

Lab 4: Heart of the (Tiny) Tiger

2 October 2015

Agenda: Writing functions to automate repetitive tasks; fitting statistical models by method of moments

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

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

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,

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

In this lab, you will fit a gamma distribution to data using the method of moments.

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

Part I

1. The data is contained in a data frame called `cats`, in the R package `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. Plot a histogram of these weights using the `probability=TRUE` option. Add a vertical line with your calculated mean using `abline(v=yourmeanvaluehere)`.
3. Write a function that takes mean and variance as inputs and returns a and s , assuming that the mean and variance came from a random sample from a gamma distribution.
4. Calculate the mean, standard deviation, and variance of the heart weights using R's existing functions for these tasks. Use your function to get estimates of a and s . What are they? Do not report them to more significant digits than is reasonable.
5. Write a function, `cat.stats()`, which takes as input a vector of numbers and returns four estimates: mean, variance, a and s .

Part II

6. Estimate the a and s separately for all the male cats and all the female cats.
7. Now, produce a histogram for the female cats. On top of this, add the shape of the gamma PDF using `curve()` with its first argument as `dgamma()`, the known PDF for the Gamma distribution. Is this distribution consistent with the empirical probability density of the histogram?
8. Repeat the previous step for male cats. How do the distributions compare?