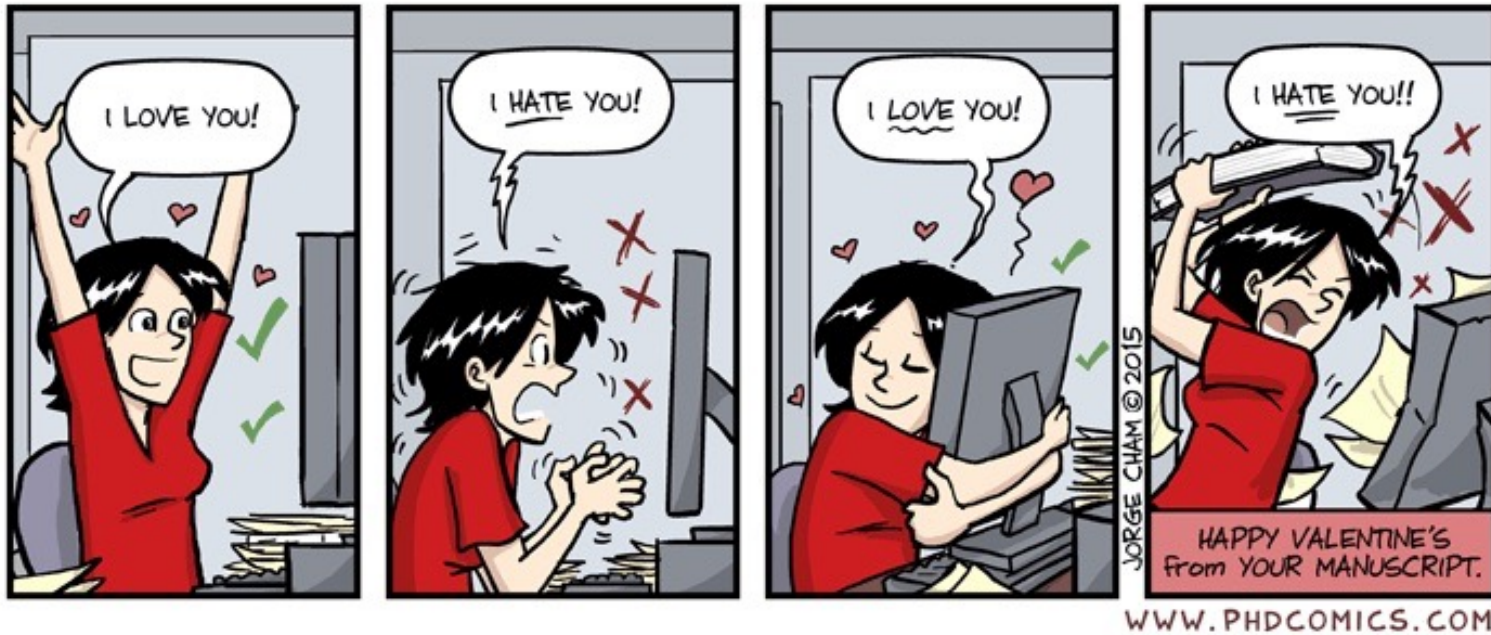


BIOL 501: Likelihood



Peer-Feedback Survey

Please get me out! It helps us all learn names

FIRST NAME
Preferred pronouns

Outline

- **Assignment #2 due 18 March**
- Probability & Distributions
- Likelihood
- Probability vs. Likelihood
- Example with wasps and binomial distribution
- Maximum likelihood estimation (MLE)
- Likelihood-based confidence intervals
- Workshop Prep

Assignment #2: Due 18 March @9pm

- 4.5 weeks, but don't wait until the last minute
- Due after model selection intentionally
- **Self-assessment with rubric *prior* to turning it in**

1. Get your dataset soon

2. Scatter plot to examine

3. Decide if it is fixed only, mixed, or generalized

Assignment #2: Data set requirements

- **Linear, mixed, or generalized linear model in R.** You can choose which type of **linear** model to use, as long as it is appropriate for your dataset.
 - Fixed → Lm, glm, gam
 - Mixed → LME, GLMM, GAMM
- Only 1 response variable
- At least 1 categorical factor
- Include at least 1, and no more than 2, additional explanatory variables

How to find data for Assignment #2

- Ask labmates, supervisor, other grad students for data
- Extract data that was used for something else other than a linear model
 - “How to find data for practicing R.docx” / Assignment #2 on Canvas
- **Not** ok to redo a figure that’s already published and analyzed as a linear model, but you can extract data from published paper that was used for something else
- **Don’t** use simulated data
- Yes it’s ok, if it’s your own thesis data, undergrad data

Discussion: Probability vs. Likelihood

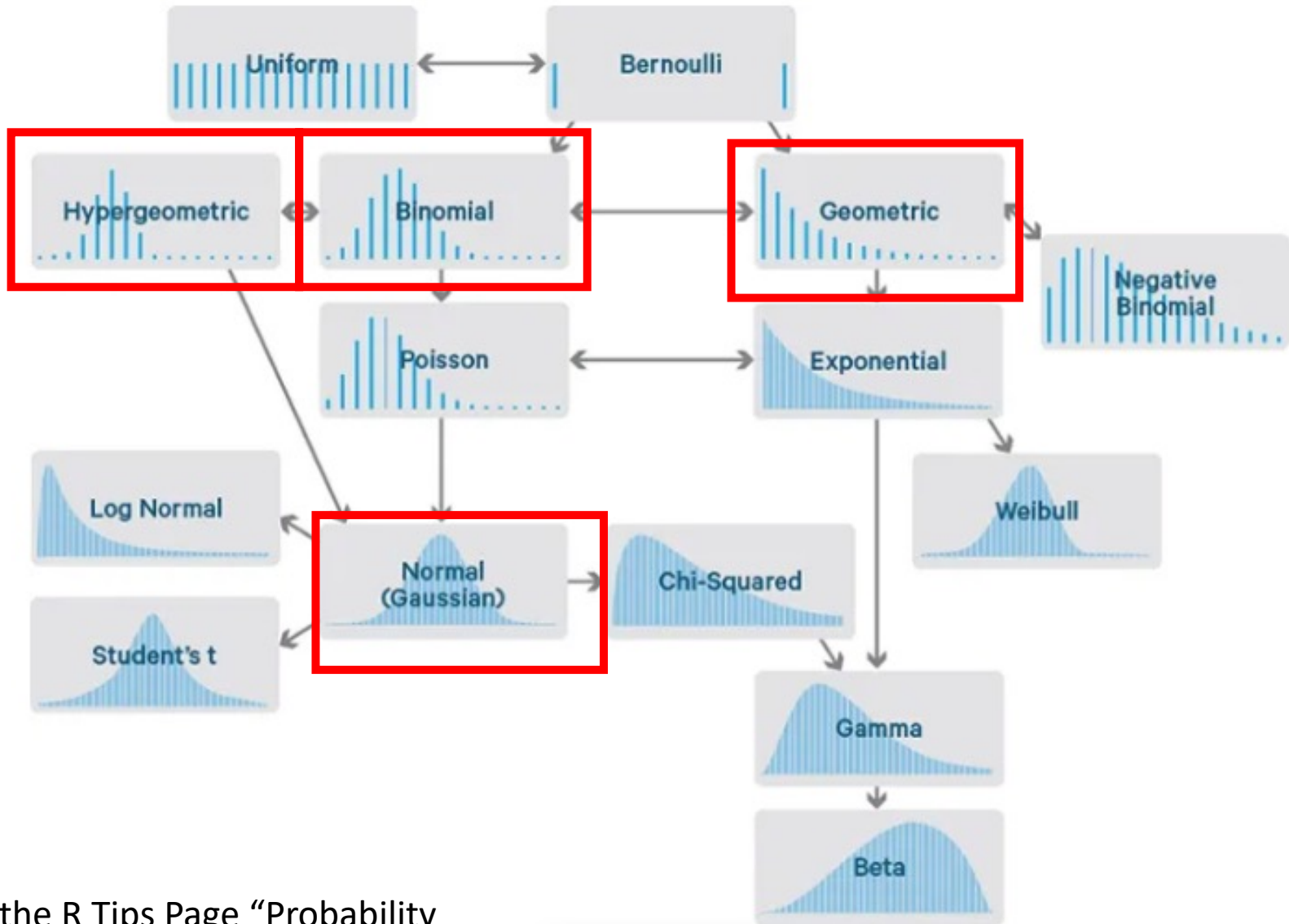
- In groups of 2-3 discuss the difference in the definitions of these terms in popular culture vs statistics

Probability & Distributions

Probability

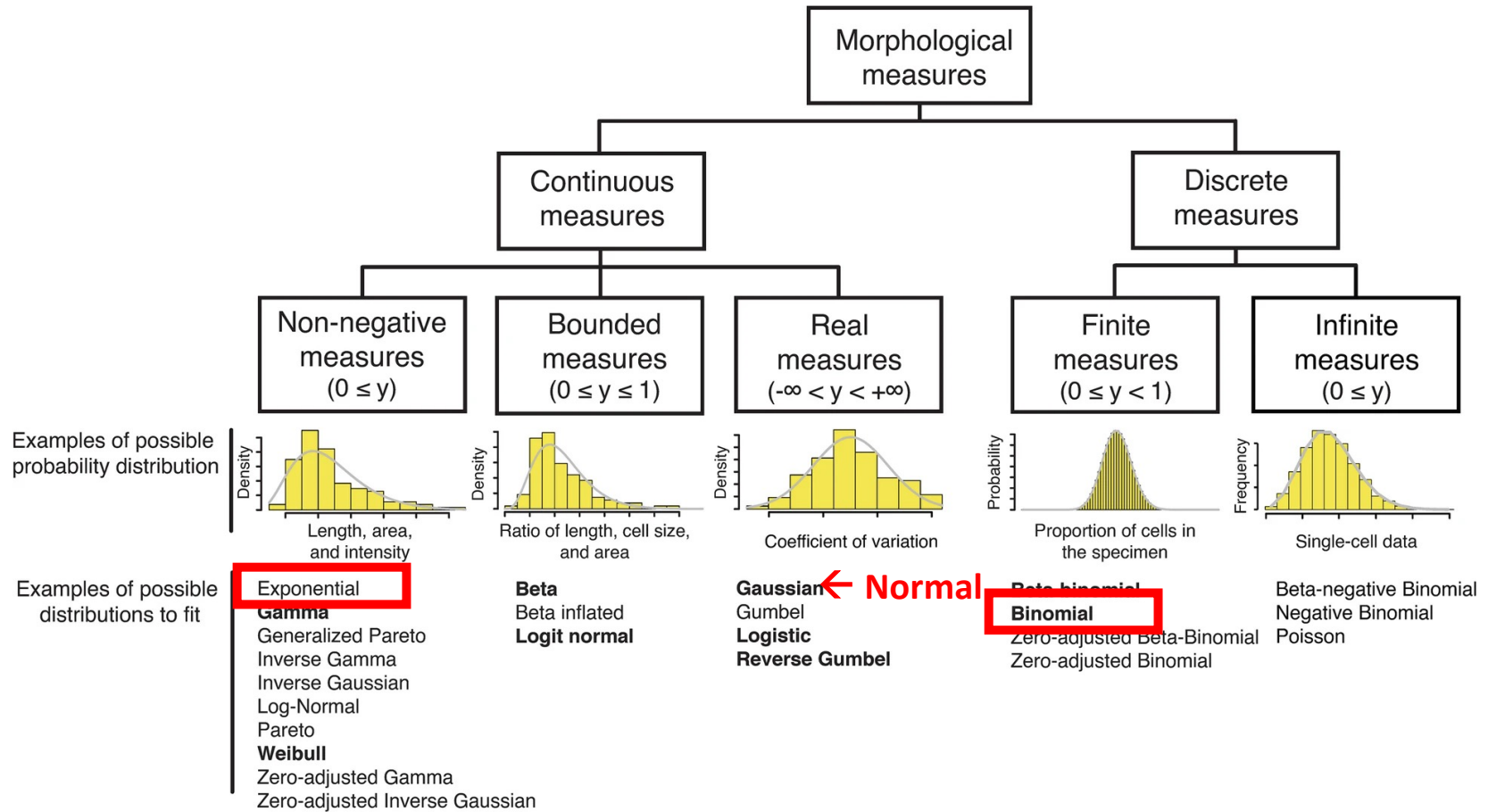
- **Probability:** Proportion of times that event would occur if a random trial is repeated over and over again under the same conditions
- **Probability distribution:** list of all mutually exclusive outcomes of a random trial and their probabilities of occurrence.

Examples of Probability Distributions



See the R Tips Page “Probability and Likelihood”

Lots of probability distributions models exist for different situations



Why care about probability distributions?

- You have to select the right probability distribution that fits your data and helps you explain what you're trying to understand.
- When creating models, change the distribution in R defaults to fit your data—otherwise many packages assume normal distribution which may or may not be true
- For example, GLMs GAMs later on

Discrete probability distributions

- It depends on the properties of your data
- Here are a few common ones, but there are many
- **Binomial:** binary outcome
- **Geometric:** binary outcome, and probability that a specified number of trials will take place before first success occurs
- **Hypergeometric:** probability of k successes in n draws without replacement (mark-recapture)
- Poisson: count data such as animal litter sizes; approaches normal as parameter increase

***Bold** indicates example used in workshop this week

Continuous probability distributions

- Here are a few common ones, but there are many
- Normal
- **Exponential**

***Bold** indicates example used in workshop this week

Likelihood

What is a conditional probability?

- Probability of an event occurring given that a condition is met.
- “Given” is represented by “|”
- **Example:** Probability that 2nd child born to a couple is a girl, given that 1st child was a girl
- $\text{Pr}[\text{second child is girl} \mid \text{first child is girl}]$

More examples of conditional probabilities

- Pr="probability"
- Condition that needs to be met "Given" = "|"
- Pr[we see an elephant today | we are in the Serengeti]
- Pr[we see an elephant today | we are in Manhattan]
- Pr[12 successes in 18 trials | $p = 0.50$]
- Pr[12 successes in 18 trials | $p = 0.10$]

Likelihood is a conditional probability

Determines if we can trust the parameters in a model based on the sample data that we've observed.

- **Law of likelihood:** extent to which data supports one parameter value or hypothesis against another is equal to the ratio of their likelihoods (difference in their log-likelihoods)
- Method invented by R. A. Fisher when a 3rd-year undergraduate.

Likelihood rephrased by a non-statistician

- Likelihood describes **the extent to which the sample provides support for any parameter value**. Higher support corresponds to a higher value of likelihood.
- We're considering the set of observations as fixed—they've happened in the past—and now we're considering under which set of model parameters we would be most likely to observe them.
- The “parameter” is the parameter of the probability distribution assumed to be underlying the data

Likelihood is a conditional probability

- Likelihood of a population parameter equaling a specific value, given the data, is the probability of obtaining the observed data *given (i.e the conditional part)* that the population parameter equals the specific value.
- Would be annotated like this

$$L[\text{parameter} = \rho \mid \text{data}] = \Pr[\text{data} \mid \text{parameter} = \rho]$$

Probability vs Likelihood

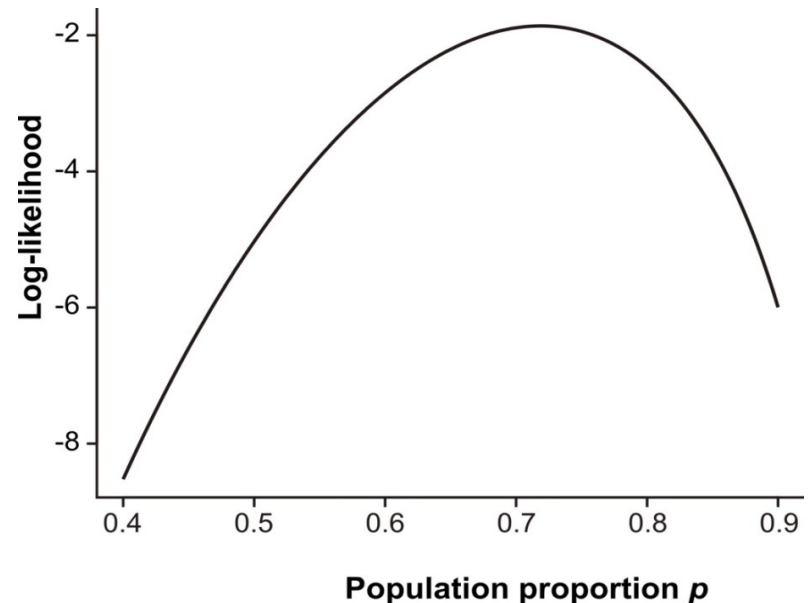
Probability vs likelihood

This is an important slide and common area of confusion

- **Probability** refers to chance that a particular outcome occurs based on the values of parameters in a model.
 - Assumes parameters are trustworthy.
- **Likelihood** refers to **how well a sample provides support for particular values of a parameter** in a model.
 - Determines if we can trust the parameters in a model based on the sample data that we've observed.
 - **Likelihood works backwards from probability**

Probability vs likelihood

- Use likelihood to estimate unknown parameters based on *known data*
- Parameters are treated as variables, the data are a constant, unvarying.
- But the likelihood function is not a probability distribution.
- The **population proportion p is the variable of the function**, but it is not a random variable (its value is not determined by random trial).



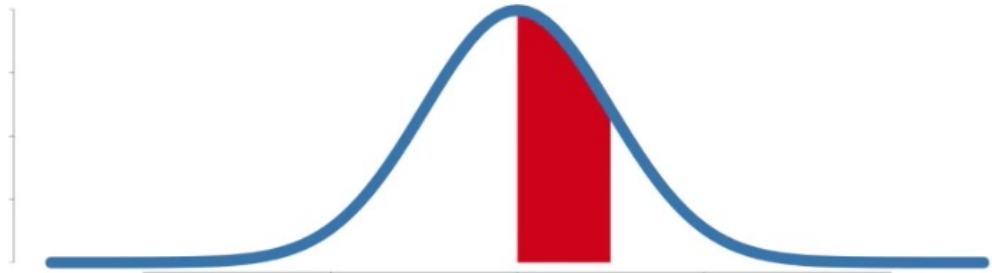
Likelihood vs. Probability related but different

<https://www.youtube.com/watch?v=pYxNSUDSFH4>

In summary...

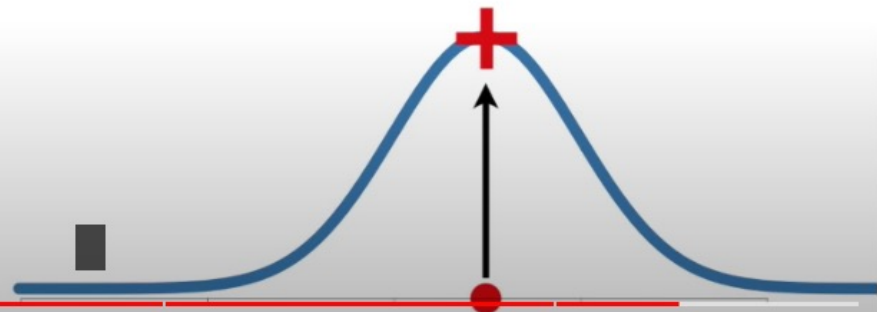
Probabilities are the areas
under a fixed distribution...

$$pr(\text{data} \mid \text{distribution})$$



Likelihoods are the y-axis
values for fixed data points
with distributions that can be
moved...

$$L(\text{distribution} \mid \text{data})$$



Coin Toss Example:binomial

- Flip coin 1 time, **probability** heads is 0.5 on a "fair" coin.
- Flip 100 times yields heads 17 times.
- Calculate likelihood in R with binomial distribution

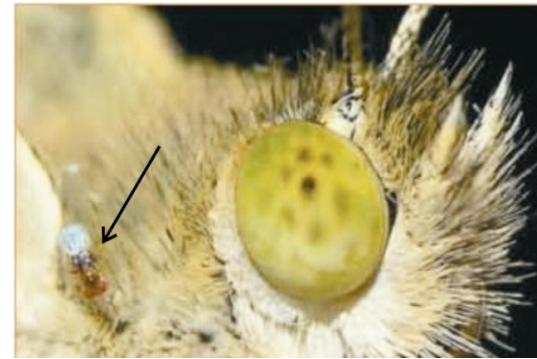
```
#If 100 coin flips, what is likelihood 17 are heads, assuming probability of heads  
on fair coin is 0.5?  
#dbinom(heads.observed, size=trials, p=assumed probability value)  
dbinom(17, size=100, p=0.50)  
##0.000000000005246031
```

- **Likelihood** that the coin is "fair" (or that p truly =0.5) quite low.
- If the coin was actually fair, we would expect it to land on heads much more often.

Coin Toss Example

- **Probability** assumes that $P(\text{heads}) = 0.5$ on a given toss if a fair coin
- **Likelihood**: trying to determine if the model parameter ($p = 0.5$) is actually correctly specified.
- Coin landing on heads only 17/100 times highly suspicious that true probability of the coin landing on heads on a given toss is actually **not** $p = 0.5$.

Example: Estimate a binomial proportion p with Wasps



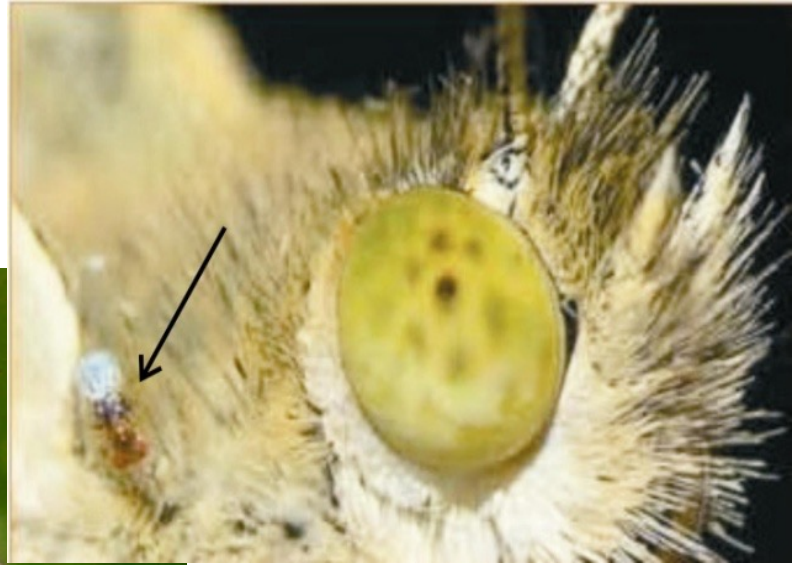
- **Data:** The tiny wasp rides on female butterflies. When a butterfly lays her eggs, the **wasp parasitizes the freshly laid butterfly eggs**.
- Can wasps distinguish mated female butterflies from unmated females? (pointless to ride on a female that already laid eggs).
- **Trials:** Wasp presented with 2 female butterflies to ride on (mated vs. unmated butterflies)
- **Observed:** 23 of the 32 wasps chose the mated female
- **Probability** of the wasp riding on a mated caterpillar in each trial was 0.50

What is the proportion of wasps in the population choosing the mated female caterpillar

Success=chose female mated caterpillar

Failure=wasp chose unmated female caterpillar

We will use this case study again in GLMS



Optional to follow-along coding wasp example

- R script posted on Canvas
- Rough outline of the steps
 1. Examine data. Which probability distribution do I use?
 2. Calculate the log-likelihood for many values of P to get the log-likelihood curve
 3. Plot the log-likelihood curve
 4. Calculate the maximum log-likelihood (where does the peak of the curve intersect y axis)
 5. Calculate the Maximum Likelihood Estimate (MLE) of the parameter (vertical line that intersects peak)
 6. Likelihood-based 95% CI

Binomial Distribution

Example: Binomial Distribution

- **Probability distribution of number of “successes” in n independent trials, when the probability of success p is the same in each trial.**
- **Binary data → only 2 outcome options**
 - 1 or 0; success or failure; alive or dead; mated or unmated, present or absent
- `dbinom()` command in R for binomial data
 - density function for a given binomial distribution

```
#Probability of obtaining  $X$  successes in  $n$  random trials
```

```
# $X$  is an integer between 0 and  $n$  ( $n$  can be 1)
```

```
 $X$ =number of successes
```

```
 $n$ =number of random trials
```

```
 $p$ =probability of success in each trial
```

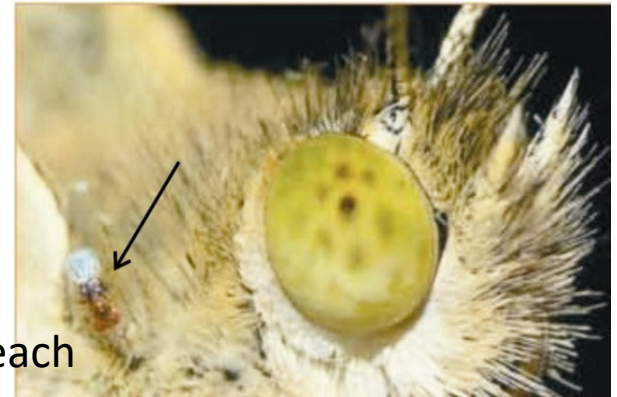
```
#Use the dbinom command below
```

```
dbinom( $X$ ,size= $n$ , prob= $p$ )
```

**This is generic example
of `dbinom` R code()**

Example: Estimate a binomial proportion p with Wasps

- Observed: 23 of the 32 wasps chose the mated female
- Probability of the wasp riding on a mated caterpillar in each trial was 0.50



What is the proportion of wasps in population choosing the mated female (i.e. success)?

```
x<-23  #number of "successes" where wasp chose mated female
n=32   #number of random trials
p=0.50 #probability that the caterpillar was mated (success)

#Use the Likelihood binomial function in R
#dbinom(X,size=n,prob=p)
dbinom(x, size=n, p=p) |
#[1] 0.0065306202

#easier to work with log-likelihoods
#this calculates the probability of ONLY 1 p when p=0.05 (which is the null 50/50)
dbinom(x, size=n, p=p,log=TRUE)
##[1] -5.031253
#this is exactly the same as below
dbinom(23, size=32, p=0.50,log=TRUE)
```

- Easier to work with log-likelihoods
- Set **dbinom(...log=TRUE)**

Why use log-likelihood instead of likelihood?

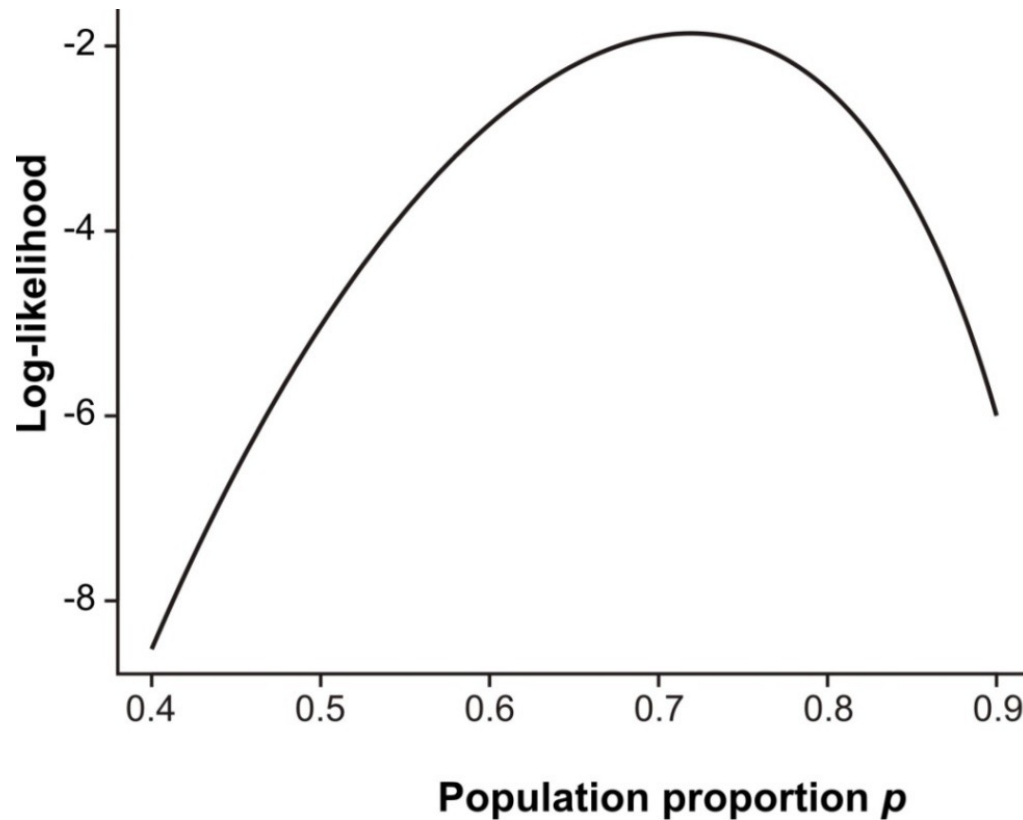
- **Short answer:** Both to the same job, but log-likelihood is computationally simpler and easier to interpret and graph

If you want to know more:

- Likelihoods are often tiny numbers (or large products) which makes them difficult to graph. Taking the natural (base e) logarithm results in a better graph with large sums instead of products.
- The log likelihood function is usually (not always!) easier to optimize.
- Natural logarithm is a monotonically increasing function. This means that if the value on the x-axis increases, the value on the y-axis also increases (see figure below). This is important because it ensures that the maximum value of the log of the probability occurs at the same point as the original probability function. Therefore we can work with the simpler log-likelihood instead of the original likelihood

Log-Likelihood Curve

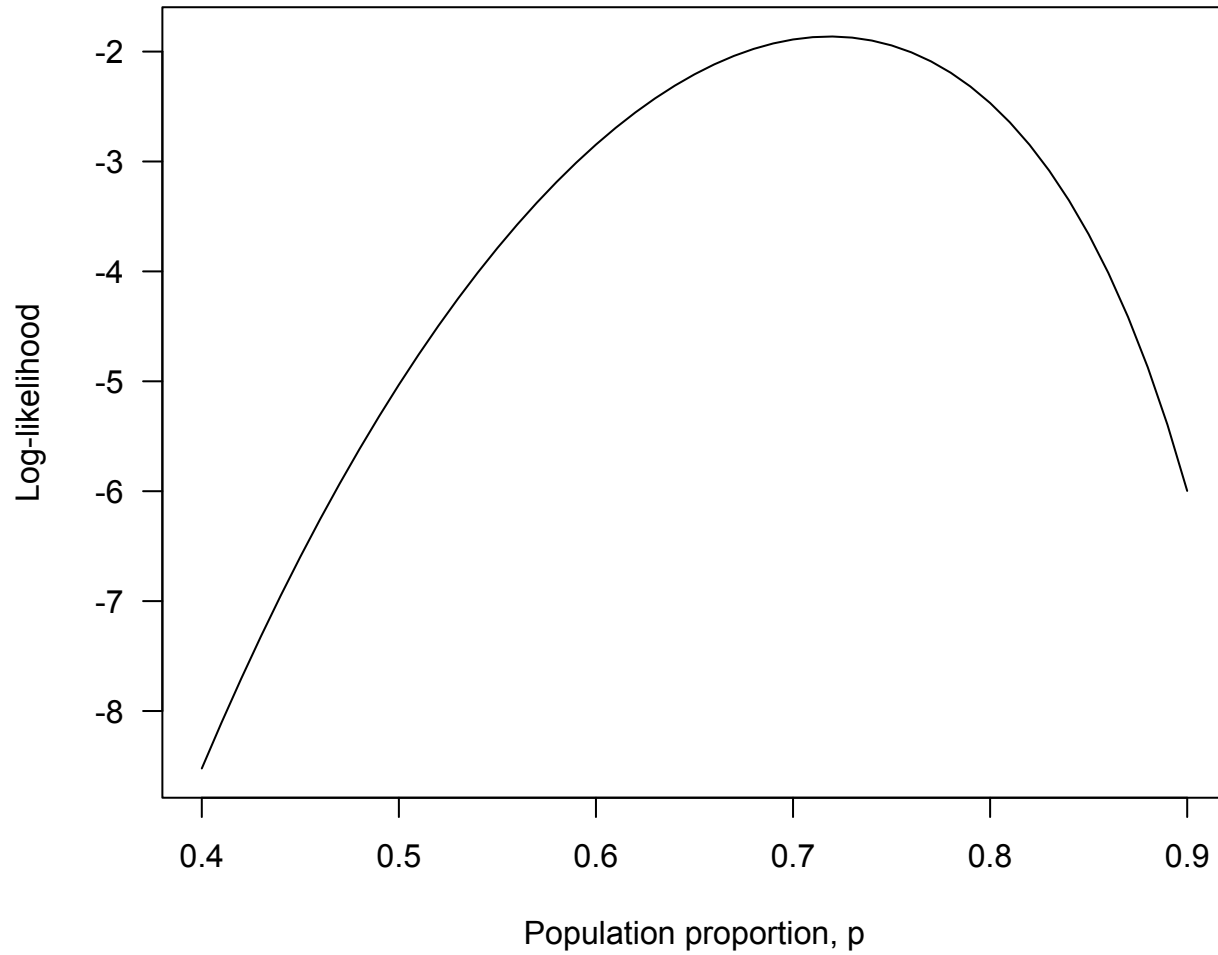
- Calculate the log-likelihood for many values of p , to get the **log-likelihood curve**



Go to Wasp R script

- Calculate the log-likelihood for many values of P to get the log-likelihood curve
 - Proportions from 0.1 to 0.9 by 0.01
- Calculate the for each value of the proportions
- Plot log-likelihood curve the **stop** before MLE

Wasp Log-likelihood Curve



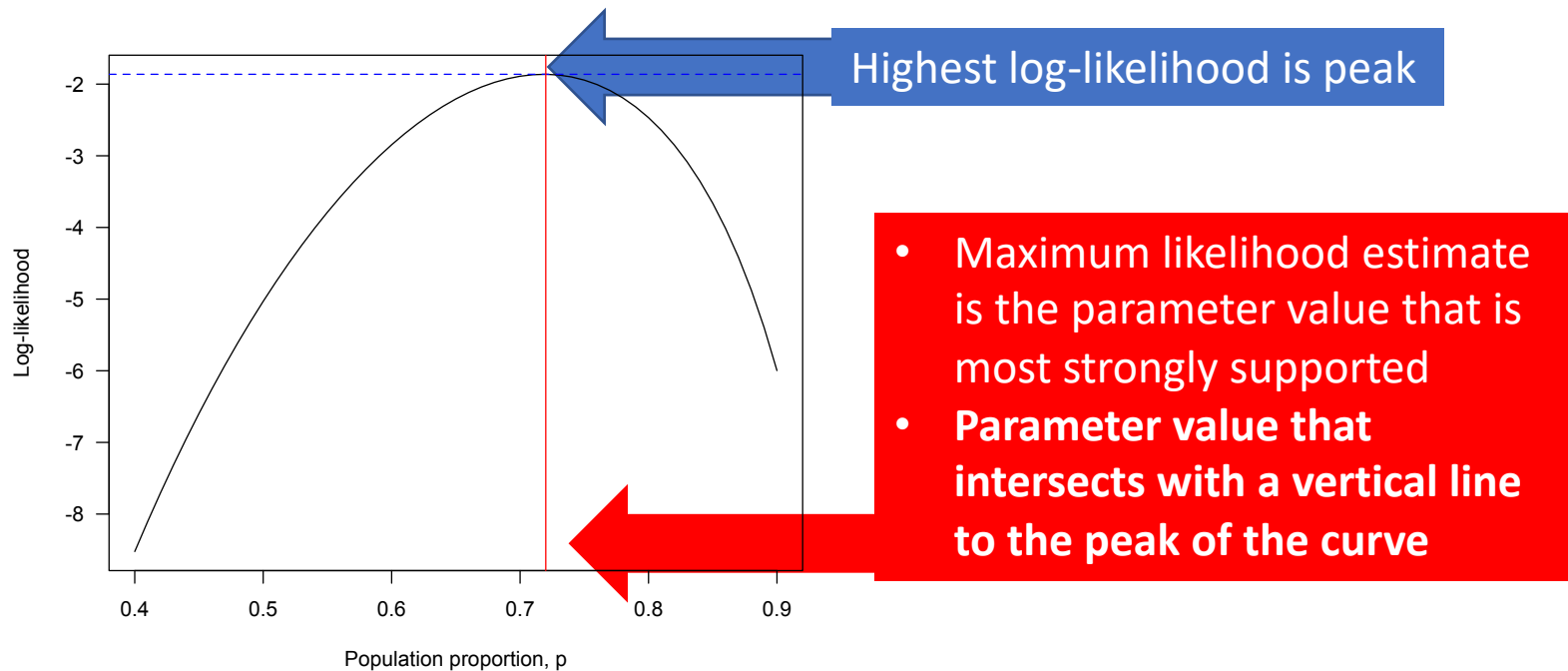
Maximum Likelihood
Estimate
MLE

Maximum Likelihood estimate

- Maximum likelihood estimation is a method that determines values for the parameters of a model. The parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were actually observed.

Maximum Likelihood Estimate

- **Likelihood ratio** (difference of log-likelihoods) measures relative support for alternative parameter values
- **Maximum likelihood estimate (MLE)**: of a parameter is the parameter value having the highest likelihood (and log-likelihood), given the data
 - MLE is the parameter value **most strongly supported** by the data

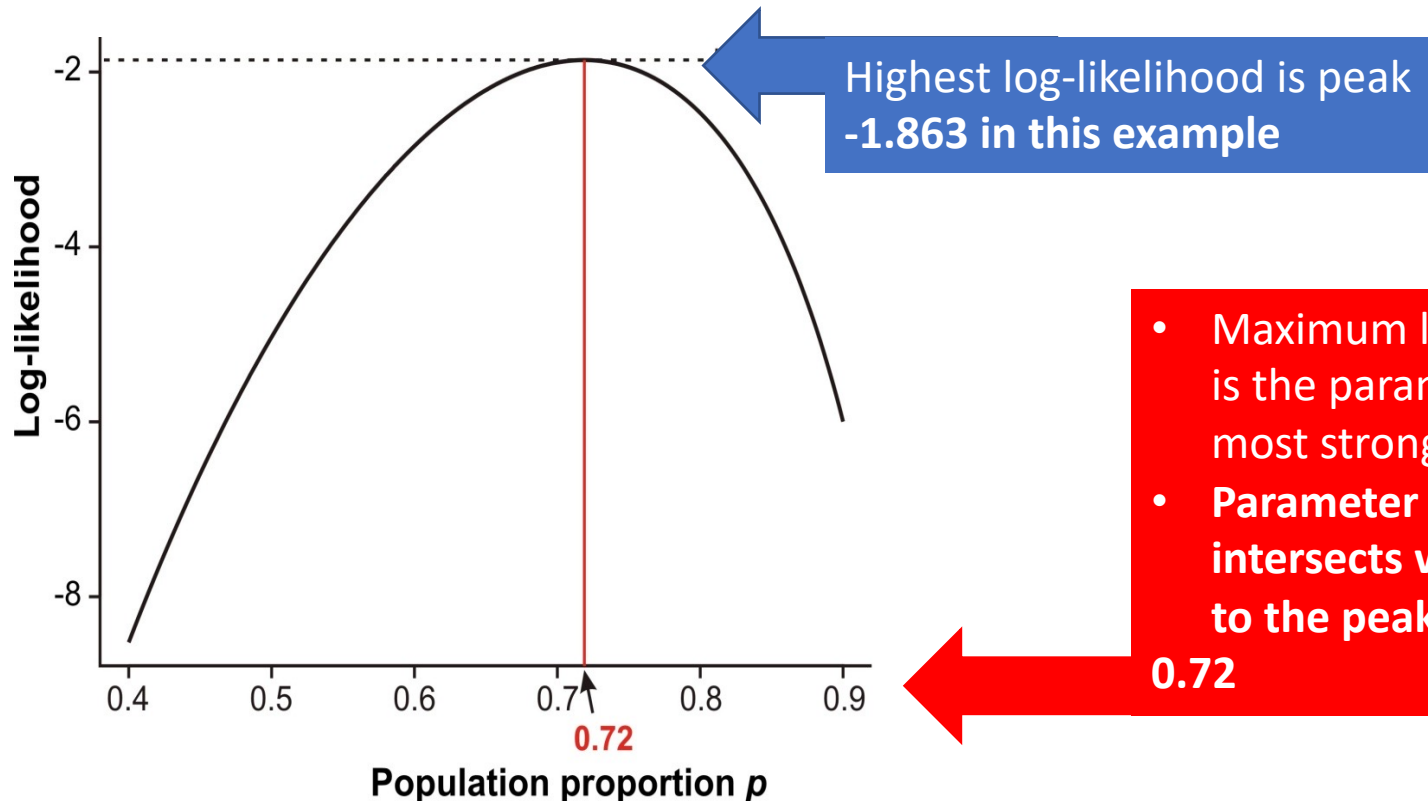


Go to Wasp R script

- Calculate the maximum log-likelihood
 - Where does the peak of the curve intersect y axis?
- Calculate the Maximum Likelihood Estimate (MLE) of the parameter
 - vertical line that intersects peak

Maximum Likelihood Estimate

- **Likelihood ratio** (difference of log-likelihoods) measures relative support for alternative parameter values
- **Maximum likelihood estimate (MLE)**: of a parameter is the parameter value having the highest likelihood (and log-likelihood), given the data
 - MLE is the parameter value **most strongly supported** by the data



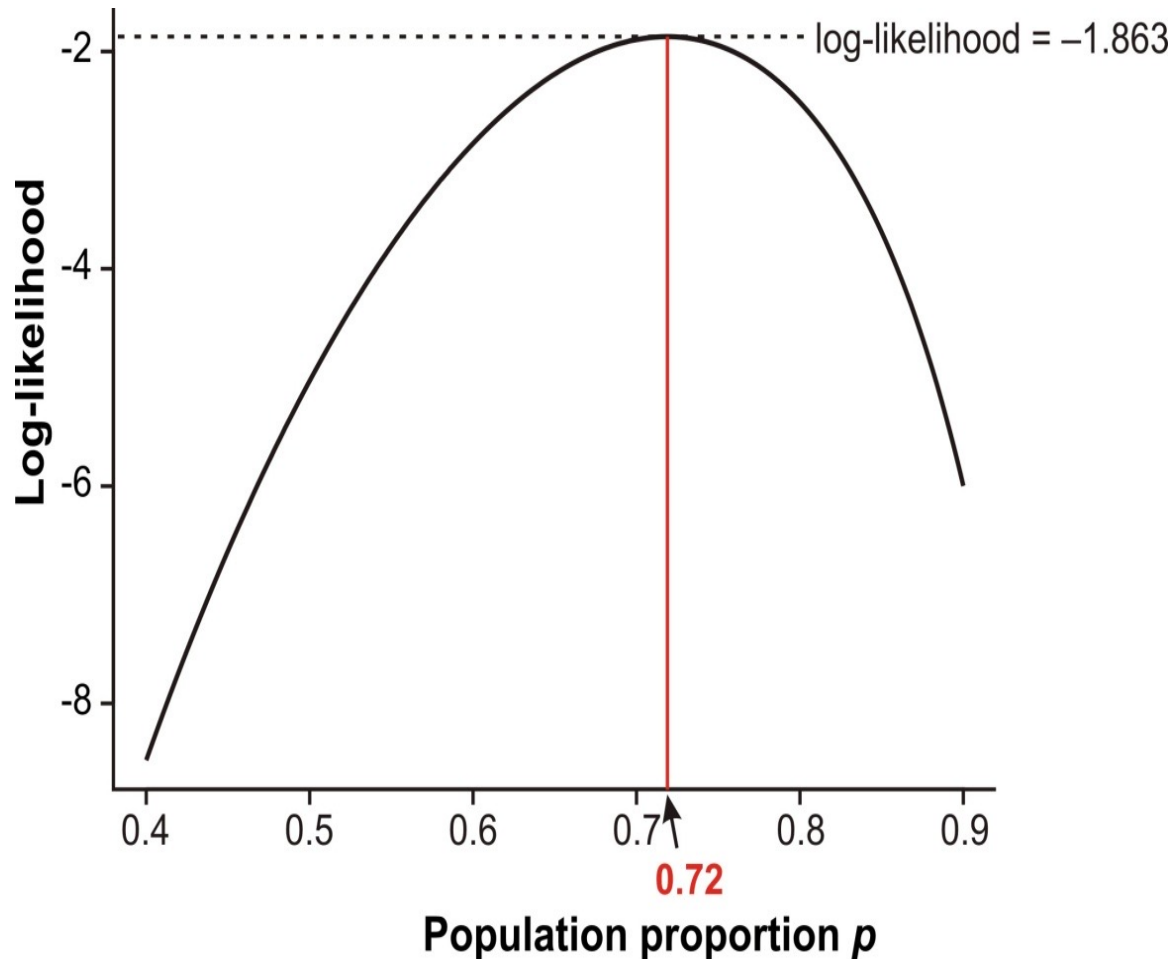
- Maximum likelihood estimate is the parameter value that is most strongly supported
- **Parameter value that intersects with a vertical line to the peak of the curve**

0.72

What does it all mean in the end?

Question: What is the proportion p of wasps in the population choosing the mated female caterpillar?

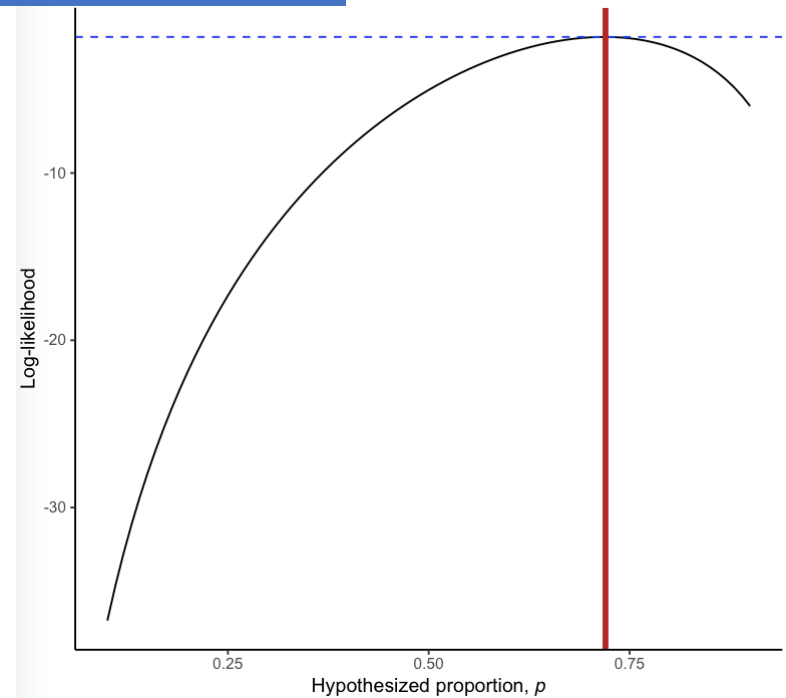
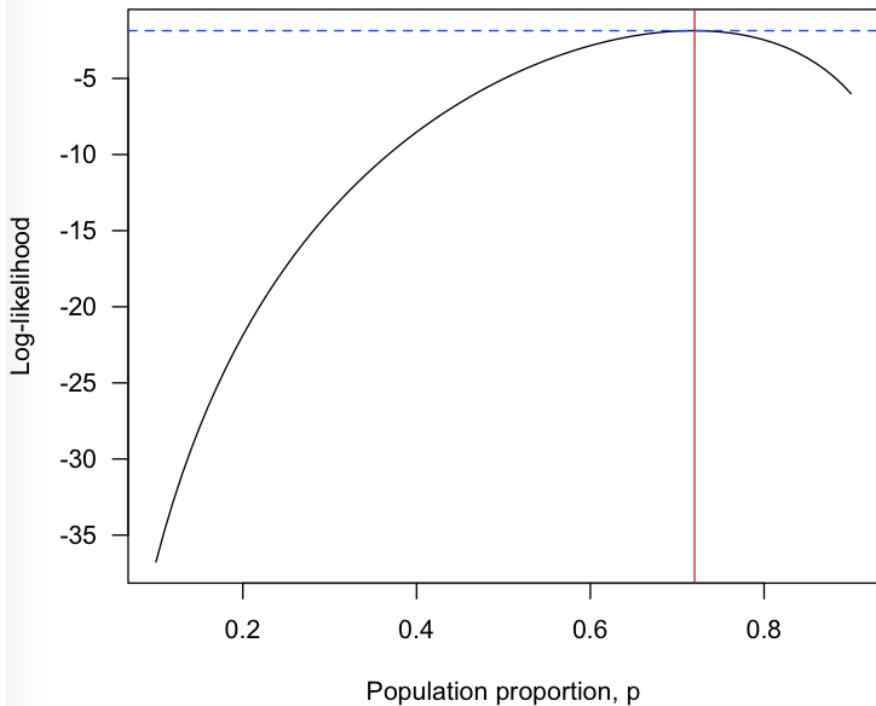
Answer: 72% of the wasps in the population choose a mated female caterpillar (the MLE estimate)



Base R

ggplot

Highest log-likelihood is peak
-1.863 in this example

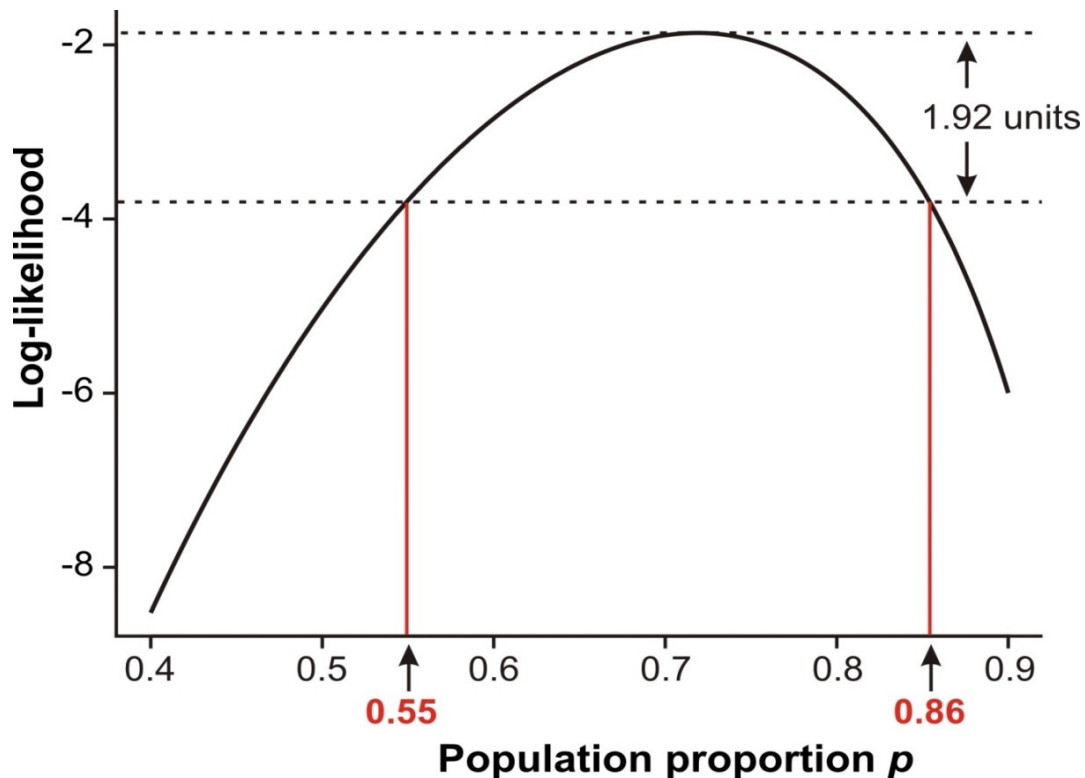


- Maximum likelihood estimate is the parameter value that is most strongly supported (**0.72**)

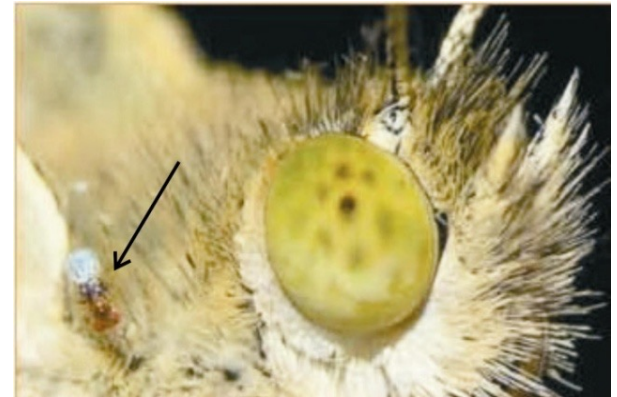
Likelihood –based
Confidence intervals

Likelihood-based Confidence Intervals

- Approximate 95% confidence interval is obtained with the values corresponding to 1.92 log-likelihood units below the maximum



Wasp example: Calculate 95% CI



- Approximate 95% CI is $0.56 \leq \text{proportion} \leq 0.85$

```
#Likelihood-based 95% CI for parameter  
# 1.92-unit support limits.  
# This method gives an interval slightly narrower than the real values  
#remember our MLE was calculated as max(loglike) #-1.862828  
range(proportion[loglike >= (max(loglike) - 1.92)])  
## [1] 0.56 0.85
```

Why care about maximum likelihood estimation?

- You don't need to be a mathematician to use likelihood in your data analysis.
- You just need to find a formula for the probability distribution of outcomes for your particular situation.

Wasp Example Finale!

- The chemical that the wasps use to distinguish mated from unmated female caterpillars is benzyl cyanide, which the male butterfly passes to the female during mating.
- The compound is an “anti-aphrodisiac”, rendering the mated female less attractive to other male butterflies (Fatouros *et al.* 2005).



Workshop

Workshop Thurs: Likelihood

I'd recommend **starting with binomial** then choose whichever distribution may be similar to your data

- Binomial → `dbinom()`
 - Left-handed flowers
- Hypergeometric → `dhyper()`
 - Counting elephants (mark-recapture)
- Geometric → `dgeom()`
 - Voyaging Vole Dispersal
- Exponential → `dexp()`
 - Life of Bees