# Lab 10: Maximum Likelihood!

Like we saw in the last lab, when we have independent samples $x_1, x_2, ...x_n$ 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 $\theta$, 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 $\theta$, not the known data $x_{1:n}$. One way to estimate the parameters is to maximize the likelihood,

$$\widehat{\theta}_{MLE} = \mathrm{argmax}_{\theta} L(\theta).$$

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

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

whose maximum is located at the same point as the maximum of $L$. By convention, optimization functions in software packages often find minimum values, so we often find maximum likelihood estimators by finding parameters to minimize the negative log-likelihood.

Maximum likelihood estimation is generally the most stastistically 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.

1. Review the previous lectures and the last lab. Fit a gamma distribtion to the cats' heart weights by maximum likelihood. (This will involve writing a function for the negative log-likelihood, its gradient, and using **optim()**. Starting values are for you to decide.)
2. Verify that the gradient of the negative log-likelihood at the maximum likelihood estiate is close to zero.
3. The negative inverse Hessian matrix of the log likelihood is an estimate of the covariance matrix of the estimated parameters. Find this for the cats' heart weights. (You may do this in one of several ways: do the calculus by hand and plug in estimates, use the results of **optim()**, or use the **grad()** function. If you are ambitious, do all three!)
4. Use the results of 3 to make large sample confidence intervals for the parameters in the gamma distribution.