Homework 7

Biology 697

10/19/2012

1   Single Parameter Likelihood

To assess the health of multiple desert bird populations, scientists sampled the abundance of a number of species. You've been asked to model the distribution. Given the shape of the data, and that these results are largely driven by similarly sized adult population minus a small amount of mortality, you select the geometric distribution (see Bolker and investigate `dgeom`). This distribution has a single parameter, p, typically the probability of failure - i.e., the probability of not observing any more birds of a given species while sampling.

1.1

Using Likelihood, find this probability. Also, plot the likelihood curve and determine the 95% CI.

1.2

Based on previous sampling events, scientists expected a value of the parameter to be 0.15. Can we reject the hypothesis that the parameter is 0.15?

1.3

One niggling question keeps coming up. Maybe a Poisson distribution would have been better, modelling lambda as the mean number of birds. Compare the likelihood of a model fitted with a Poisson distribution to that of the Geometric. How different are they? Which would provide a better description of the data?

2   Multiparameter Likelihood - Brute Force Style

Another data set provided by W&S regards the running speed of male spiders before and after they amputate one of their spermatophores after sex. The idea is that lighter male spiders can run away faster from their hungry mate.

2.1

Using brute force (iteration) use likelihood to estimate the mean and SD of the male spiders' running speed pre-amputation. Plot the Log-Likelihood surface using the function of your choice - countour, persp, or anything you find interesting. Note, for contour and persp, the x and y arguments are vectors of values in order along an axis, and z is a matrix where, at every cell [i,k], you have a function calculated using x[i] and y[j]. You could also use ggplot2 with a color argument to lay out a surface. It might help to plot the distribution to guess at what values you should iterate over.

2.2

Now, do the same fit with mle2, just to check yourself.

2.3

Now for the fun. Using brute force, calculate the profile Log-Likelihood for each parameter, and use them to determine the 95% CI. As we only touched on this briefly in class, here's a reminder of how to get the 95% CI for any paramater of interest.

1. For the parameter of interest, iterate over a many values and calculte the log-likelihood at each iteration for a model where all other parameters have been maximized.

2. This curve is your profile, and the log-likelihood at each value of your parameter is the log-likelihood you'll use when determining the 95% CI

3. Now determine all values of the parameter that fall within 1.92 log-likelihood units of the MLE. This is your CI.

For those of you who have created a matrix above, well, you've already done all of the fitting you'll need to determine the profile. You can write some code that will generate a profile using that matrix. Or, if you haven't fit a matrix (or don't want to use the one you've fit), you can brute-force create a profile. Once you have the profile in hand, determine the bounds of the 95% CI. 2 Extra points for adding those profile lines to your previous plot of the likelihood surface. Notice anything interesting?

2.4

Check yourself by plotting the profiles from the mle2 fit and examining the confidence intervals.

3 Linear Regression and Likelihood

So the question still remains, does it work? Do spiders get faster after they amputate? One way to investigate this is to see if there is a relationship between running speed before and after. If there is, and the slope is greater than 1, then amputation may be beneficial.

3.1 Regression with Functions

Write a function that will give you the negative log-likelihood of a linear regression fit to the data. Then use `mle2` with that function to get the maximum likelihood estimates of those parameters. Note, you may need to play with start values a bit to get a fit.

3.2 Regression with Just mle2

Huh. Why did you have all of those NA issues? Sometimes, using a function, you can search into areas of parameter space which give bad results - i.e., if the SD is less than 0. You can write a function to avoid these issues. But that can get hairy. This is one reason that mle2 implements a number of standard distributions - to help you avoid getting on the wrong track. So, let's see what MLE2 will cool up if you let it do its work. Fit the same linear relationship just using `mle2` and no external function. How do the estimates compare to linear regression?

3.3 Confidence

What is the 95% CI? How do the profile and Wald estimates compare? Plot the profiles to help you answer.

3.4 Is there a slope?

Test two alternate hypotheses against your MLE fit. First, that there is actually no relatioship between before and after speed. Second, that there is a relationship, and that it is 1.2 (so, greater than 1). What do your results from these tests tell you about spider amputation?