To Do List

Read Chapter 2, Sections 2.1-2.3

Do Problems 1- 4 in Chapter 2

Today’s Class: Parameter Estimation

1) Definition of a (Point) Estimate of an Unknown Parameter

2) Method of Maximum Likelihood

i) Definition of the Likelihood Function

ii) Definition of the Maximum Likelihood Estimate

iii) Definition of the Relative Likelihood Function

iv) Definition of the Log Likelihood Function

From Last Day: Method of Maximum Likelihood

Suppose we have observed data y and we assume a statistical model $f(y; \theta)$, $\theta \in \Omega$ which is completely known except for an unknown parameter θ.

The method of maximum likelihood is a method for estimating this unknown parameter θ.

The method is based on the idea that values of θ that make the observed data probable are the values of θ which are most plausible.

Definitions from Last Day

1) The **likelihood function** for θ is

$$L(\theta) = L(\theta; y)$$
$$= P(\text{observing the data } y;\ \theta) \quad \text{for } \theta \in \Omega$$

2) The **maximum likelihood estimate** is the value of θ which maximizes $L(\theta)$.

Likelihood Function and Maximum Likelihood Estimate for Binomial Data - Summary

Let *Y* = number of successes in *n* Bernoulli trials with P(Success) = θ. Then *Y* ~ Binomial(*n*, θ).

Suppose a Binomial experiment is conducted and *y* successes are observed. The likelihood function for θ based on the observed data is

$$L(\theta) = P(Y = y; \theta) = P(\text{observing } y \text{ successes in } n \text{ trials})$$

$$= \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \quad for \;\; 0 < \theta < 1.$$

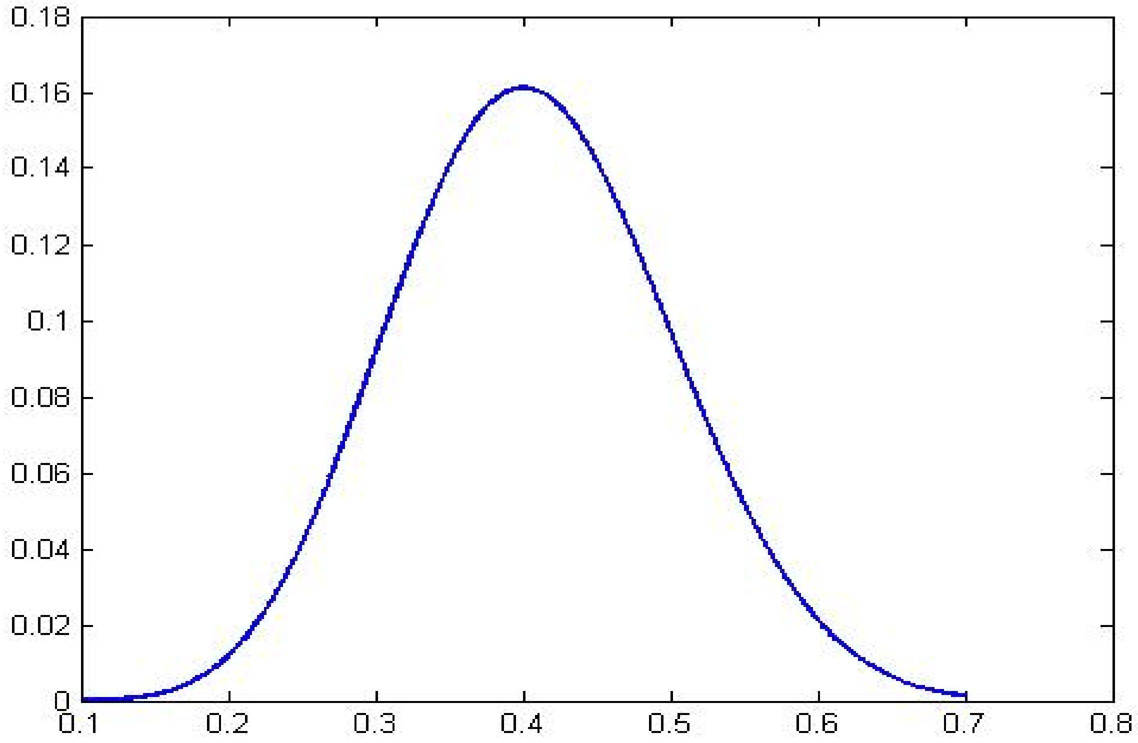
and the maximum likelihood estimate of θ is

$$\hat{\theta} = \frac{y}{n}$$

$L(\theta)$ for Coin Example

Graph of $L(\theta)=\binom{25}{10}\theta^{10}(1-\theta)^{15}$

$L(\theta)$



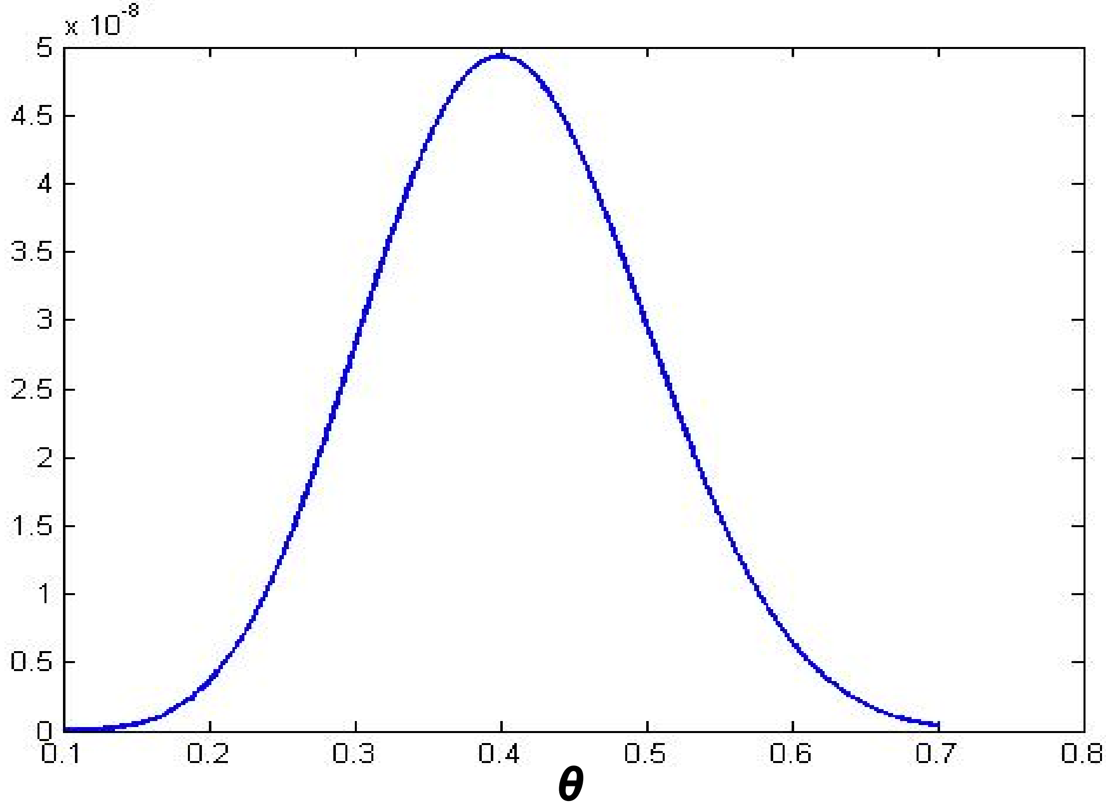
θ

L(θ) is maximized at θ = 10/25 = 0.4

Coin Example

Graph of $\theta^{10}(1-\theta)^{15}$

$$\theta^{10}(1-\theta)^{15}$$



L(*θ*) is maximized at *θ* = 10/25 = 0.4

The Likelihood Function

The shape of L(θ) and the value of θ at which the maximum value occurs are not affected if we multiply L(θ) by a constant.

Only the scale of the y axis changes.

The Likelihood Ratio

The value of the ratio

$$\frac{L(\theta_1)}{L(\theta_2)}$$

is also unaffected if we multiply L(*θ*) by a constant.

If you're a likelihood value, (relative) size matters.

The relative value at two different values of the parameter, e.g. $\frac{L(\theta_1)}{L(\theta_2)}$ **, is what is most important.**

This ratio indicates how much more or how much less consistent the data are with the value $\theta = \theta_1$ **as compared to the value** $\theta = \theta_2$**.**

Definition of the Likelihood Function

Therefore we can also define the likelihood function for $\boldsymbol{\theta}$ as

$$L(\theta) = L(\theta; y)$$

$$= kP(\text{observing the data } y;\ \theta) \quad \text{for } \theta \in \Omega$$

where k is a positive constant, not depending on θ, which can be chosen to simplify $L(\theta)$.

Binomial Likelihood Function

Therefore for Binomial data we can define L(*θ*) as

$$L(\theta)=\binom{n}{y}\theta^{y}(1-\theta)^{n-y} \quad \text{and} \quad \theta^{y}(1-\theta)^{n-y}$$

or more simply as

$$L(\theta)=\theta^{y}(1-\theta)^{n-y} \quad \text{for } 0<\theta<1.$$

Both are maximized at *θ* =*y*/*n* and have the same shape.

(Only the *y* axis gets relabeled!)

Relative Likelihood Function

The relative likelihood function is defined as

$$R(\theta)=\frac{L(\theta)}{L(\hat{\theta})}, \quad \theta \in \Omega.$$

$$\text{Note: } 0 \le R(\theta) \le 1 \quad \text{for all } \theta \in \Omega$$

$$\text{and } R(\hat{\theta}) = 1.$$

Relative Likelihood Function for Binomial

For Binomial data the relative likelihood function is

$$R(\theta) = \frac{\theta^y (1-\theta)^{n-y}}{\hat{\theta}^y (1-\hat{\theta})^{n-y}} \quad \text{for } 0 < \theta < 1$$

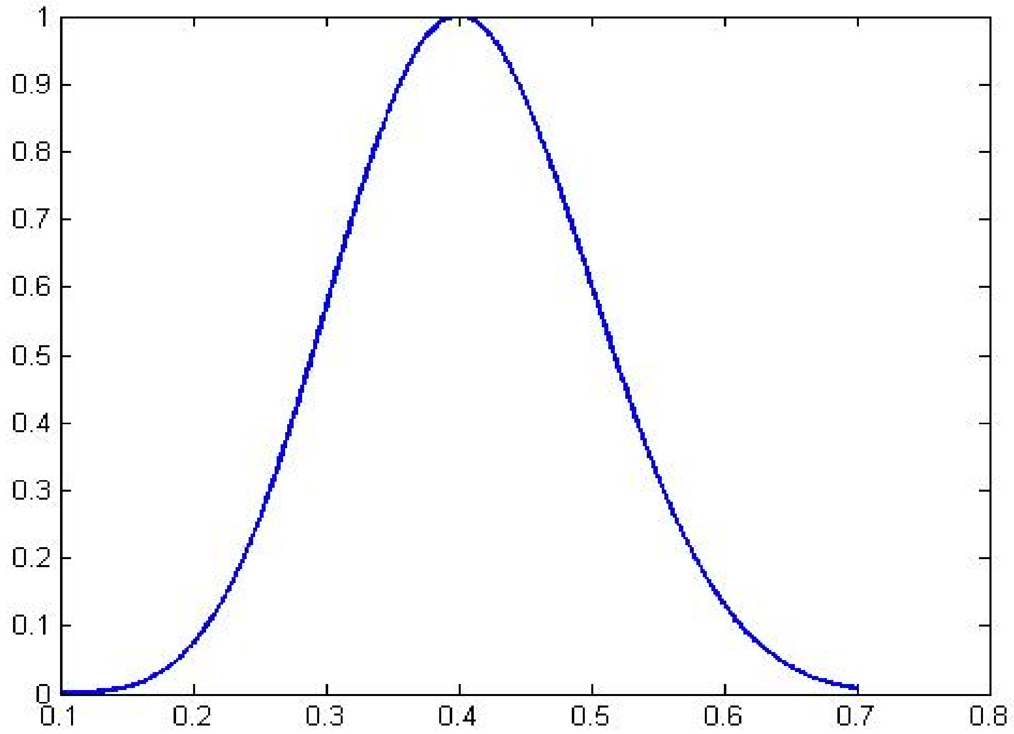
$$\text{where} \quad \hat{\theta} = \frac{y}{n}$$

R(*θ*) for the Coin Example

Graph of

$$R(\theta) = \frac{\theta^{10}(1-\theta)^{15}}{(0.4)^{10}(0.6)^{15}}$$

R(*θ*)



θ

Log Likelihood Function

The log likelihood function is defined as

$$l(\theta) = \log L(\theta), \quad \theta \in \Omega.$$

Note:

log = ln = natural log

(Mathematicians only ever use natural log. Why?)

Binomial Log Likelihood

Since the Binomial likelihood function is

$$L(\theta) = \theta^{y}(1-\theta)^{n-y} \quad \text{for } 0 < \theta < 1$$

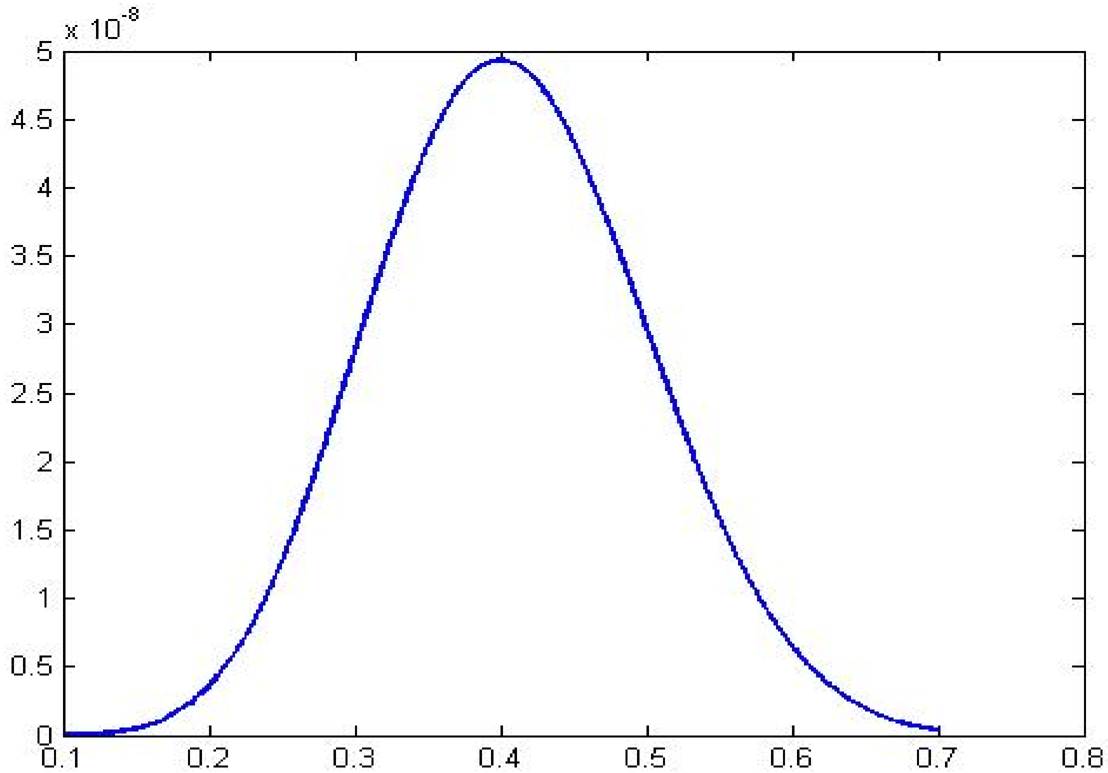
therefore the Binomial log likelihood function is

$$\begin{aligned} l(\theta) &= \log\left[L(\theta)\right] \\ &= \log[\theta^{y}(1-\theta)^{n-y}] \\ &= y\log\theta + (n-y)\log(1-\theta) \quad \text{for } 0 < \theta < 1 \end{aligned}$$

Coin Example

Graph of $L(\theta) = \theta^{10}(1-\theta)^{15}$

$L(\theta)$

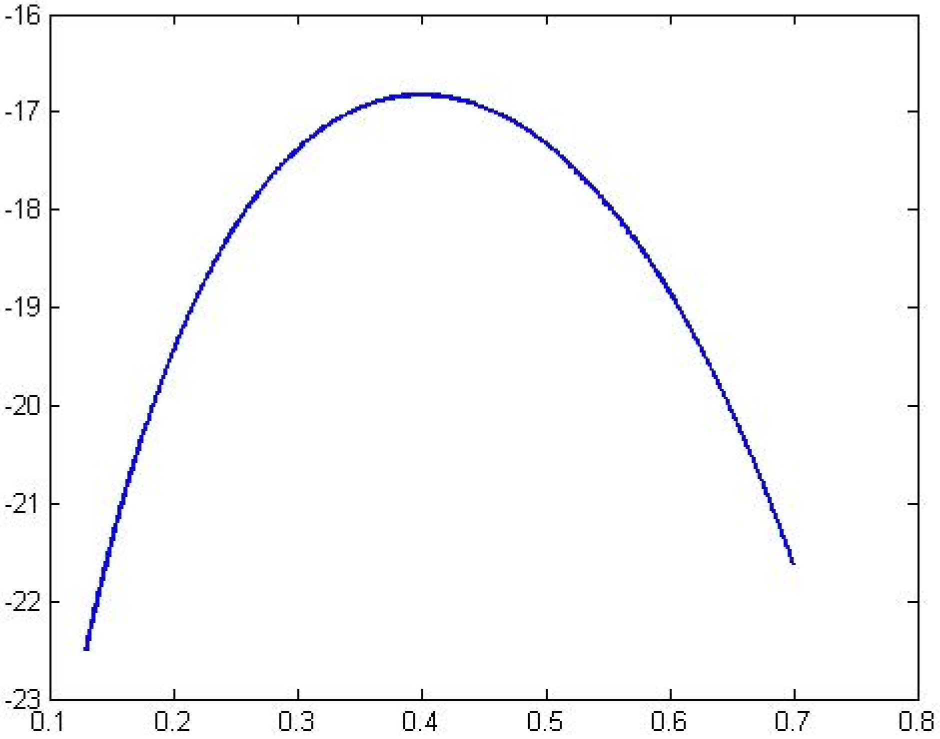


θ

Graph of *l*(*θ*) for Coin Example

Graph of $l(\theta)=10\log\theta+15\log(1-\theta)$

$l(\theta)$



θ

Log Likelihood Function

The graph of the log likelihood function $l(\theta)$ is typically quadratic in shape.

Often it is easier to maximize the log likelihood function $l(\theta)$ rather than the likelihood function $L(\theta)$.

(Sum rule for differentiation is easier to use than product rule.)

Likelihood Function for Independent Experiments

Suppose we have two independent data sets $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ corresponding to independent random variables $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$.

Since P($\boldsymbol{Y}_1 = \boldsymbol{y}_1$,$\boldsymbol{Y}_2 = \boldsymbol{y}_2$; $\boldsymbol{\theta}$)

= P($\boldsymbol{Y}_1 = \boldsymbol{y}_1$; $\boldsymbol{\theta}$) P($\boldsymbol{Y}_2 = \boldsymbol{y}_2$; $\boldsymbol{\theta}$)

the (combined) likelihood function for $\boldsymbol{\theta}$ based on the data $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ is

$$L(\theta) = L_1(\theta) \times L_2(\theta) \quad \theta \in \Omega$$

$$\text{where } L_i(\theta) = P(\boldsymbol{Y}_i = \boldsymbol{y}_i ; \theta)$$

Likelihood Function and Maximum Likelihood Estimate for Poisson Model

Suppose we observe data $y_1, y_2, \ldots, y_n$.

Suppose also that from past experience we know that it is reasonable to assume that these data represent a set of independent and identically distributed observations from a Poisson(θ) model.

We want to find the maximum likelihood estimate of θ based on the data $y_1, y_2, \ldots, y_n$.

Poisson Relative Likelihood Function (Course Notes page 53)

For Poisson data $y_1, y_2, \ldots, y_n$

$$L(\theta) = \theta^{n\bar{y}} e^{-n\theta} \quad for \;\; \theta > 0$$

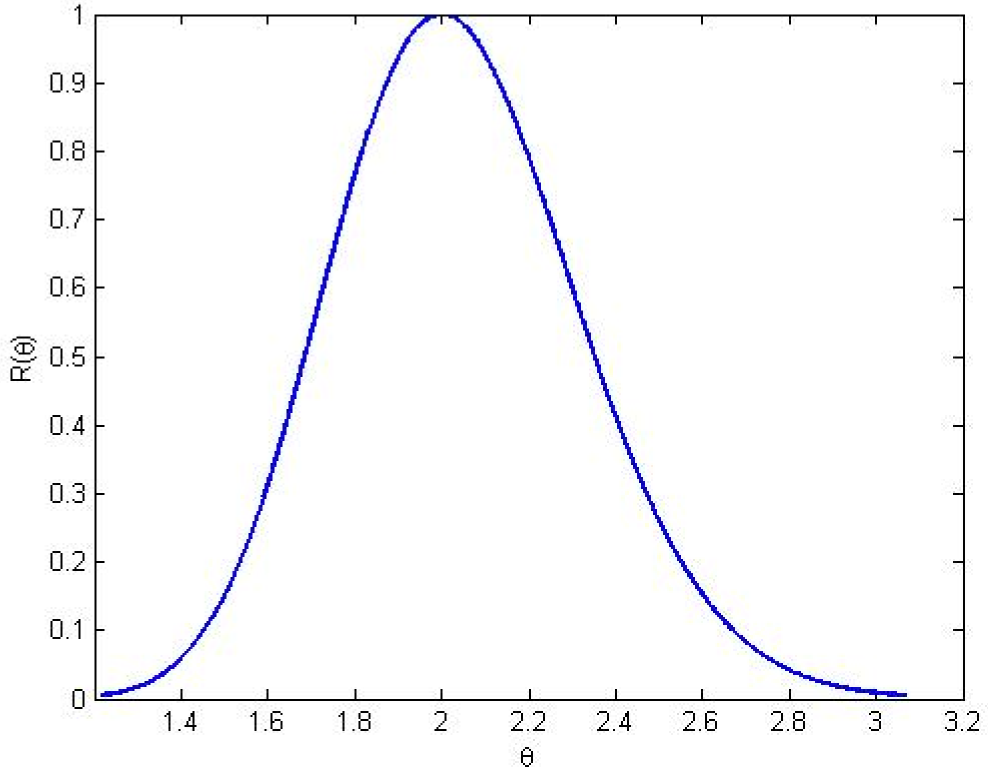
(ignoring constants with respect to θ) and

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \frac{\theta^{n\bar{y}} e^{-n\theta}}{\hat{\theta}^{n\bar{y}} e^{-n\hat{\theta}}} \quad for \;\; \theta > 0$$

$$\text{where } \hat{\theta} = \bar{y}$$

Poisson Relative Likelihood

$$n = 25 \text{ and } \bar{y} = 2$$



Example 2.2.4

Please see Example 2.2.4, pages 55-56 of the Course Notes.

A Poisson type example in which the maximum likelihood estimate must be found numerically.