Brief introduction to maximum likelihood estimation

Some of this material is taken from Appendix B of my *Analysis of Categorical Data with R* textbook.

Suppose the success or failure of a field goal in football can be modeled with a Bernoulli(π) distribution. Let Y = 0 if the field goal is a failure and Y = 1 if the field goal is a success. Then the probability distribution for Y is:

P(Y = y) = 

where π denotes the probability of success.

Suppose we would like to estimate π for a 40 yard field goal. Let y1, …, yn denote a random sample of observed field goal results at 40 yards. Thus, these yi’s are either 0’s or 1’s. Given the resulting data (y1, …, yn), the “likelihood function” measures the plausibility of different values of π:





Suppose  = 4 and n = 10. Given this observed information, we would like to find the corresponding parameter value for π that produces the largest probability of obtaining this particular sample. The following table can be formed to help find this parameter value:

|  |  |
| --- | --- |
| π |  |
| 0.2 | 0.000419 |
| 0.3 | 0.000953 |
| 0.35 | 0.001132 |
| 0.39 | 0.001192 |
| 0.4 | 0.001194 |
| 0.41 | 0.001192 |
| 0.5 | 0.000977 |

Calculations in R (LikelihoodFunction.R):

> sum.y <- 4

> n <- 10

> pi <- c(0.2, 0.3, 0.35, 0.39, 0.4, 0.41, 0.5)

> Lik <- pi^sum.y\*(1-pi)^(n-sum.y)

> data.frame(pi, Lik)

 pi Lik

1 0.20 0.0004194304

2 0.30 0.0009529569

3 0.35 0.0011317547

4 0.39 0.0011918935

5 0.40 0.0011943936

6 0.41 0.0011919211

7 0.50 0.0009765625

> #Likelihood function plot

> curve(expr = x^sum.y\*(1-x)^(n-sum.y), from = 0, to =

 1, xlab = expression(pi), ylab = "Likelihood

 function")

> abline(h = 0)



Note that π = 0.4 is the “most plausible” value of π for the observed data because this maximizes the likelihood function. Therefore, 0.4 is the maximum likelihood estimate (MLE).

In general, the MLE can be found as follows:

1. Find the natural log of the likelihood function, 
2. Take the derivative of  with respect to π.
3. Set the derivative equal to 0 and solve for π to find the MLE. Note that the solution is the maximum of  provided certain conditions hold, like the MLE is not at the boundaries of possible parameter values.

For the field goal example:



where log means natural log.









Therefore, the maximum likelihood estimator of π is the proportion of field goals made. To avoid confusion between a parameter and a statistic, one denotes the estimator as  = /n.

Example: MLEs for μ and σ from a normal distribution

Let x1, …, xn be a random sample from a N(μ,σ2) distribution. Find the MLE of μ and σ2. Remember that the normal distribution is:



Then the likelihood function is:



Taking the log of the likelihood function produces:



where I drop the i = 1 to n part of the summation symbol to make it easier to write out. To find the maximum likelihood estimate of μ, take the derivative with respect to μ and set equal to 0.



Solving for μ produces:



To find the maximum likelihood estimate of σ, take the derivative with respect to σ and set equal to 0.



Solving for σ2 produces:



Likelihood Ratio Test (LRT)

The LRT is a general way to test hypotheses. The LRT statistic, Δ, is the ratio of two likelihood functions. Most often this symbol is given as **Λ**, but I will use Δ here because **Λ** is used in a different way in the factor analysis part of the course.

The numerator of Δis the likelihood function maximized over the parameter space restricted under the null hypothesis. The denominator is the likelihood function maximized over the unrestricted parameter space. The test statistic is written as:



Wilks (1935, 1938) shows that –2log(Λ) can be approximated by a  for a large sample and under H0 where u is the difference in dimension between the alternative and null hypothesis parameter spaces. See Appendix B of my book for more background on the LRT if needed.

Example: Continuing the field goal example; suppose the hypothesis test H0: π = 0.5 vs. Ha: π ≠ 0.5 is of interest.

The numerator of Δ is the maximum possible value of the likelihood function under the null hypothesis. Because π = 0.5 is the null hypothesis, the maximum can be found by just substituting π = 0.5 in the likelihood function:



Then



The denominator of Δ is the maximum possible value of the likelihood function under the null OR alternative hypotheses. Because this includes all possible values of π here, the maximum is achieved when the MLE is substituted for π in the likelihood function! As shown previously, the maximum value is 0.001194.

Therefore,



Then –2log(Λ) = -2log(0.8179) = 0.4020 is the test statistic value. The critical value is  = 3.84 using α = 0.05:

> qchisq(p = 0.95, df = 1)

[1] 3.841459

There is not sufficient evidence to reject the hypothesis that π = 0.5.

In general for any level α, the chi-square distribution and the critical value plotted are



Questions:

* Suppose the ratio is close to 0, what does this say about H0? Explain.
* Suppose the ratio is close to 1, what does this say about H0? Explain.