Lab 6: I Can Has Likelihood Surface?

36-350, Statistical Computing

Friday, 5 October 2012

Agenda: Using functions as arguments; plotting functions; likelihood.

Instructions: Save all your answers in a single plain text file (Word files will not be graded), and upload it to Lore. When asked to do something, give the command you used. When asked to explain, write a short explanation in coherent, complete sentences. When asked to make a figure, give the command you used, and upload the figure as a separate file. (We will see all the files you upload.)

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

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

3. (25) Write a function, `gamma.loglike`, which takes in a vector, containing a shape and a scale parameter, and returns the log-likelihood of the cats' hearts' masses. Check that when you run `gamma.loglike` with the estimate from question 1, you get the log-likelihood from question 2.

4. (25) 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 Monday's lecture. Also, you will probably want to increase the number of levels on the contour plot above the default of 10.

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

6. (15) Based on your plots, how much higher can the likelihood get than -326? Where, *roughly*, does it reach this maximum?