**Section 1.1.2 – Inference for the probability of success**

Introduction to maximum likelihood estimation

Some of this material is taken from Appendix B.

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 w = 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 in Appendix B):

> w <- 4

> n <- 10

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

> Lik <- pi^w\*(1-pi)^(n-w)

> 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

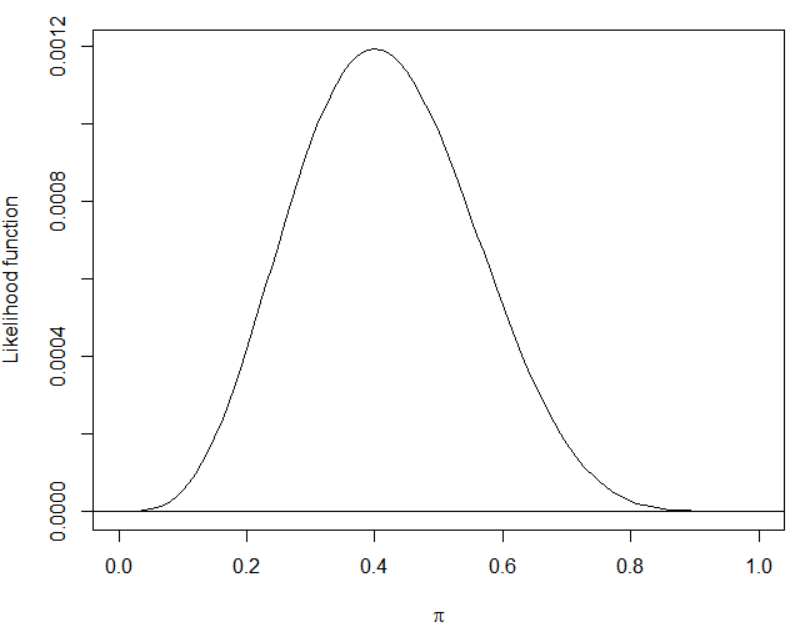
> #Likelihood function plot

> curve(expr = x^w\*(1-x)^(n-w), xlim = c(0,1),

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 maximum likelihood estimate. 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, we will denote the estimator as  = w/n.

Maximum likelihood estimation will be extremely important in this class!!!

For additional examples with maximum likelihood estimation, please see my “Introduction to Mathematical Statistics” course notes available through [www.chrisbilder.com](http://www.chrisbilder.com).

Properties of maximum likelihood estimators

 will vary from sample to sample. We can mathematically quantify this variation for maximum likelihood estimators in general as follows:

* For a large sample, maximum likelihood estimators can be treated as normal random variables.
* For a large sample, the mean (expected value of the estimator) is the parameter the estimator is estimating and the variance of the estimator can be computed from the second derivative of the log likelihood function.

Thus, in general for a maximum likelihood estimator  for θ, we can say that the estimator has an approximate normal distribution with mean θ and an estimated variance



for a large sample Y1, …, Yn. Why do you think this is important to know?

The use of “for a large sample” can also be replaced with the word “asymptotically”. You will often hear these results talked about using the phrase “asymptotic normality of maximum likelihood estimators”.

You are not responsible to do derivations as shown in the next example. This example is helpful for understanding how R will be doing these and more complex calculations.

Example: Field goal kicking



The large sample variance of  is

.

To find this, note that,





 because only the Yi’s are random variables (so W is a random variable)





Since π is a parameter, we replace it with its corresponding estimator to obtain .

Thus, = .

This same result is derived on p. 474 of Casella and Berger (2002).

See Chapter 18 of Ferguson (1996) for more on the “asymptotic normality” of maximum likelihood estimators.