scaling factor that ensures that all the probabilities sum to 1.

The tricky thing here is coming up with a prior probability for *r*. What if you have no idea how many unicorns are out there. What basis do you have for assigning a prior probability for any particular value of *r*? Not much. In this case, you would use a *non-*

informative prior, for which P(*r*) is a constant value for a range of values of *r*. In other words, P(*r*) is going to be a uniform distribution.

Open up the spreadsheet called Bayesian unicorn, and go the page called "uniform prior". On this spreadsheet you'll see the likelihood, the prior probability and the posterior probability. Compare the posterior probability and the likelihood graph. Do they look similar? Based on your understanding Bayes theorem, why does this happen when you have a uniform prior?

Now you see that likelihoods can be interpreted as posterior probabilities if you have an uninformative prior about the parameter being estimated. All you need to do is scale them so that they sum to 1. Easy, eh?

Now we can make some statements about the probability of different parameter values. What is the probability that *r* is less than 0.1? You can calculate this by summing all of the probabilities for r greater than 0.1.

What is the probability that *r* is between 0.15 and 0.25?

Find a range of *r* that has a 95% probability of containing the true value. This is often called a 95% credibility interval, to distinguish it from the confusing confidence interval of frequentist statistics. Note the different interpretation of a 95% credibility interval vs. a 95% confidence interval.

Suppose that a few days before your survey, somebody surveyed 40 km^2 and spotted 12 unicorns. You want to incorporate this information into your estimate. How would you do this?

Switch to the worksheet called "informative prior". Here you'll again see the likelihood from your data (which should look the same as before), the prior probability based on the previous study, and the resulting posterior probability distribution. What is the most probable value for *r*? Now what is the probability that *r* is less than 0.1? That *r* is between 0.15 and 0.25? What is the 95% credibility interval?

One Final Unicorn Example

Recall that the above calculations were assuming homogeneous unicorn distributions, and we saw that patchiness in unicorns in space and/or time produces a lot of uncertainty in our estimate. Now we'll examine how a Bayesian would approach this problem.

Remember that we are using a negative binomial distribution, which is based on the Poisson model but it has an addition parameter, n, that dictates how dispersed unicorns are. When n is high (e.g., >100), the negative binomial is approximately equal to the Poisson distribution. In this exercise we are going to consider the case where we don't have prior information on unicorn densities, but we do have prior information on the dispersion parameter.

Open up the Excel file called "Patchy Bayes". On the top left hand corner you can specify your data. For now, lets just pretend that we saw 8 unicorns in 40 km^2 of searching. We also have two cells that control our prior probability of unicorn dispersion parameter, n. We are going to assume in this model that the ln(n) is normally distributed, with some mean, μ, and some standard deviation, σ. For now, make sure that the mean is set to 2.3 and the standard deviation is set to 150. On the graphs below, you can see the prior probability of n, and the posterior probability of unicorn density.

How did I calculate this? It was similar to the calculation of likelihood profiles that we used last week. Recall that the likelihood profile looks at many combinations of the parameters *r* and *n*, and for each *r* assigns the largest possible likelihood that was found over the many values of *n* considered.

To calculate a Bayesian posterior, I instead want to consider the sum of the likelihoods over all of the different values of n. To show this graphically, move over to the spreadsheet called "2-d plot", where you'll see a surface plot of unicorn density on the x-axis, dispersion parameter on the y, and the probability of the data under each combination of *r* and *n* in the "z" direction. To calculate the total amount of probability associated with each value of r, I simply summed up all of the probabilities for that value of *r*. That creates the posterior probability plot shown below the 2-d plot.

Move back over to the page called "Priors and Posteriors". Based on the graph of the prior probability of n, would you say that this is an "informative' or "non-informative" prior? Explain.

What is the most probable value of unicorn density, and what is the 95% credibility interval?

Suppose the U.S. Dept. of Imaginary Animals (DIA) was considering listing unicorns as a threatened species, but being rather risk-prone, they will only list them if there is a 90% chance that unicorn density is below 0.5 unicorns /km^2. Do your data provide sufficient evidence to list unicorns on the DIA threatened species list?

Iteratively adjust the unicorn density in cell "B8" until you find a value for which there is a 90% chance that unicorns are below this density.

After conducting you study, an additional study on unicorn foraging and dispersal patterns was conducted. They used all types of fancy methods to describe the patchiness in unicorn spatial distributions, and generated a probability distribution for the ln(dispersion parameter). This distribution was normally distributed, with a mean of 2.3 and standard deviation of 0.5.

Change the value of the prior probability standard deviation, and see how the prior probability distribution changed.

Tab over to the page "2-d Graph" to look at how the two-dimensional posterior probability plot changed, and how this, in turn, changed the posterior probability of unicorn density. Pretty cool, huh?

Now, move back over to the "Prior and Posteriors" page, and identify the most probable value of *r* and the 95% credibility interval. Did the either of these change much?

Lastly, reconsider whether there is sufficient evidence to list unicorns on the DIA threatened species list.

Feedback

You (hopefully) found that incorporating prior information led to different conclusions. This is both a strength and a weakness of Bayesian statistics, depending on who you talk to. Which do you think? Provide some justification for your stance.