Lab 3: Of Big- and Small- Hearted Cats

36-350, Statistical Computing

Friday, 16 September 2011

Agenda: Writing functions to automate repetitive tasks; fitting statistical models.

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.)

The **gamma** distributions are a family of probability distributions defined by the density functions,

$$f(x) = \frac{x^{a-1}e^{-x/s}}{s^a \Gamma(a)} \tag{1}$$

where the **gamma function** $\Gamma(a) = \int_0^\infty u^{a-1}e^{-u}du$ is chosen so that the total probability of all non-negative x is 1. The parameter a is called the **shape**, and s is the **scale**. When $a = 1$, this becomes the exponential distributions we saw in the first lab. The gamma probability density function is called `dgamma()` in R. You can prove (as a calculus exercise) that the expectation value of this distribution is as, and the variance as^2. In this lab, you will fit a gamma distribution to data, and estimate the uncertainty in the fit.

Our data today are measurements of the weight of the hearts of 144 cats.

1. The data is contained in a data frame called `cats`, in an R **library** (or **package**) called `MASS`. (This package is part of the standard R installation.) This records the sex of each cat, its weight in kilograms, and the weight of its heart in grams. Load the data as follows:

   ```
   library(MASS)
   data(cats)
   ```

   Run `summary(cats)` and explain the results. (5)

2. Calculate the mean, standard deviation, and variance of the heart weights. (5)

3. Write a formula for the scale parameter s of a gamma distribution in terms of its mean and variance. *Hint*: Use algebra, not R. (8)

4. Write a formula for the shape parameter a of a gamma distribution in terms of its scale s and its mean. *Hint*: More algebra. (7)

5. Plug the mean and variance of the cats' hearts into your formulas and get estimates of a and s. What are they? Do not report them to more significant digits than is reasonable. (5)

6. Write a function, `gamma.est()`, which takes as input a vector of numbers, calculates the mean and variance, and returns the estimate of a and s. (30)

7. What estimates does `gamma.est` give on the cats' hearts weight? Should it agree with your answer in question 5 or not? (5)

8. Plot a histogram of these weights. Using `curve` and `dgamma`, add the gamma pdf with the estimated shape and scale parameters. Be sure to turn in the exact commands you used. Turning in a PDF is optional. (15)

9. Estimate the a and s separately for all the male cats and all the female cats, using `gamma.est`. Give the commands you used and the results. (10)

10. Do the estimates for the two sexes agree with each other? How can you tell? (10)