Lab 6: Likelihood

36-350, Statistical Computing

Friday, 7 October 2011

Agenda: Using functions as arguments and return values; plotting functions; maximum likelihood estimation.

Instructions: Save all your answers in a single plain text file (Word files will not be graded), and upload it to Blackboard, using the page for this assignment. (There is no general digital dropbox any more.) When a question asks you to do something, give the command you use to do it. For questions which ask you to explain, write a short explanation in coherent, complete sentences. (You will be graded on your written explanation, not what you might say to the TA.)

When we have independent samples $x_1, x_2, \ldots x_n$ from a common probability density $p(x)$, the joint probability density of the whole sample is

$$\prod_{i=1}^{n} p(x_i)$$

When we are not sure what the right density p is, but we know it belongs to some family (like the Gaussian, the exponential, the gamma, etc.), we write the parameters of the family as θ, and say that the **likelihood function** is

$$L(\theta) = \prod_{i=1}^{n} p(x_i; \theta)$$

Notice that the likelihood is a function of the unknown parameters θ, not the known data $x_{1:n}$. One way to estimate the parameters is to maximize the likelihood,

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$$

For several reasons, including numerical stability, we usually work with the log-likelihood instead,

$$\ell(\theta) \equiv \log L(\theta) = \sum_{i=1}^{n} \log p(x_i; \theta)$$

whose maximum is located at the same point as the maximum of L. (Why?) Maximum likelihood estimation is generally the most efficient way to find the parameters of a probability density, when true density really is in the family we've guessed.

In this lab, we begin working with likelihood functions, continuing to use the data on the heart weight of cats from previous labs. (Load it now, please, as in Lab 3.)

1. Fit the gamma distribution to the cats' hearts', using the method from Lab 3. You can use code from that lab's solutions. (5)

2. Calculate the log-likelihood of the shape and scale parameters you just estimated. The answer, rounded to the nearest integer, should be -326. (15)

3. Write a function, `make.gamma.loglike`, which takes in a data vector `x` and returns the gamma log-likelihood function for that data. The returned function should take one argument, a vector of length 2, and treat its first component as the shape parameter and the second component as scale. Check that `make.gamma.loglike`, when run on `cats$Hwt`, returns a function which matches your calculation in the previous question. (25)

4. Make a contour plot of the log-likelihood, with the shape parameter on the horizontal axis (range 1 to 40) and the scale parameter on the vertical (range 0.01 to 1). Add a point indicating the location of your moment-based estimate from question 1. *Hints*: Consider `surface.0` from Lecture 11. Also, you will probably want to increase the number of levels on the contour plot above the default of 10. (15)

5. Use the plot from the previous question to locate the region where the likelihood seems to be maximized. Make a new plot which zooms in on this region. (10)

6. Using the version of `gradient.descent` given for Homework 6, find the parameters which maximize the log-likelihood, with your estimate from question 1 as the initial guess. *Hint*: Think carefully about what `gradient.descent` does when setting the `step.scale` argument. (20)

7. Is the maximum likelihood estimate close to where you thought the maximum would be from the plots? Is it close to the moment-based estimate from question 1? Should it be close to either? (10)