The Binomial Likelihood Function

The **likelihood function** for the binomial model is

$$\mathcal{L}(p \mid n, y) = \binom{n}{y} p^y (1-p)^{n-y} \;.$$

This function involves the parameter p, given the data (the n and y). The discrete data and the statistic y (a count or summation) are known. The likelihood function is not a probability function; but it is a positive function and $0 \le p \le 1$. The left hand side is read "the likelihood of the parameter p, given n and y. <u>Likelihood theory and the likelihood function are fundamental in the statistical sciences.</u>

Note the similarity between the probability function and the likelihood function; the right hand sides are the same. The difference between the two functions is the conditioning of the left hand sides. The probability function returns probabilities of the data, given the sample size and the parameters, while the likelihood function gives the relative likelihoods for different values of the parameter, given the sample size and the data.

Some experience will allow an understanding of relative likelihood. Consider n flips of an unfair coin, whereby y are "HEADS." Let n = 11 flips and y = 7 heads. Thus, 4 are tails (by subtraction or observation). The conceptual motivation behind parameter estimation is to pick that value of the parameter that is "most likely", given the data. So, we can use the likelihood to evaluate some relative likelihoods for several values of p. We find that $\binom{n}{y}$ = 330. Then,

p	Likelihood
0.3	0.0173
0.5	0.1611
0.7	0.2201
0.8	0.1107.

Given the data (n and y), the value p = 0.3 is relatively unlikely as the underlying parameter. The values of p = 0.5 or 0.8 are far more likely, relative to p = 0.3. Of the values consider (above), the value p = 0.7 is the most likely. So, of the 4 values considered above, given the data observed, which would you select as the estimate of the underlying parameter? Why?

While this function is quite useful, there are a host of reasons why the *log-likelihood function* is often more useful. The **log-likelihood** for the binomial model is

$$\log_e(\mathcal{L}) = \log_e\left(\mathcal{L}(p \mid n, y)\right) = \log_e \binom{n}{y} + y \cdot \log_e(p) + (n - y) \cdot \log_e(1 - p).$$

One advantage of the log-likelihood is that the terms are additive. Note, too, that the binomial coefficient does not contain the parameter p. We will see that this term is a constant and can often be omitted. Note, too, that the log-likelihood function is in the negative quadrant because of the logarithm of a number between 0 and 1 is negative.

An Example:

Consider the example above; n flips of an unfair coin, whereby y are "HEADS." Let $n = 11$ flips and $y = 7$ heads. Thus, 4 are tails (by subtraction or observation).

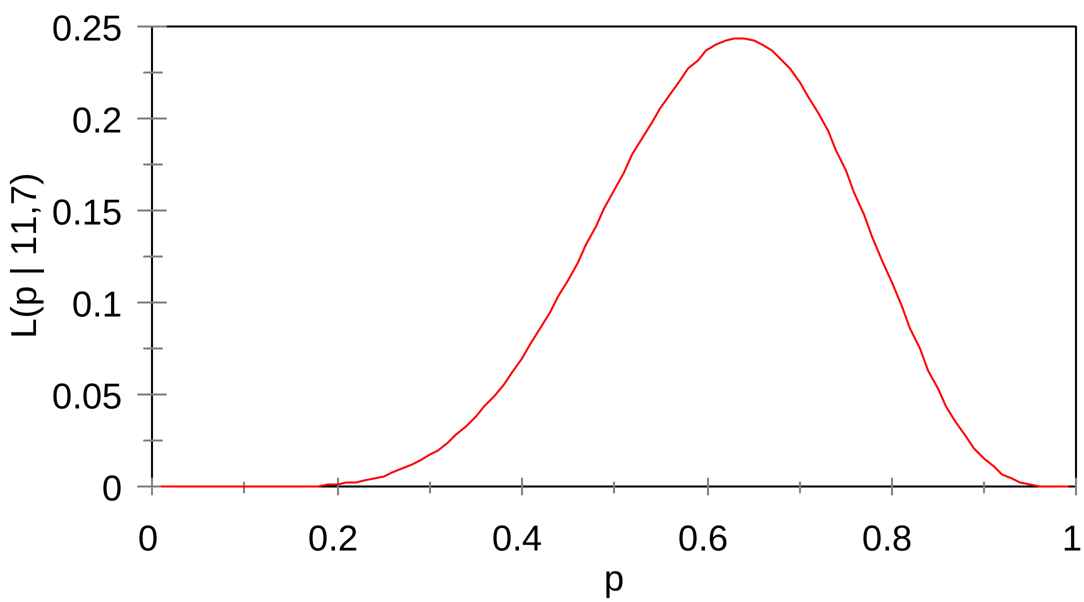
The log-likelihood function is

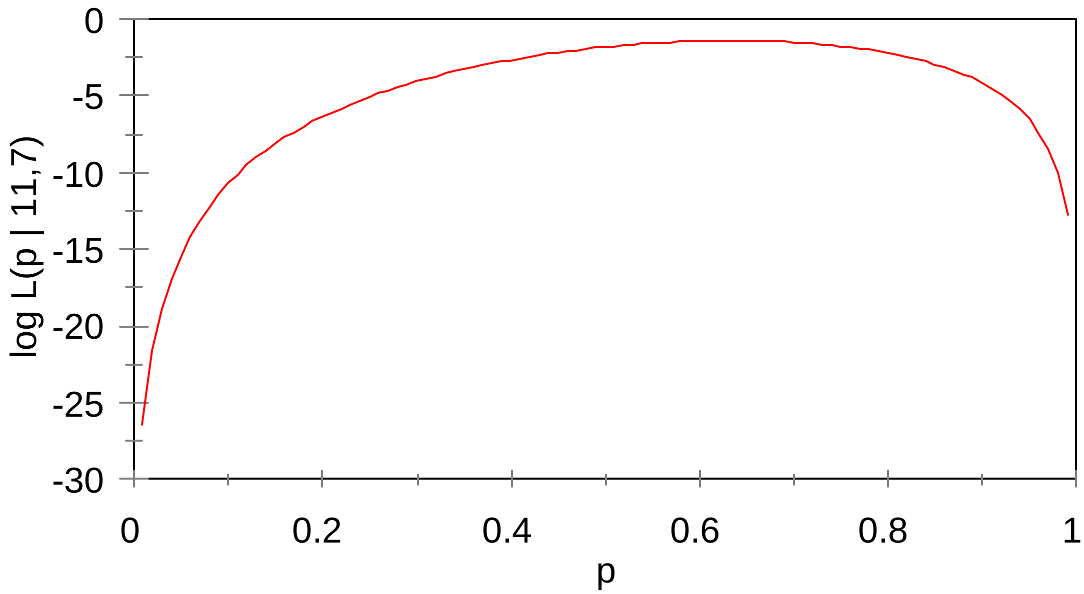
$$\log_e(\mathcal{L}(p \mid n, y) = \log_e(C) + y \log_e(p) + (n - y) \log_e(1 - p)$$

or

$$\log_e(\mathcal{L}(p \mid 11, 7) = \log_e(C) + 7 \log_e(p) + (4) \log_e(1 - p),$$

where $0 \leq p \leq 1$ and represents the underlying probability of a head.





0
-5
-10
-15
-20
-25
-30
log L(p | 11,7)
0
0.2
0.4
0.6
0.8
1
p