Introduction to Categorical Data Analysis

DATA SCIENCE BOOTCAMP

Categorical Data Analysis

Methods for response *(y)* variable having scale that is categorical

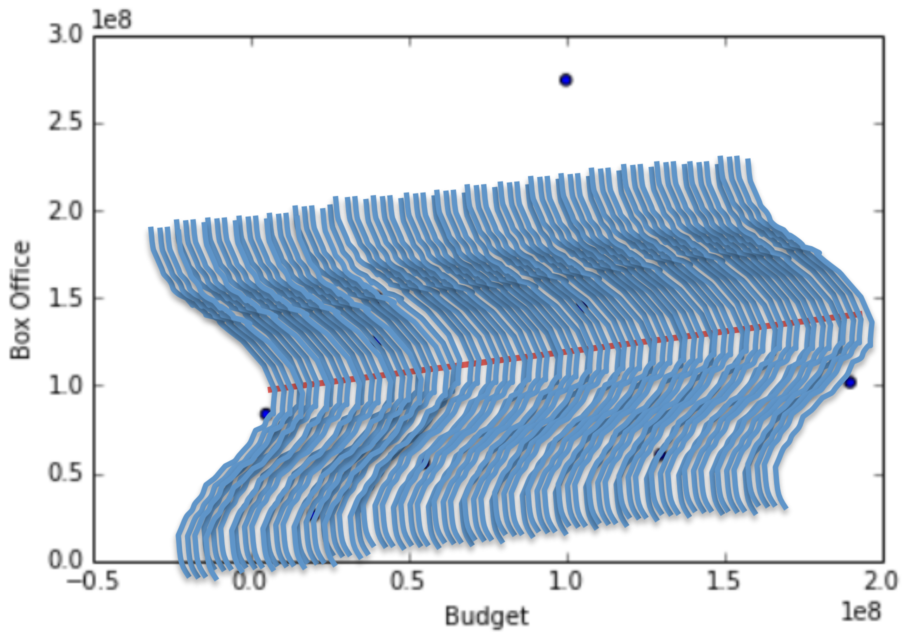
Examples:

- Binary (Democrat or Republican)
  (Success or Not Success)
- Nominal (Favorite music genre: rock, classical, jazz)
- Ordinal (Rank: 1^{st}, 2^{nd}, 3^{rd})

For these problems, linear regression is not the correct approach.

Probability Distributions for Categorical Data

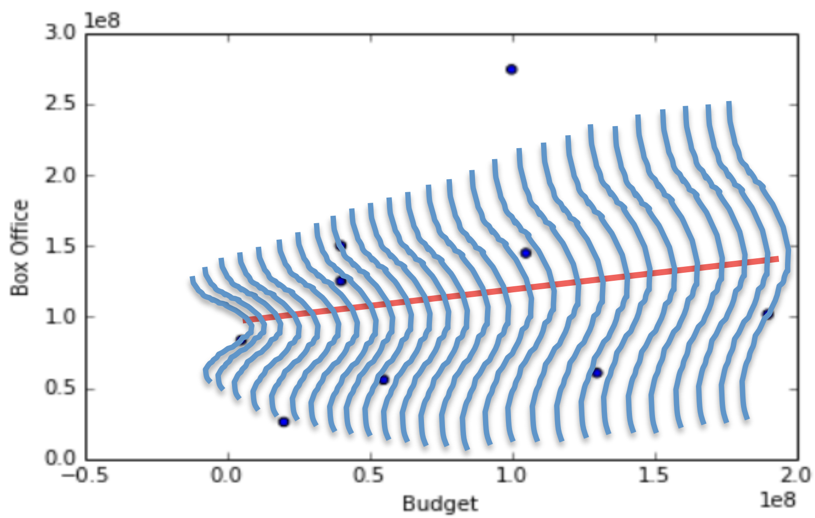
In linear regression: response y is continuous, and the (conditional) response $y \mid x$ is normally distributed.



Probability Distributions for Categorical Data

In linear regression: response y is continuous, and the (conditional) response $y \mid x$ is normally distributed.

In models where y is categorical, the response $y \mid x$ can take other distributions.



Let’s first start simple.

Binomial Experiments

The experiment of tossing a coin n times is an example of a binomial experiment.

Binomial Experiments

The experiment of tossing a coin n times is an example of a binomial experiment.

Three conditions:

- n independent trials

- The outcome of a single trial is classified as 1 of 2 types:
    - *Success*
    - *Not Success*

- For each trial, the probability of success is the same:
    - $P(Success) = \pi$
    - Note: π is called the parameter of the experiment.

Binomial Experiments

Y: random variable, represents the # of successes out of n trials

Example:

- 200 voters
- Each person votes:

  Democrat – with $\mathrm{p} = 0.55$

  Republican – with $\mathrm{p} = 0.45$

Binomial Experiments

Y: random variable, represents the # of successes out of n trials

Example:

- 200 voters
- Each person votes:

  Democrat – with p = 0.55

  Republican – with p = 0.45
- Then # Democrat votes ~ $Binomial\ (n=200, \pi = 0.55)$

Binomial Experiments

Y: random variable, represents the # of successes out of n trials

Example:

- 200 voters
- Each person votes:

  Democrat – with p = 0.55

  Republican – with p = 0.45
- Then # Democrat votes ~ $Binomial\ (n=200, \pi = 0.55)$
- Note: if $n = 1$ (just one trial), then $Y \sim Bernoulli$ distribution

Binomial Experiments

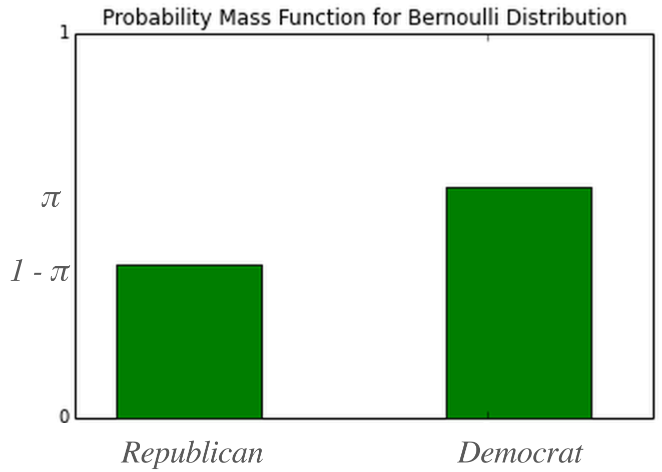
Y: random variable, represents the # of successes out of n trials

Example:

- 200 voters
- Each person votes:

  Democrat – with p = 0.55

  Republican – with p = 0.45
- Then # Democrat votes ~ *Binomial* (n=200, π = 0.55)
- Note: if $n = 1$ (just one trial), then Y ~ *Bernoulli* distribution



The Binomial Distribution

More generally, the random variable

$$Y \sim Binomial\ (n, p)$$

has probability mass function:

$$P(Y = y) = \frac{n!}{y!(n-y)!}\ \pi^{y} (1-\pi)^{n-y}$$

The Binomial Distribution

More generally, the random variable

$$Y \sim Binomial\ (n, \pi)$$

has probability mass function:

$$P(Y = y) = \frac{n!}{y!(n-y)!} \pi^{y} (1-\pi)^{n-y}$$

In general, for a binomial random variable Y,

- The <u>Expected Value</u> is:

$$E(Y) = \mu_Y = n\pi$$

- The <u>Variance</u> is:

$$Var(Y) = \sigma_Y^2 = n\pi(1-\pi)$$

The Binomial Distribution

More generally, the random variable

$$Y \sim Binomial\ (n, \pi)$$

has probability mass function:

$$P(Y = y) = \binom{n}{y} \pi^{y} (1 - \pi)^{n-y}$$

In general, for a binomial random variable Y,

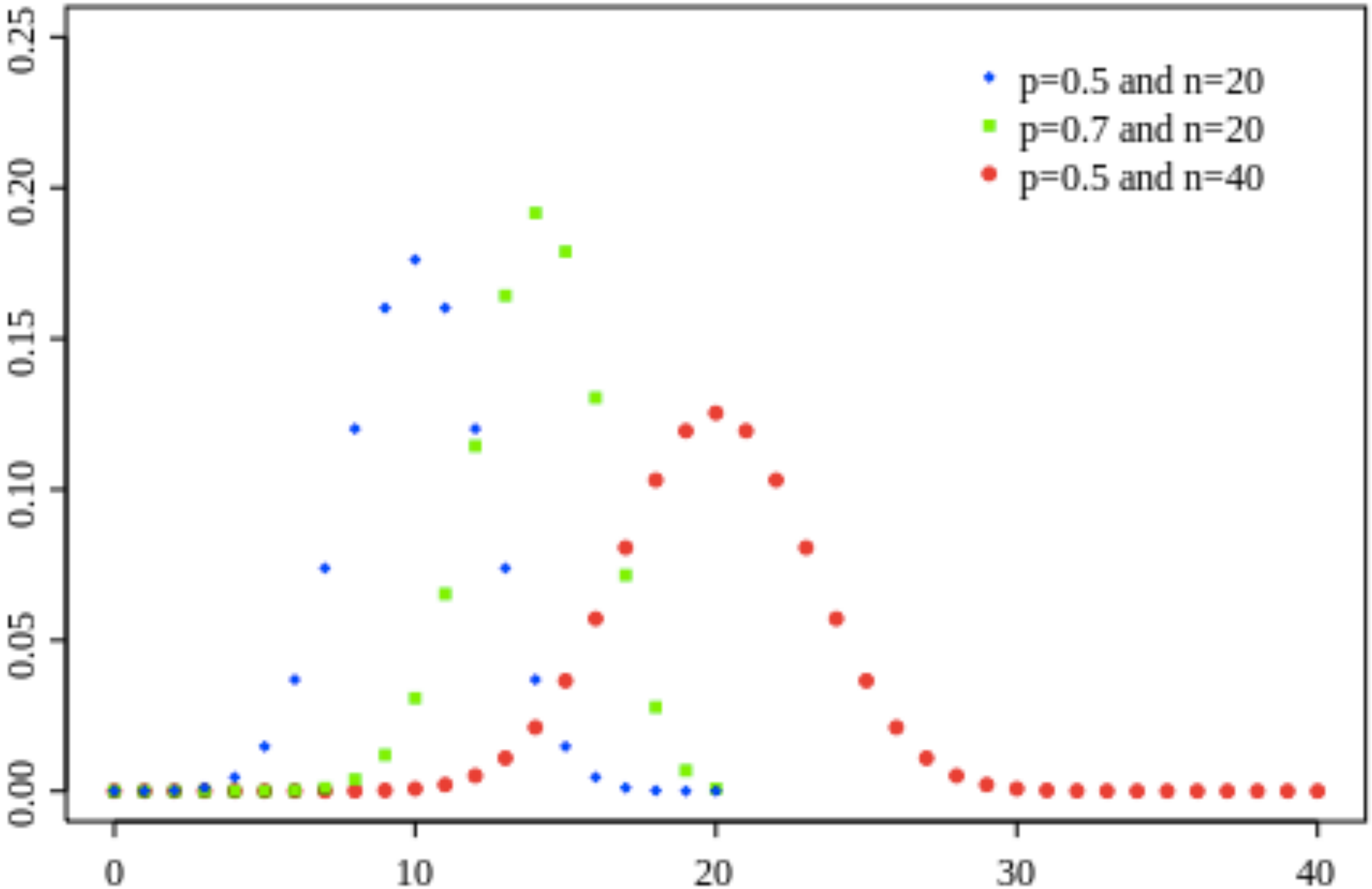
- The Expected Value is:

$$E(Y) = \mu_Y = n\pi$$

- The Variance is:

$$Var(Y) = \sigma_Y^2 = n\pi(1-\pi)$$

The Binomial Distribution



Example of a Binomial Distribution

Example of a Binomial Distribution

Let Y = # of people who voted Democrat.

$n = 3$ voters.

Suppose $P(vote\ Dem) = 0.5$

Possible values of Y are:

Example of a Binomial Distribution

Let Y = # of people who voted Democrat.

$n = 3$ voters.

Suppose $P(vote\ Dem) = 0.5$

Possible values of Y are: $0, 1, 2, 3$

Example of a Binomial Distribution

Let Y = # of people who voted Democrat.

$n = 3$ voters.

Suppose $P(vote\ Dem) = 0.5$

Possible values of Y are: $0, 1, 2, 3$

Then Y follows a binomial distribution:

$$Y \sim Binomial\ (n = 3, \pi = 0.5)$$

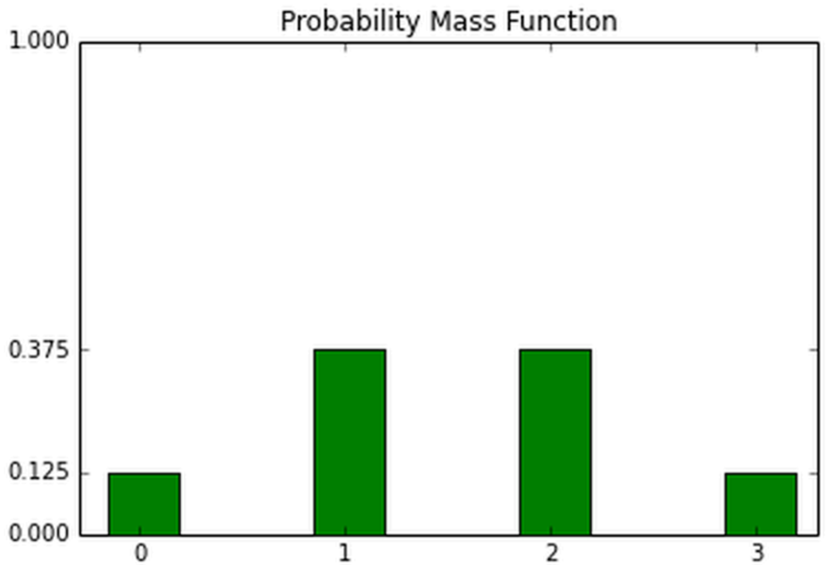
Example of a Binomial Distribution

Let Y = # of people who voted Democrat.
$n = 3$ voters.
Suppose $P(vote\ Dem) = 0.5$

Possible values of Y are: $0, 1, 2, 3$

Then Y follows a binomial distribution:

$Y \sim Binomial\ (n = 3, \pi = 0.5)$



Example of a Binomial Distribution

Let Y = # of people who voted Democrat.

$n = 3$ voters.

Suppose $P(vote\ Dem) = 0.5$

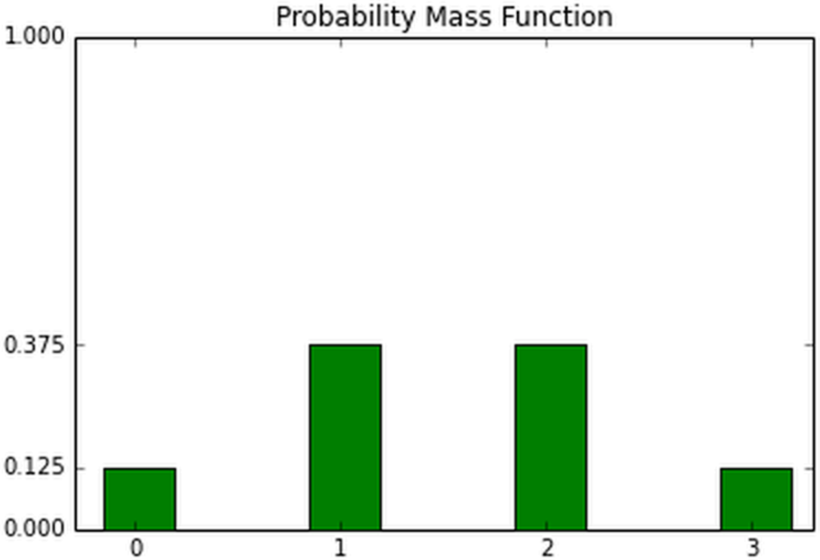
Possible values of Y are: $0, 1, 2, 3$

Then Y follows a binomial distribution:

$$Y \sim Binomial\ (n = 3, \pi = 0.5)$$

$$P(Y = y) = \frac{3!}{y!(3 - y)!}\ (1/2)^{y}\,(1/2)^{3-y}$$

$$= \frac{3!}{y!(3 - y)!}\ (1/8)$$



Thinking with Likelihoods

DATA SCIENCE BOOTCAMP

Thinking with Likelihoods

Before, we were thinking about the probability of various possible outcomes of Y, given a parameter π (and n data points).

Given π, what is is the probability of seeing *y # successes*, $P(Y=y)$?

Given $P(vote\ Dem) = 0.5$, what is is the probability of seeing *y # Democrat votes*, $P(Y=y)$?

Thinking with Likelihoods

Before, we were thinking about the probability of various possible outcomes of Y, given a parameter π.

Given π, what is is the probability of seeing y # *successes*, $P(Y=y)$?

In real life, we're often more interested in going the other way. So now let's flip our thinking:

Given $Y = y$ # *successes*, what is the underlying proportion π?

Thinking with Likelihoods

Before, we were thinking about the probability of various possible outcomes of Y, given a parameter π.

Given π, what is is the probability of seeing y *# successes*, $P(Y=y)$?

In real life, we're often more interested in going the other way. So now let's flip our thinking:

Given $Y = y$ *# successes*, what is the underlying proportion π?

Example: on Wall St, I count 100 people, 15 of whom are women. What is the proportion of women on Wall St?

Thinking with Likelihoods: The Idea

Thinking with Likelihoods: The Idea

- There is an unknown parameter π (like a "state of nature")
- Have data, i.e. observations Y (or Y_1, Y_2, Y_3, …)
- Assume a model, i.e. $Y \sim Binomial\ (n, \pi)$

Thinking with Likelihoods: The Idea

- There is an unknown parameter π (like a "state of nature")
- Have data, i.e. observations Y (or Y_1, Y_2, Y_3, …)
- Assume a model, i.e. $Y \sim Binomial\ (n, \pi)$

Goal: want $P(\pi \mid Y = y)$

Thinking with Likelihoods: The Idea

- There is an unknown parameter π (like a "state of nature")
- Have data, i.e. observations Y (or Y_1, Y_2, Y_3, …)
- Assume a model, i.e. $Y \sim Binomial\ (n, \pi)$

Goal: want $P(\pi \mid Y = y)$

After we have data, we want to know the probability of seeing various values of π.

Thinking with Likelihoods: The Idea

- There is an unknown parameter π (like a "state of nature")
- Have data, i.e. observations Y (or Y_1, Y_2, Y_3, …)
- Assume a model, i.e. $Y \sim Binomial\ (n, \pi)$

Goal: want $P(\pi \mid Y = y)$

After we have data, we want to know the probability of seeing various values of π.

In particular, we'd like to find the probability that is *maximal*.

Thinking with Likelihoods: The Idea

Goal: want $P(\pi \mid Y = y)$

After we have data, we want to know the probability of seeing various values of π.

Two ways:

Thinking with Likelihoods: The Idea

Goal: want $P(\pi \mid Y = y)$

After we have data, we want to know the probability of seeing various values of π.

Two ways:

1. Bayes Theorem.

Thinking with Likelihoods: The Idea

Goal: want $P(\pi \mid Y = y)$

After we have data, we want to know the probability of seeing various values of π.

Two ways:

1. Bayes Theorem.

   $P(\pi \mid Y = y)$ *is proportional to* $P(\pi)\, P(Y = y \mid \pi)$

Thinking with Likelihoods: The Idea

Goal: want $P(\pi \mid Y = y)$

After we have data, we want to know the probability of seeing various values of π.

Two ways:

1. Bayes Theorem.

$$P(\pi \mid Y = y) \textit{ is proportional to } \underbrace{P(\pi)}_{\text{prior}} \, \underbrace{P(Y = y \mid \pi)}_{\text{likelihood}}$$

Thinking with Likelihoods: The Idea

Goal: want $P(\pi \mid Y = y)$

After we have data, we want to know the probability of seeing various values of π.

Two ways:

1. Bayes Theorem.

$$P(\pi \mid Y = y) \text{ is proportional to } \underbrace{P(\pi)}_{\text{prior}} \; \underbrace{P(Y = y \mid \pi)}_{\text{likelihood}}$$

   Some controversy in getting the prior:
   - How to get it, what it means, subjective disagreements, subjective biases

Thinking with Likelihoods: The Idea

2. Maximum Likelihood

   No priors; more limited, working only with the likelihood

Thinking with Likelihoods: The Idea

2. Maximum Likelihood

   No priors; more limited, working only with the likelihood

   We can conduct inferences about parameters using maximum likelihood.

Thinking with Likelihoods: The Idea

2. Maximum Likelihood

   No priors; more limited, working only with the likelihood

   We can conduct inferences about parameters using maximum likelihood.

Definition. The <u>likelihood function</u> is the probability of the observed data occurring, expressed as a function of the parameter value.

Thinking with Likelihoods: The Idea

2. Maximum Likelihood

   No priors; more limited, working only with the likelihood

   We can conduct inferences about parameters using maximum likelihood.

Definition. The <u>likelihood function</u> is the probability of the observed data occurring, expressed as a function of the parameter value.

Example.
Of the N = 2 people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = ?)$

Maximum Likelihood Example, continued.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = \boldsymbol{?})$

Maximum Likelihood Example, continued.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = \mathbf{?})$

Binomial Distribution:

$$P(Y = y \mid n=2) = \frac{2!}{y!(2-y)!} (\pi)^{y} (1 - \pi)^{2-y}$$

Maximum Likelihood Example, continued.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = \mathbf{?})$

Binomial Distribution:

$$P(Y = y \mid n=2) = \frac{2!}{y!(2-y)!} (\pi)^{y} (1 - \pi)^{2-y}$$

$$P(Y = 1 \mid n=2) = 2\pi(1 - \pi)$$

Maximum Likelihood Example, continued.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = ?)$

Binomial Distribution:

$$P(Y = y \mid n=2) = \frac{2!}{y!(2-y)!} (\pi)^{y} (1 - \pi)^{2-y}$$

$$P(Y = 1 \mid n=2) = 2\pi(1 - \pi)$$

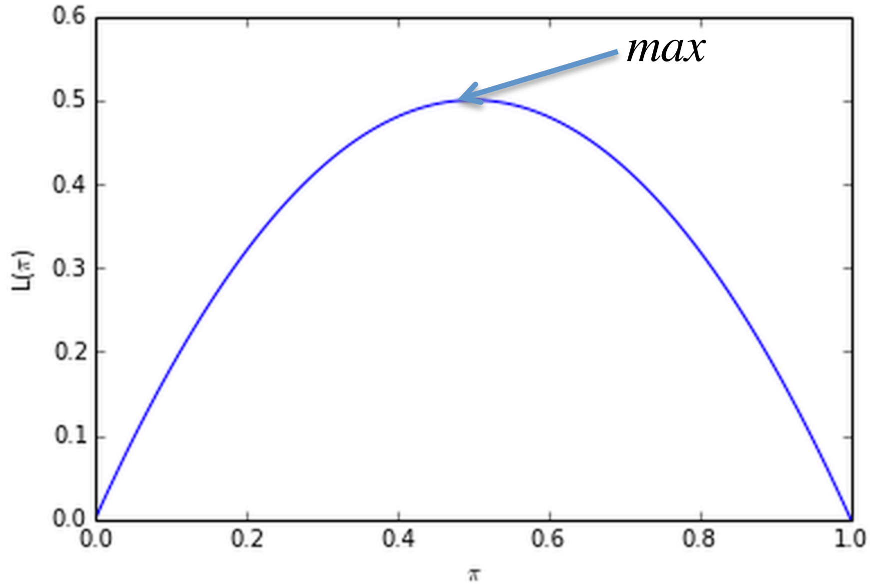
The likelihood function, below, is the probability of observing $y=1$, expressed as a function of parameter value π:

$$L(\pi) = 2\pi(1 - \pi)$$

Maximum Likelihood Example, continued.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = ?)$

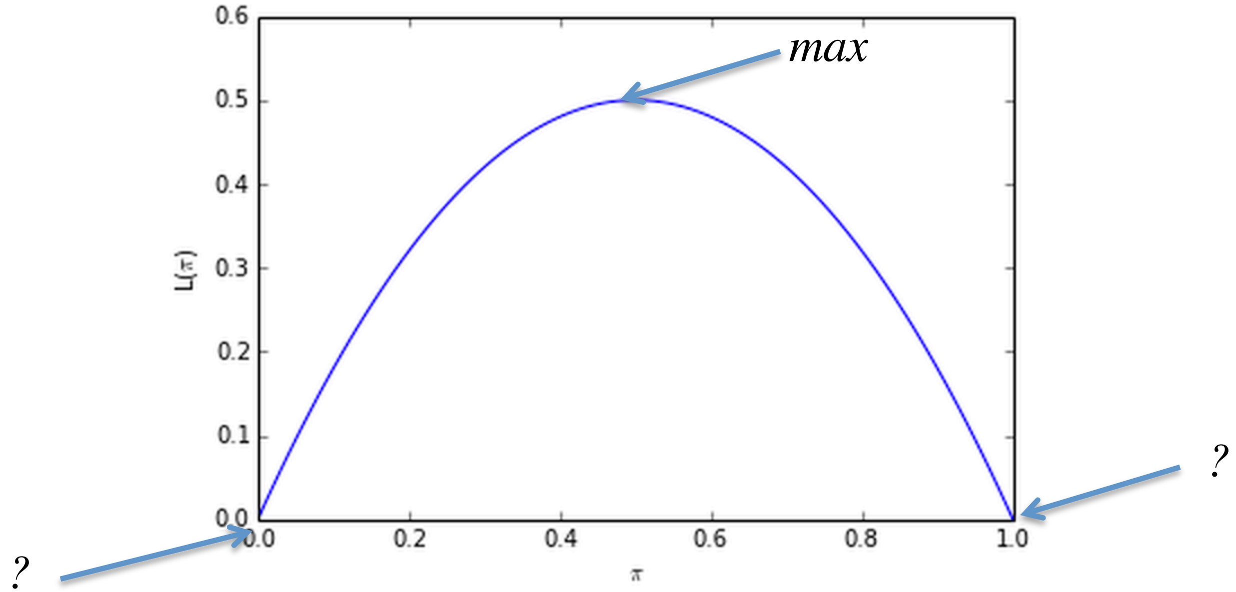
$$L(\pi) = 2\pi(1 - \pi)$$



Maximum Likelihood Example, continued.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = ?)$

$$L(\pi) = 2\pi(1 - \pi)$$



Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = ?)$

$$L(\pi) = 2\pi(1 - \pi)$$

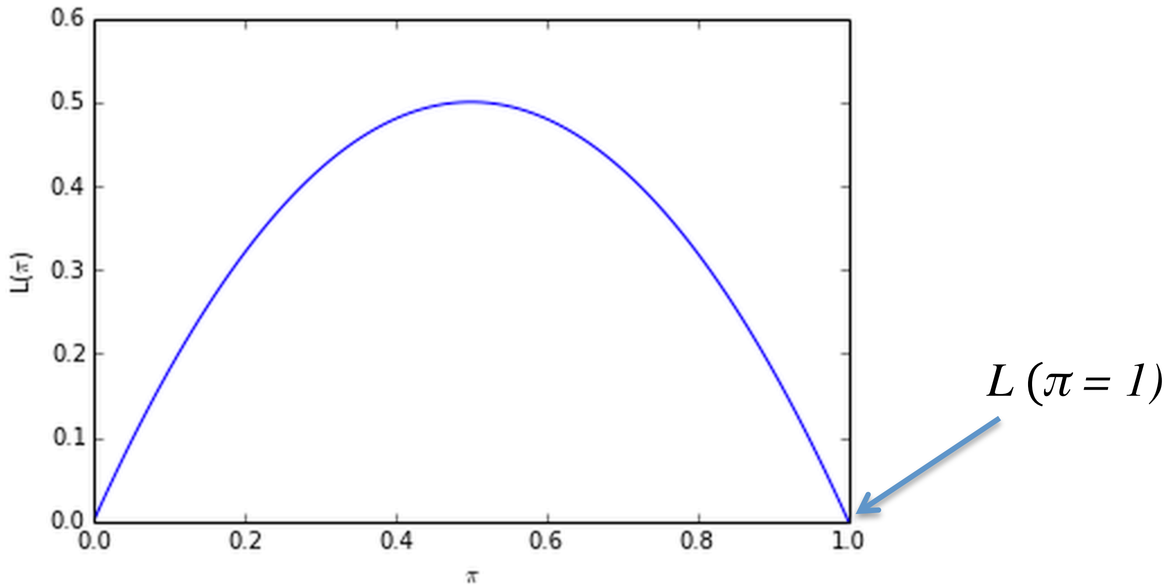
If $\pi = 1$, *the likelihood function represents the probability of observing* $y = 1$.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = ?)$

$$L(\pi) = 2\pi(1 - \pi)$$

If $\pi = 1$ *, the likelihood function represents the probability of observing* $y = 1$.

$$L\ (\pi = 1) = 2 \times 1 \times (1 - 1)$$
$$= 0$$



Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = \boldsymbol{?})$

$$L(\pi) = 2\pi(1 - \pi)$$

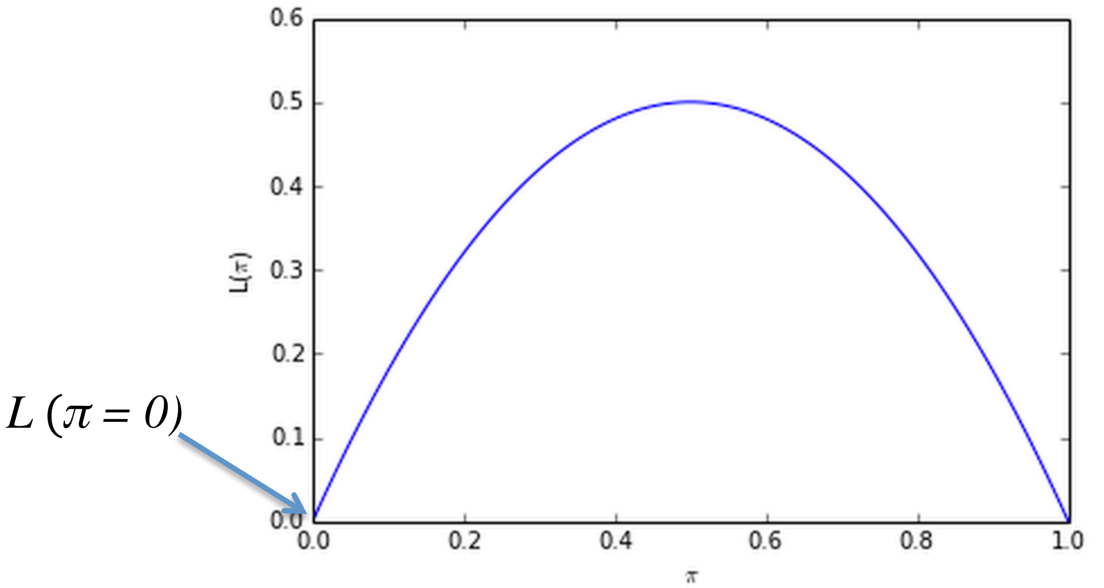
If $\pi = 0$, *the likelihood function represents the probability of observing* $y = 1$.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = \boldsymbol{?})$

$$L(\pi) = 2\pi(1 - \pi)$$

If $\pi = 0$ *, the likelihood function represents the probability of observing* $y = 1$.

$$L\ (\pi = 0) = 2 \times 0 \times (1 - 0)$$
$$= 0$$



Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = \boldsymbol{?})$

$$L(\pi) = 2\pi(1 - \pi)$$

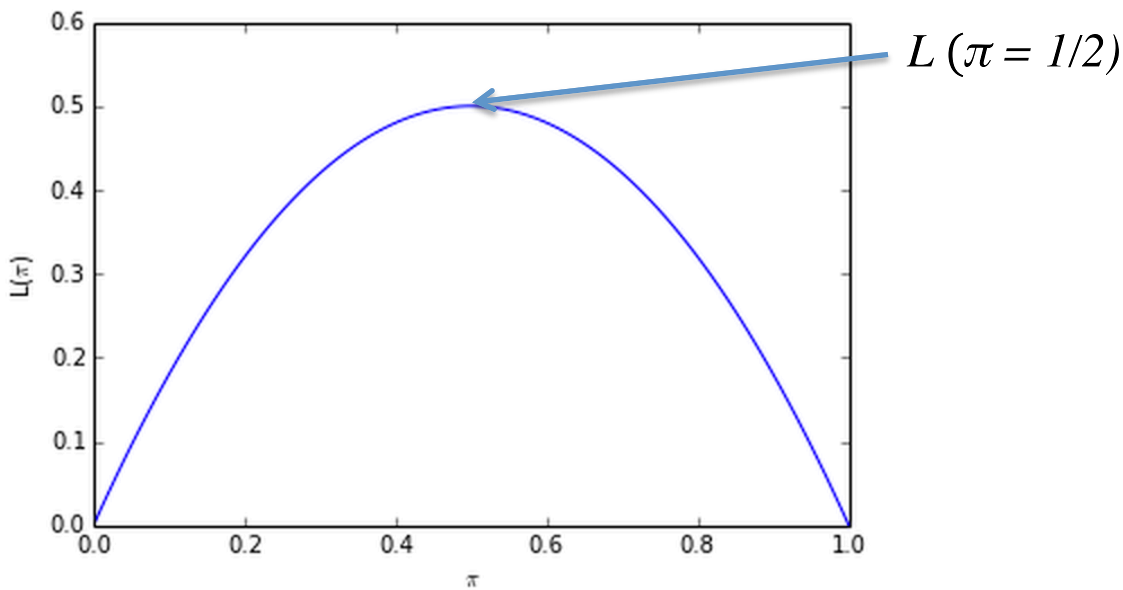
If $\pi = 1/2$, *the likelihood function represents the probability of observing* $y = 1$.

Of the $n = 2$ people on Wall Street, 1 was a woman.
Then $Y = \#\ successes = 1$.
$Y \sim Binomial\ (n = 2, \pi = \boldsymbol{?})$

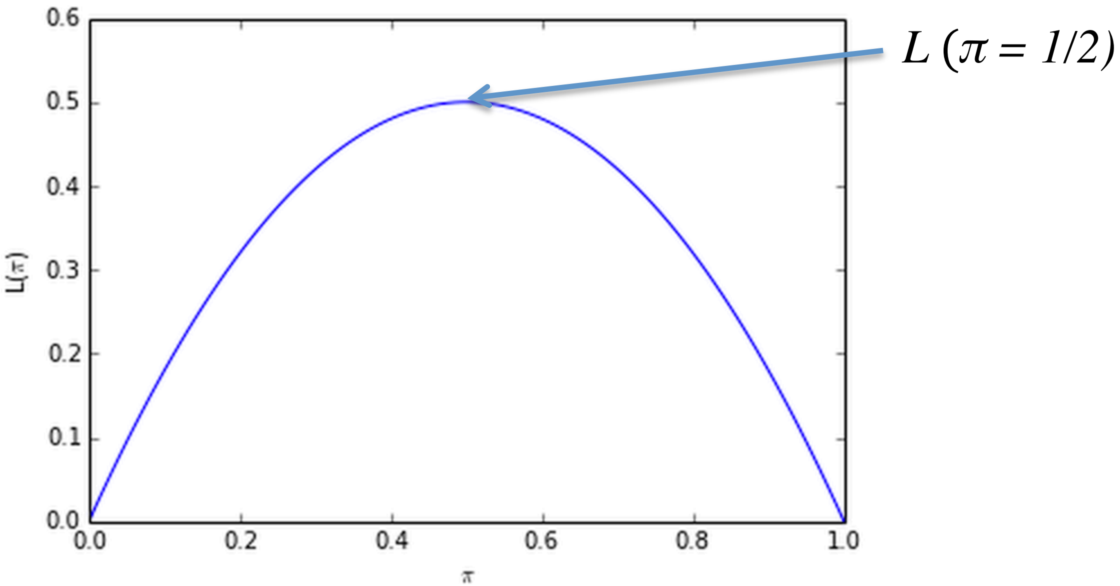
$$L(\pi) = 2\pi(1 - \pi)$$

If $\pi = 1/2$ *, the likelihood function represents the probability of observing* $y = 1$.

$$L\ (\pi = 1/2) = 2 \times 1/2 \times (1 - 1/2)$$
$$= 1/2$$



This is the maximum likelihood estimate.



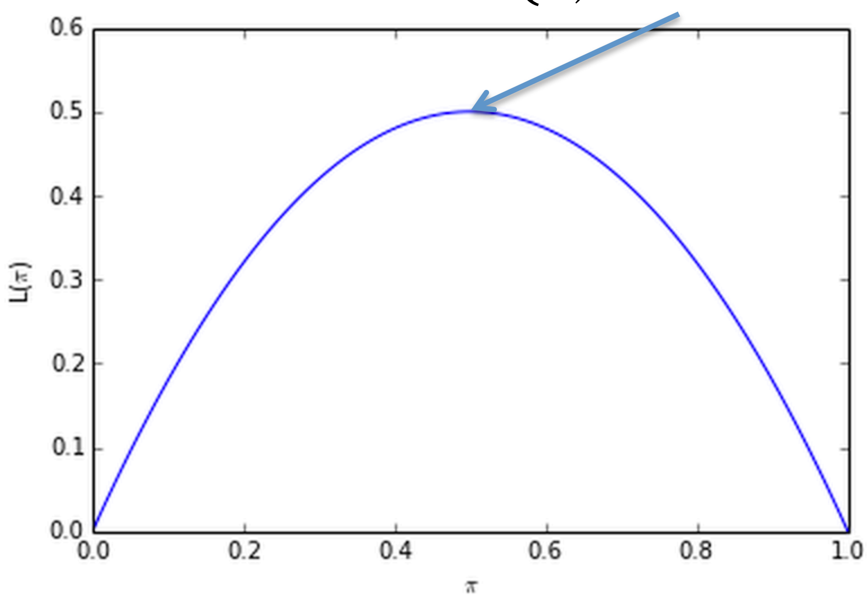
Definition. The Maximum Likelihood (ML) estimate of a parameter π is the parameter value at which the likelihood function takes its maximum.

Intuitively, the Maximum Likelihood estimate $\hat{\pi}$ is your guess for the "state of nature" that best explains the data. It is the estimate for which the data is *most likely* to occur.

Definition. The Maximum Likelihood (ML) estimate of a parameter π is the parameter value at which the likelihood function takes its maximum.

Intuitively, the Maximum Likelihood estimate $\hat{\pi}$ is your guess for the "state of nature" that best explains the data. It is the estimate for which the data is *most likely* to occur.

Example: $L(\pi) = 2\pi(1 - \pi)$ is maximized at $\pi = ½$.
Thus the Maximum Likelihood estimate for $L(\pi)$ is $\hat{\pi} = ½$.



Notes.

For Binomial random variable Y: $\hat{\pi} = y/n$

Mathematically, in Ordinary Least Squares ($Y|X \sim Normal$), the least squares estimates for the β parameters are the Maximum Likelihood estimates.