Lab 3 - Maximum Likelihood

Thomas Campbell
Victoria Peechatt

February 23, 2024

1 Problem

Coates and Burton (1999) studied the influence of light availability on the growth increment of saplings of species of conifers in northwestern interior cedar-hemlock forests of British Colombia. They used the deterministic model:

$$\mu_i = \frac{\alpha(L_i - C)}{\frac{\alpha}{\gamma} + (L_i - c)}$$

1) **Write a model:**
Our model for tree growth: $\mu_i \sim Normal(f(\alpha, \gamma), \sigma^2)$

2 Setting Up The Spreadsheet

2) How could you use the data to help you find good initial conditions for model parameters?
Using the data, we can adjust the values of the parameters which would change the predicted values with our function, and produce a model minimizing the sum of the least squares

3) Adjust the values of α, γ, and c until you get predictions that look reasonable in your plot. How could we get a better initial value for σ?
We can get a better value of sigma by maximizing the log likelihoods, which depends on the likelihood, which is calculated by the equation of the normal distribution, $f(x, \mu, \sigma)$ where we have x and μ.

4) Write the mathematics (the full equation) that is implemented in the formula in column E.
Normal Distribution PDF:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

PDF of Our Data:

$$f(\mu_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - f(\alpha,\gamma))^2}{2\sigma^2}}$$

5) What is the reason for the argument "FALSE" in the Excel formula in column E? Using the "FALSE" argument in Excel returns the probability density function of the distribution rathe than the probability cumulative distribution.

6) What does the function return when the argument is "TRUE"? Using the "TRUE" argument in Excel returns the cumulative distribution function of the distribution.

7) In column F we take the logs of the likelihoods which are summed in cell K2. If we had not taken the logs and instead, worked directly with the likelihoods, what formula would we use in K2?

$$L(\theta|\mu_i) = c\prod_{i=1}^{n}[\mu_i|f(\theta, x_i), \sigma^2]$$

where θ is some (α, γ) and x_i contains our data.

8) What are some potential computational problems with using the individual likelihoods rather than the log-likelihoods to estimate the total likelihood? There are two primary computational problems with using individual likelihoods instead of log-likelihoods. First, when we use log likelihoods, we are adding them together instead of multiplying them. Adding them is computationally more simple and more efficient. The second is that depending on parameters, multiplying extremely small/large numbers will produce unusable likelihoods. Using logs (addition) will produce more reasonable and usable likelihoods. If the likelihood in one instance was zero, it would make the total likelihood zero when using individual likelihoods rather than log likelihoods.

9) This model violates a fundamental assumption of traditional regression analysis. What is that assumption? How might you fix the problem? (Hint: think about what we are assuming about the covariate, light availability). We are violating the independence assumption between our variables. That is, our model using light availability as a covariate creates a dependency on the total light a sapling receives thus producing a violation of a key assumption.

3 Using the Excel Solver

10) How might you use the squared error column D to compute α? Make this computation and compare with your maximum likelihood estimate of α obtained using Solver. The formula for standard deviation is: $\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{n}}$. If we calculate this with out data, we get $\sigma = 7.356$ which is very close to the cell in our sheet where $\sigma = 7.370$.

4 Using R To Do The Same Thing

Below is our graph from the sample data given to us.

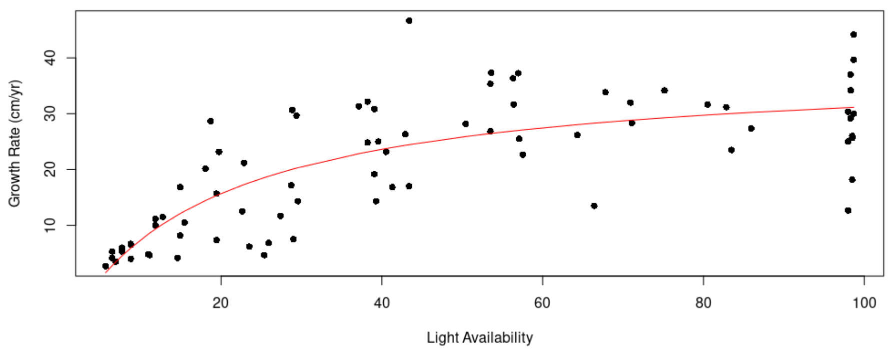

Figure 1: Using R for Finding MLE

5 Incorporating Prior Information in an MLE

Suppose that a previous study reported a mean value of $\alpha = 35$ with a standard deviation of the mean = 4.25. You may use a normal distribution to represent the prior information on α.

11) Write a model for the data that includes the prior information. Incorporate these prior data in your new MLE estimate of α. Hint: Create a likelihood function for the probability of the new value of α conditional on the previous value and its standard deviation.
$\mu_i \sim Normal(f(\alpha, \gamma), \sigma^2)$
$\alpha \sim Normal(\mu, \sigma^2)$

```r

## Incorporaating Priors into MLE

## keeping other parameters as best one from model
c = c.hat
s = s.hat
sigma2 = summary(model)$sigma

## initializing alpha values and total likelihood
aseq = seq(1,100,by=0.001)
llh = numeric(length(aseq))
i = 1

## Looping through all alpha values and adding likelihoods
## to find parameters that maximize it

for(a in aseq){
  #calculate expected value with function
  expect = a*(x-c)/(a/s+x-c)
  #likelihood of data
  likeli = prod(dnorm(y,expect,sd = sigma2,log = FALSE))
  #log likelihood of data
  likeli = log(likeli)
  #likelihood of alpha
  alikeli = dnorm(a, mean = 35, sd = 4.25, log = FALSE)
  #log likelihood of alpha
  alikeli = log(alikeli)
  #total likelihood
  llh[i] = likeli + alikeli

  #index
  i = i + 1

}

#Maximum likelihood
a_best = aseq[which.max(llh)]
a_best

[1] 37.849

```

12) How do you combine likelihood (or log-likelihoods) to obtain a total likelihood)? Using just likelihoods you multiply the likelihoods together to get to the total likelihood. Using log-likelihoods you sum the log-likelihoods to get to the total likelihood.

13) Describe what happens to the estimate of α relative to the one you obtained earlier. What is going on? The estimate for α based on both the prior and current data. With the prior data of α with $\mu = 35$ and $\sigma^2 = 4.25$ the estimate for α shifted from 38.5009 to 37.849, closer to the mean of the prior.

14) What is the effect of increasing the prior standard deviation on the new estimate? What happens when it shrinks? *Increasing* the prior's standard deviation means that there is a less accurate/more uncertain and a more flat distribution. A larger standard deviation in the prior means that the prior is less informative. *Decreasing* the prior's standard deviation means that the prior distribution has a more pronounced peak and its effect on the likelihood increases.

15) There is a single log-likelihood for the prior distribution but the sum of many for the data. This seems "unfair". Explain how the prior distribution can overwhelm the data and vice versa. If there is a strong prior it can overwhelm the new data especially if there is limited data, because the prior will have a greater influence on the estimation of α than the data itself. If there is a lot of data, however, it can easily obliterate the effect of a single prior. This means that many, good priors and a lot of data are required to get the most accurate estimations of any parameters, in this case α.