WILD 595 - 1: Lab 5 Maximum Likelihood.

OBJECTIVES

- Learn the basics of maximum likelihood estimation
- Become familiar with the optim() function
- Understand diagnostic tools available for MLE

Maximum Likelihood Estimation

Maximum likelihood estimation is much as it sounds… parameter estimates are obtained by maximizing the log-likelihood function of the data given a model. In this lab, we will explore the optim() function as a method for obtaining MLEs and variances. Let's begin with a data set of random normal values

```
ndata <- rnorm( 100, 10, 3 ) # what does this represent?
```

The normal distribution has two parameters, the mean (μ) and standard deviation (σ). Let's use MLE to estimate μ and sigma from the data given above. First we must define the likelihood function.

```
n.likelihood <- function( p, data ){
     ll <- sum( dnorm( data, p[1], p[2], log=TRUE ) )
     return(ll)
}  # what does this function do?
```

Now we can optimize the log-likelihood. We will use the optim function. We need:

- an initial value vector (of what length?)
- the function name
- a scaling option to maximize rather than minimize
- turn the Hessian matrix on
- pass the data to the likelihood through optim()

Assignment due 10/5/15

1) Generate a data set of 100 values each of which is the average of two random numbers from a uniform(0,1) distribution. Assuming a normal distribution for the likelihood, estimate the mean and standard deviation of the data. Estimate the standard error of each parameter using the Hessian matrix. Now, use the mean() and sd() functions and compare the results. Plot a histogram of the data, do they look "normal"?

2) Generate a data set consisting of 250 values from a Normal(10, 3) and 750 values from a Normal(20,2). You now have a data set that contains 1,000 values. Plot a histogram of the data, what are its characteristics? Use a normal likelihood to estimate the mean and standard deviation. Use the mean() and sd() functions and compare the results. How do these estimates compare to your plot?

3) Use the data generated in 2). Now assume a normal-mixture model where:

$$\log L(\mathbf{p} \mid x) = \pi N(\mu_1, \sigma_1) + (1-\pi) N(\mu_2, \sigma_2)$$

Use this likelihood to estimate the five parameter (π, μ_1, σ_1, μ_2, σ_2). Use at least two different starting values, one <10 and one >20. Provide the parameter estimates and standard errors for each optimization. What is causing the results you see.