Stat 704 Data Analysis I
Likelihood

Dr. Yen-Yi Ho

Department of Statistics, University of South Carolina

Outline

- Defining Likelihood
- Interpreting Likelihoods
- Plotting Likelihoods
- Maximum Likelihood
- Interpreting Likelihood ratios

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

Interpretations of Likelihoods

The likelihood has the following properties:

- Ratios of likelihood values measure the relative **evidence** of one value of the unknown parameter to another.
- 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.
- If $\{X_i\}$ are independent events, 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.

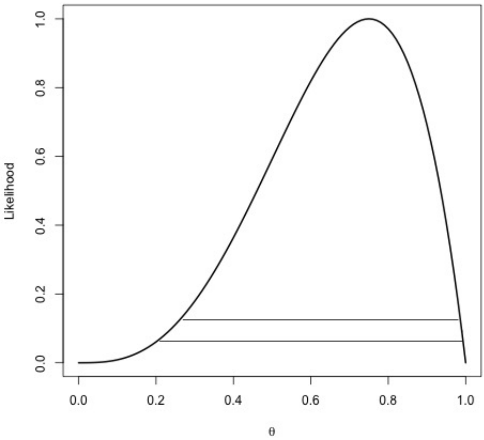
Example

- Suppose we flip a coin 4 times and get the sequence 1, 0, 1, 1
- The likelihood is: $L(\theta; \textit{data}) = L(\theta; 1, 0, 1, 1) = \theta^3(1-\theta)^1$
- Now consider $\frac{L(0.5; \textit{data})}{L(0.25; \textit{data})} = 5.33$
- There is over five times as much evidence supporting the hypothesis that $\theta = 0.5$ versus that $\theta = 0.25$

Plotting Likelihood

- Generally, we want to consider all possible values of θ.
- A **likelihood plot** displays θ by $L(\theta; data)$
- Usually, it is divided by its maximum value so that its height is 1
- Because the likelihood measure relative evidence, dividing the curve by its maximum value (or another other value) does not change its interpretation

Likelihood



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

$$MLE = \mathrm{argmax}_\theta L(\theta; data).$$

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

MLE, Coin Example

- The maximum likelihood estimate for θ in the coin example is always that proportion of heads
- Proof: Let x be the number of head and n be the number of trials
- Recall
$$L(\theta, x) = \theta^x (1-\theta)^{n-x}$$
- 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