Mathematical Biostatistics Boot Camp: Lecture 6, Likelihood

Brian Caffo

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

September 19, 2012

Table of contents

Likelihood

- A common and fruitful approach to statistics is to assume that the data arises from a family of distributions indexed by a parameter that represents a useful summary of the distribution
- The **likelihood** of a collection of data is the joint density evaluated as a function of the parameters with the data fixed
- Likelihood analysis of data uses the likelihood to perform inference regarding the unknown parameter

Likelihood

Given a statistical probability mass function or density, say $f(x, \theta)$, where θ is an unknown parameter, the **likelihood** is f viewed as a function of θ for a fixed, observed value of x.

Interpretations of likelihoods

The likelihood has the following properties:

1. Ratios of likelihood values measure the relative **evidence** of one value of the unknown parameter to another.
2. Given a statistical model and observed data, all of the relevant information contained in the data regarding the unknown parameter is contained in the likelihood.
3. If $\{X_i\}$ are independent random variables, then their likelihoods multiply. That is, the likelihood of the parameters given all of the X_i is simply the product of the individual likelihoods.

Example

- Suppose that we flip a coin with success probability θ
- Recall that the mass function for x

$$f(x, \theta) = \theta^x (1 - \theta)^{1-x} \quad \text{for} \quad \theta \in [0, 1].$$

where x is either 0 (Tails) or 1 (Heads)

- Suppose that the result is a head
- The likelihood is

$$\mathcal{L}(\theta, 1) = \theta^1 (1 - \theta)^{1-1} = \theta \quad \text{for} \quad \theta \in [0, 1].$$

- Therefore, $\mathcal{L}(.5, 1)/\mathcal{L}(.25, 1) = 2$,
- There is twice as much evidence supporting the hypothesis that $\theta = .5$ to the hypothesis that $\theta = .25$

Example continued

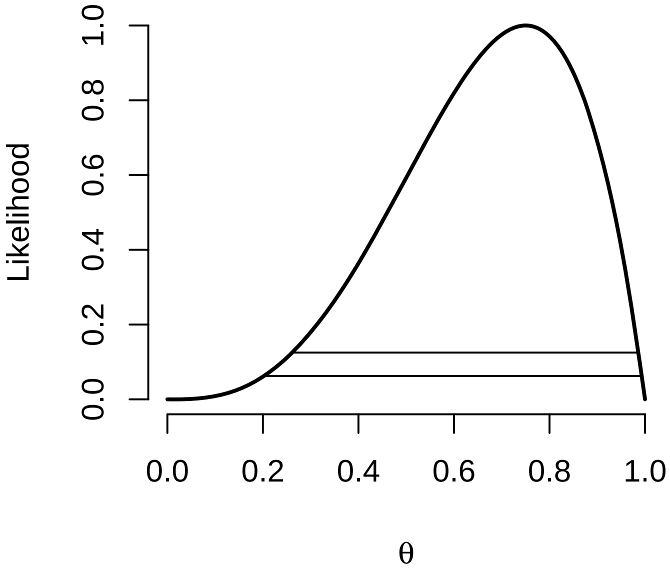
- Suppose now that we flip our coin from the previous example 4 times and get the sequence 1, 0, 1, 1
- The likelihood is:

$$\begin{aligned}\mathcal{L}(\theta, 1, 0, 1, 1) &= \theta^1(1-\theta)^{1-1}\theta^0(1-\theta)^{1-0} \\ &\times \theta^1(1-\theta)^{1-1}\theta^1(1-\theta)^{1-1} \\ &= \theta^3(1-\theta)^1\end{aligned}$$

- This likelihood only depends on the total number of heads and the total number of tails; we might write $\mathcal{L}(\theta, 1, 3)$ for shorthand
- Now consider $\mathcal{L}(.5, 1, 3)/\mathcal{L}(.25, 1, 3) = 5.33$
- There is over five times as much evidence supporting the hypothesis that $\theta = .5$ over that $\theta = .25$

Plotting likelihoods

- Generally, we want to consider all the values of θ between 0 and 1
- A **likelihood plot** displays θ by $\mathcal{L}(\theta, x)$
- Usually, it is divided by its maximum value so that its height is 1
- Because the likelihood measures *relative evidence*, dividing the curve by its maximum value (or any other value for that matter) does not change its interpretation


Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
θ

Maximum likelihood

- The value of θ where the curve reaches its maximum has a special meaning
- It is the value of θ that is most well supported by the data
- This point is called the **maximum likelihood estimate** (or MLE) of θ

$$MLE = \mathrm{argmax}_\theta \mathcal{L}(\theta, x).$$

- Another interpretation of the MLE is that it is the value of θ that would make the data that we observed most probable

Maximum likelihood, coin example

- The maximum likelihood estimate for θ is always the proportion of heads
- Proof: Let x be the number of heads and n be the number of trials
- Recall

$$\mathcal{L}(\theta, x) = \theta^x (1-\theta)^{n-x}$$

- It's easier to maximize the **log-likelihood**

$$l(\theta, x) = x \log(\theta) + (n - x) \log(1 - \theta)$$

Continued

- Taking the derivative we get

$$\frac{d}{d\theta} l(\theta, x) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

- Setting equal to zero implies

$$(1 - \frac{x}{n})\theta = (1-\theta)\frac{x}{n}$$

- Which is clearly solved at $\theta = \frac{x}{n}$
- Notice that the second derivative

$$\frac{d^2}{d\theta^2} l(\theta, x) = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} < 0$$

provided that x is not 0 or n (do these cases on your own)

What constitutes strong evidence?

- Again imagine an experiment where a person repeatedly flips a coin
- Consider the possibility that we are entertaining three hypotheses: $H_1 : \theta = 0$, $H_2 : \theta = .5$, and $H_3 : \theta = 1$

Outcome X	$P(X \mid H_1)$	$P(X \mid H_2)$	$P(X \mid H_3)$	$\mathcal{L}(H_1)/\mathcal{L}(H_2)$	$\mathcal{L}(H_3)/\mathcal{L}(H_2)$
H	0	.5	1	0	2
T	1	.5	0	2	0
HH	0	.25	1	0	4
HT	0	.25	0	0	0
TH	0	.25	0	0	0
TT	1	.25	0	4	0
HHH	0	.125	1	0	8
HHT	0	.125	0	0	0
HTH	0	.125	0	0	0
THH	0	.125	0	0	0
HTT	0	.125	0	0	0
THT	0	.125	0	0	0
TTH	0	.125	0	0	0
TTT	1	.125	0	8	0

Benchmarks

- Using this example as a guide, researchers tend to think of a likelihood ratio
  - of 8 as being moderate evidence
  - of 16 as being moderately strong evidence
  - of 32 as being strong evidence

  of one hypothesis over another
- Because of this, it is common to draw reference lines at these values on likelihood plots
- Parameter values above the $1/8$ reference line, for example, are such that no other point is more than 8 times better supported given the data