Lab 3: Of Big- and Small- Hearted Cats

36-350, Statistical Computing

Friday, 14 September 2012

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

Instructions: Save all your answers in a single plain text file (Word or PDF will not be graded), and upload it to Lore. 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 complete, coherent 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. If the mean and variance are known, μ and σ^2, then we can solve for the parameters,

$$a = \frac{a^2s^2}{as^2} = \frac{\mu^2}{\sigma^2} \tag{2}$$

$$s = \frac{as^2}{as} = \frac{\sigma^2}{\mu} \tag{3}$$

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. (5) 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.

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

3. (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.

4. (30) 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.

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

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

7. (10) 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.

8. (5) Plot the two estimated density functions for the two sexes on the *same* plot, using `curve` and `dgamma`.

9. (15) Randomly divide the cats into two groups, of the same size as the male and female cats. Estimate the gamma parameters separately for your random groups. Do this multiple times. Report the typical range of differences in parameters between the two artificial groups (as well as the commands you used). *Hint:* Recipe 8.5, and/or 8.7.

10. (5) What can you conclude about whether the distribution of heart weights for female cats is different from the distribution for male cats? Explain.