Likelihood

Likelihood

- Likelihood is a general approach to statistics that can be used for:
  - Estimating parameters
  - Building confidence intervals
  - Testing hypotheses
  - Comparing hypotheses against one another
- Likelihood appears to be similar to probability, but has a very different interpretation

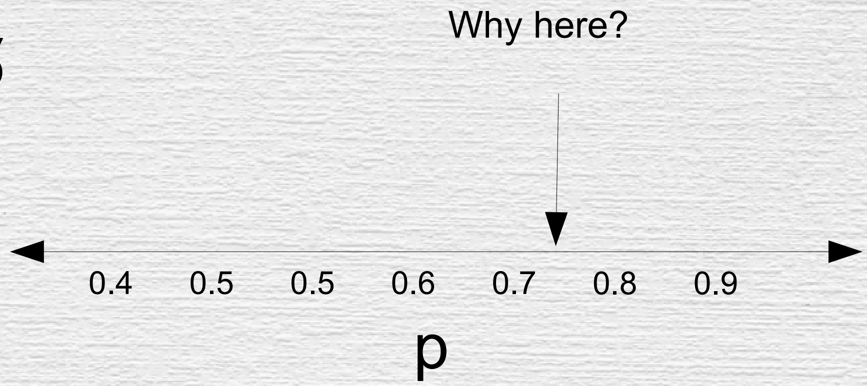
Maximum likelihood estimation

- The value of a parameter that has the greatest chance of having produced the data is the ***maximum likelihood estimate***
- Example: parasitic wasps, *Trichogramma brassicae*
  - Lay eggs on butterfly eggs
  - Ride on legs of butterflies
  - Can they tell mated butterflies from virgins?
  - Present them with mated and virgin butterflies simultaneously, see which they climb on to
  - Result: 23 out of 32 chose mated butterflies
  - What's the best estimate for the probability (p) that a wasp will choose a mated butterfly?

Selecting a best value for p

- We can use the estimator:

$$\hat{p} = \frac{successes}{trials} = \frac{23}{32} = 0.71875$$



- But, how do we know this is the best estimate for p?
  Why not 0.72, or 0.71?
- We can use maximum likelihood – but first we need to pick a likelihood function
- Likelihood functions are derived from probability distributions – we need a good probability distribution for these data

The binomial probability distribution

- Describes probabilities of X successes out of n trials, when the probability of success on a single trial is p
- The binomial distribution allows us to ask “What's the probability of getting 23 successes out of 32 trials if p is equal to 0.71875?”
  - Symbolically: P(23 out of 32 | p)
- Conditional = the term on the right of “ | ” is considered known, or “given”
- When we use this as a probability distribution, the value of p is treated as known, and the data are treated as subject to random variation

The binomial distribution

$$P\left(23\, out\, of\ 32 \middle| p\right) = \binom{32}{23} \underbrace{p^{23}\left(1-p\right)^{(32-23)}}$$

The "counting" part

i.e. the number of different ways to get 23 "successes" out of 32 "trials"

The "and probability" part

The “and probability” part

MMMMMMMMMMMMMMMMMMMMMMMVVVVVVVVV

Probability of selecting a mated butterfly is p

Probability of selecting mated butterflies 23 times: p and p and p and p.... = p^{23}

Probability of selecting unmated butterflies (32-23) = 9 times
(1-p)(1-p).... = $(1-p)^9$

Probability of M 23 times and V 9 times is $p^{23}(1-p)^9 = 5.53\times10^{-9}$

$$p^{23}(1-p)^{(32-23)}$$

Question: does the order of M’s and V’s matter?

The counting part

- There are many ways to get 23 M and 9 V
  1. MMMMMMMMMMMMMMMMMMMMMMMVVVVVVVVV
  2. MMMMMMMMMMMMMMMMMMMMMMVMVVVVVVVV

  ...

  ?. MVMVMMVMMVMMVVMMVMMVMMMMVMMMMMMM

  ...

  ?. VVVVVVVVVMMMMMMMMMMMMMMMMMMMMMMM
- The formula for the number of different ways to get 23 M out of 32 trials is:

$$\frac{n!}{r!(n-r)!} = \frac{32!}{23!(9!)} = 28{,}048{,}800$$

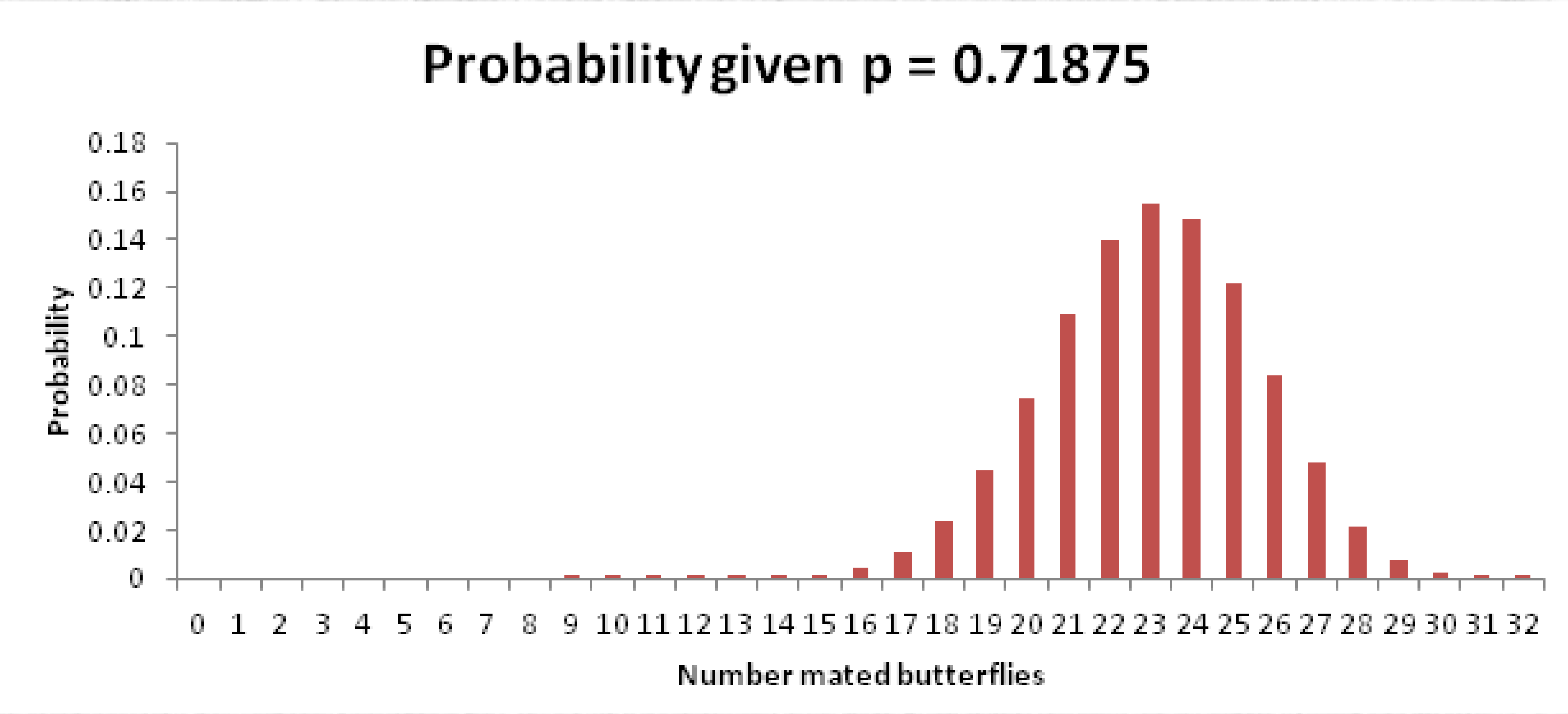
Combining the counting part and probability part

- Counting part: there are 28048800 ways to get 23 out of 32 mated flies
- Probability part: each has a probability of occurring of 5.53507×10^{-9}
- The probability of obtaining 23 out of 32 mated flies is the product of these two numbers

$$P(23\,out\,of\;32|\,p)=\binom{32}{23}p^{23}(1-p)^{(32-23)}=(28048800)(5.53507\times10^{-9})$$

$$P(23\,out\,of\;32|\,p)=0.155$$

The distribution of probabilities of all possible outcomes, given p



This is a discrete probability distribution – each bar is the probbility of one outcome, and the sum of the bar heights is 1

Likelihoods are not probabilities

- Likelihoods invert what we treat as known and what we treat as unknown
- Likelihoods ask: “What's the likelihood that p is 0.71875, given that 23 of 32 mated females were selected in our observed data set?”
  - Symbolically: L(p | 23 out of 32)
- We will use the binomial distribution as a **likelihood function**
  - Instead of plugging in different values for successes, we will hold successes and trials constant and vary p
  - This treats the data as a given, and treats p as unknown

The binomial likelihood function

- Looks just like a probability, but here we're treating p as unknown, and the data as a given
- What value of p should we use? Try 0.71875:

$$L(0.71875 | 23\, out\ of\ 32) = \binom{32}{23} 0.71875^{23} (1-0.71875)^{(32-23)} = 0.155$$

- But, 0.71875 is just one of infinite possibilities – how does it compare to other possible values of p?
- What value of p has the highest possible likelihood?

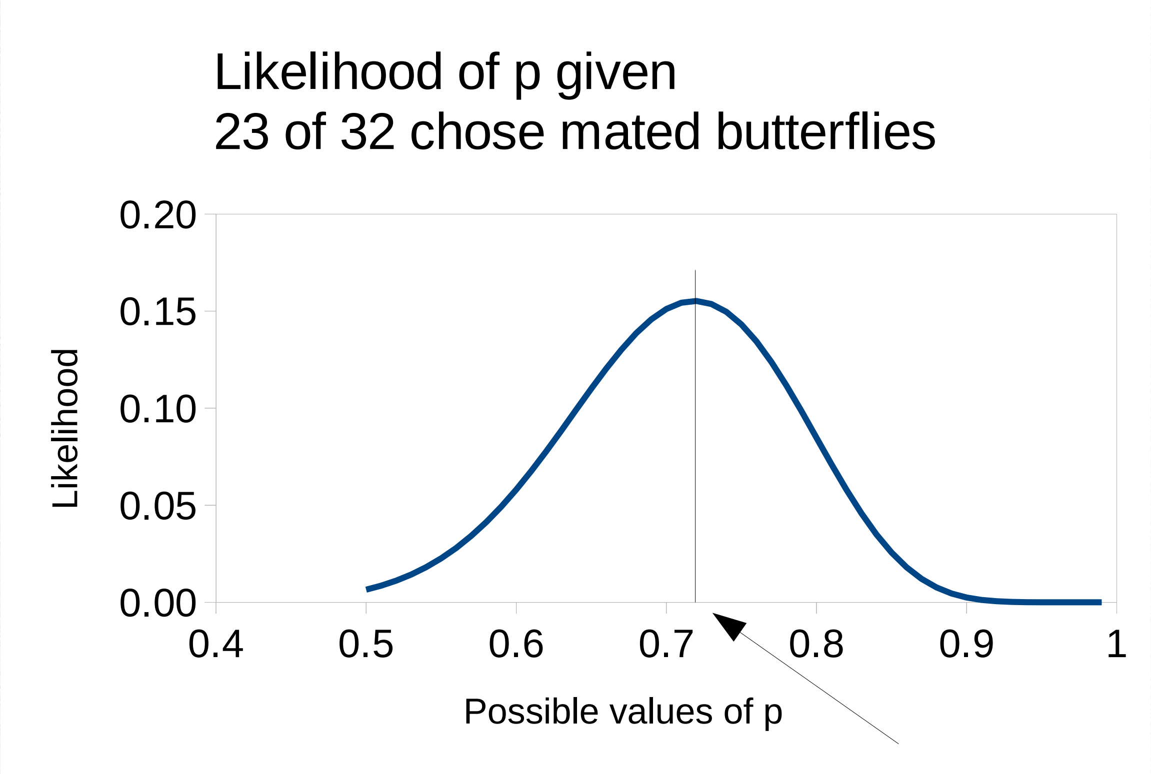
Finding the maximum likelihood

- Try out different possible values of p
- The value of p that produces the largest likelihood is the **maximum likelihood estimate**
- For these data,
  that value is p = 0.71875

	A	B	C	D	E	F
1	Possible p	Likelihood				
2	0.1	=BINOMDIST(23,32,A2,0) – gives 1.0867E-016				
3	0.2	3.1580E-010				
4	0.3	1.0656E-006				
5	0.4	0.0002				
6	0.5	0.0065				
7	0.6	0.0581				
8	0.7	0.1511				
9	0.71875	0.1553				
10	0.8	0.0848				
11	0.9	0.0025				
12	1	0.0000				
13						

- Thus, or formula for p of (# successes)/(# trials) is the maximum likelihood estimator for the population proportion

Likelihood function graphed



p is varied, while the outcome (23/32) is held constant

Very different from binomial probability distribution!

Continuous, smooth curve

Maximum likelihood estimate of p

This is not a probability distribution, and the area under the curve is not equal to 1

Working with log-likelihoods

- As we work with more complex problems, it's better to work with logs of likelihoods
  - Likelihoods are often less than 1 (often they are probabilities)
  - Combining likelihoods involves multiplying them, which quickly leads to very small numbers
  - Logs are exponents, which make the numbers we work with larger, and combining log-likelihoods means adding them
  - Log-likelihoods won't produce numbers too small for your calculator (or computer) to represent accurately
- Taking the log of the likelihood function changes its shape, but the maximum is still at the same value of p

Binomial log-likelihood

$$L(p|23\,out\,of\,32) = \binom{32}{23} p^{23} (1-p)^{(32-23)}$$

$$\ln[L(p|23\,out\,of\,32)] = \ln[\binom{32}{23} p^{23} (1-p)^{(32-23)}]$$

$$\ln[L(p|23\,out\,of\,32)] = \ln[\binom{32}{23}] + 23\ln p + 9\ln(1-p)$$

Drop unneeded terms?

- Notice that p doesn't appear in the first term (the “counting” part)

$$\ln\left[L\left(p|23\, out\, of\; 32\right)\right]=\ln\left[\binom{32}{23}\right]+23\ln p+9\ln\left(1-p\right)$$

- The counting part is an additive constant that doesn’t change shape of the likelihood function – can drop it without changing where the likelihood function maximizes
- If we drop the term, the likelihood function becomes different from the probability distribution
  - So, all likelihood functions are based on probability distributions, but not all likelihood functions are probability distributions

With or without the unneeded term

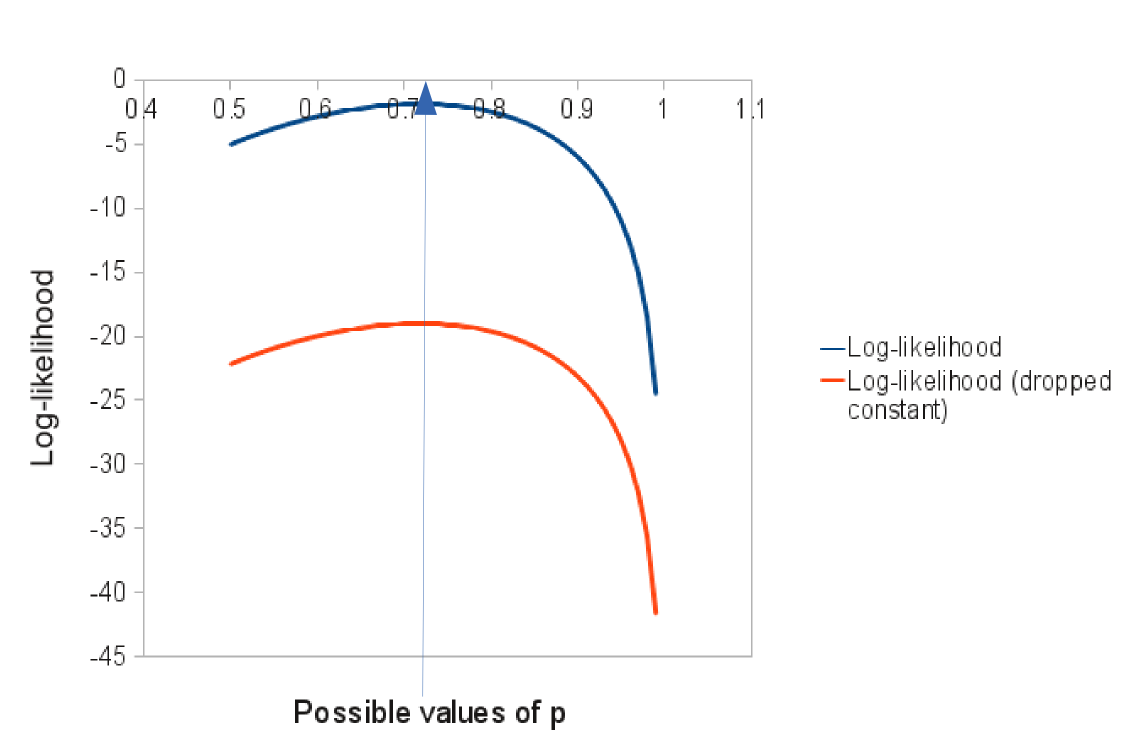
	A	B	C	D
1	Possible p	Likelihood	Log-likelihood	
2	0.1	0.00	=LN(B2) – gives -36.758	
3	0.2	0.00	-21.87591	
4	0.3	0.00	-13.75199	
5	0.4	0.00	-8.52266	
6	0.5	0.01	-5.03125	
7	0.6	0.06	-2.84615	
8	0.7	0.15	-1.88982	
9	0.71875	0.16	-1.86270	
10	0.8	0.08	-2.46779	
11	0.9	0.00	-5.99710	
12				

Take the log of the binomial likelihoods (keeps the counting term in the formula)

	A	B	C	D
1	Possible p	Likelihood	Log-likelihood	Log-likelihood (dropped constant)
2	0.1	0.00	-36.75825	=23*LN(A2) + 9*LN(1-A2) - gives -53.90770
3	0.2	0.00	-21.87591	-39.02536
4	0.3	0.00	-13.75199	-30.90145
5	0.4	0.00	-8.52266	-25.67212
6	0.5	0.01	-5.03125	-22.18071
7	0.6	0.06	-2.84615	-19.99561
8	0.7	0.15	-1.88982	-19.03928
9	0.71875	0.16	-1.86270	-19.01216
10	0.8	0.08	-2.46779	-19.61724
11	0.9	0.00	-5.99710	-23.14656

Drop the unneeded counting term

Likelihood functions with or without unneeded term

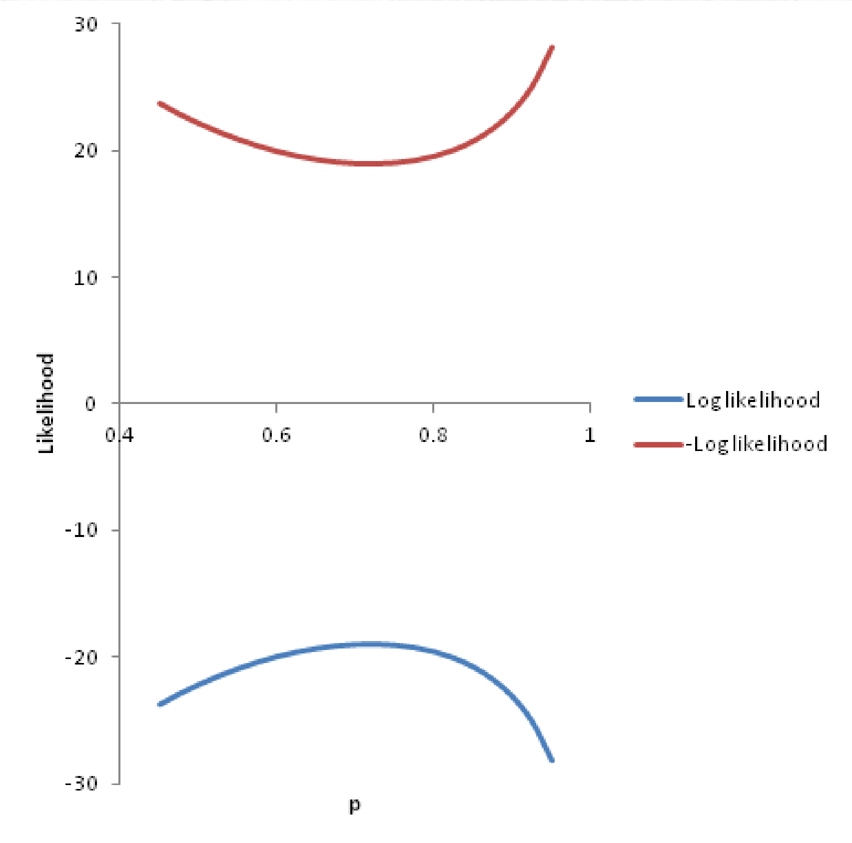


Same shapes, differ by a constant amount across the whole curve, both identify the same best value for the estimate of p

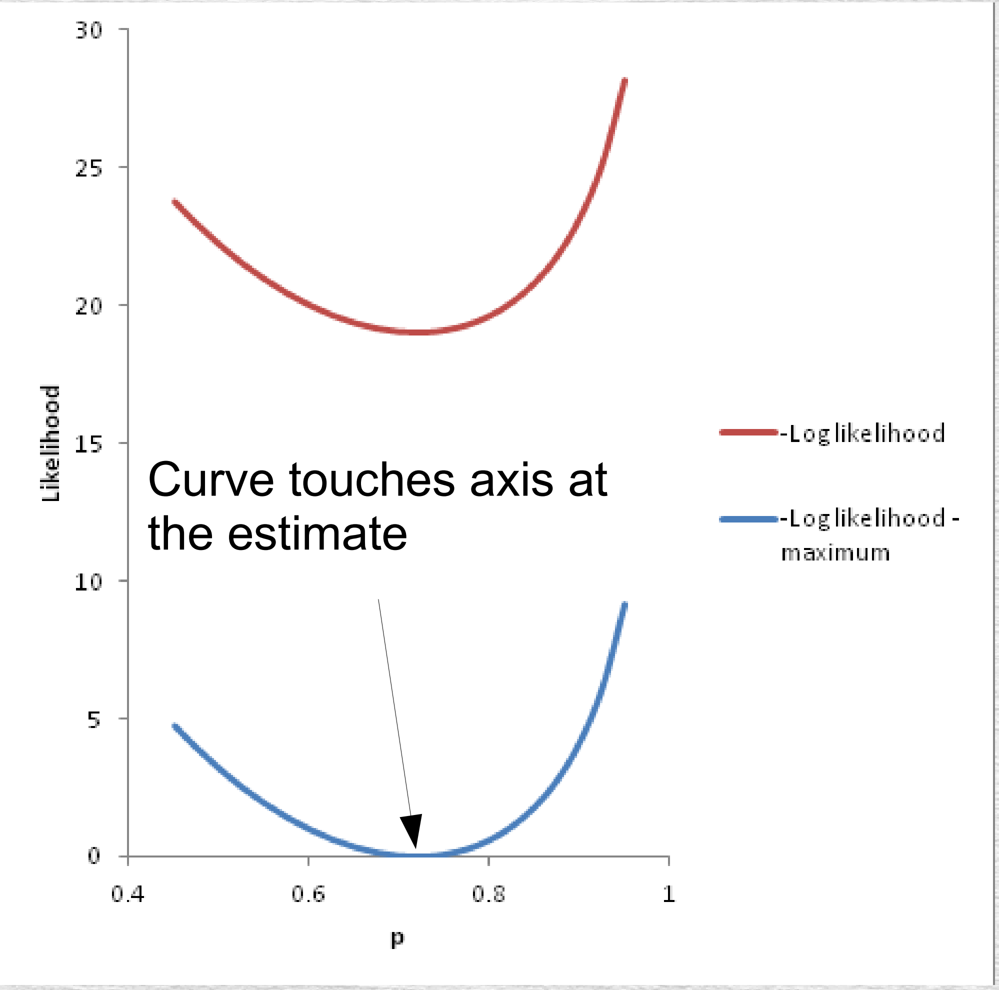
-Log Likelihood

- Using the negative of the log-likelihood changes the direction of the curve (flips it around the x-axis)
- The maximum likelihood estimate becomes the minimum of the -log likelihood
- It is also common to subtract the minimum from the likelihood function so that the maximum likelihood estimate has a value of 0
- This has some advantages
  - We're better at interpreting relationships between positive numbers
  - The -log likelihood follows a Chi-square distribution, which we can use to calculate confidence intervals

Switching from log likelihood to -log likelihood



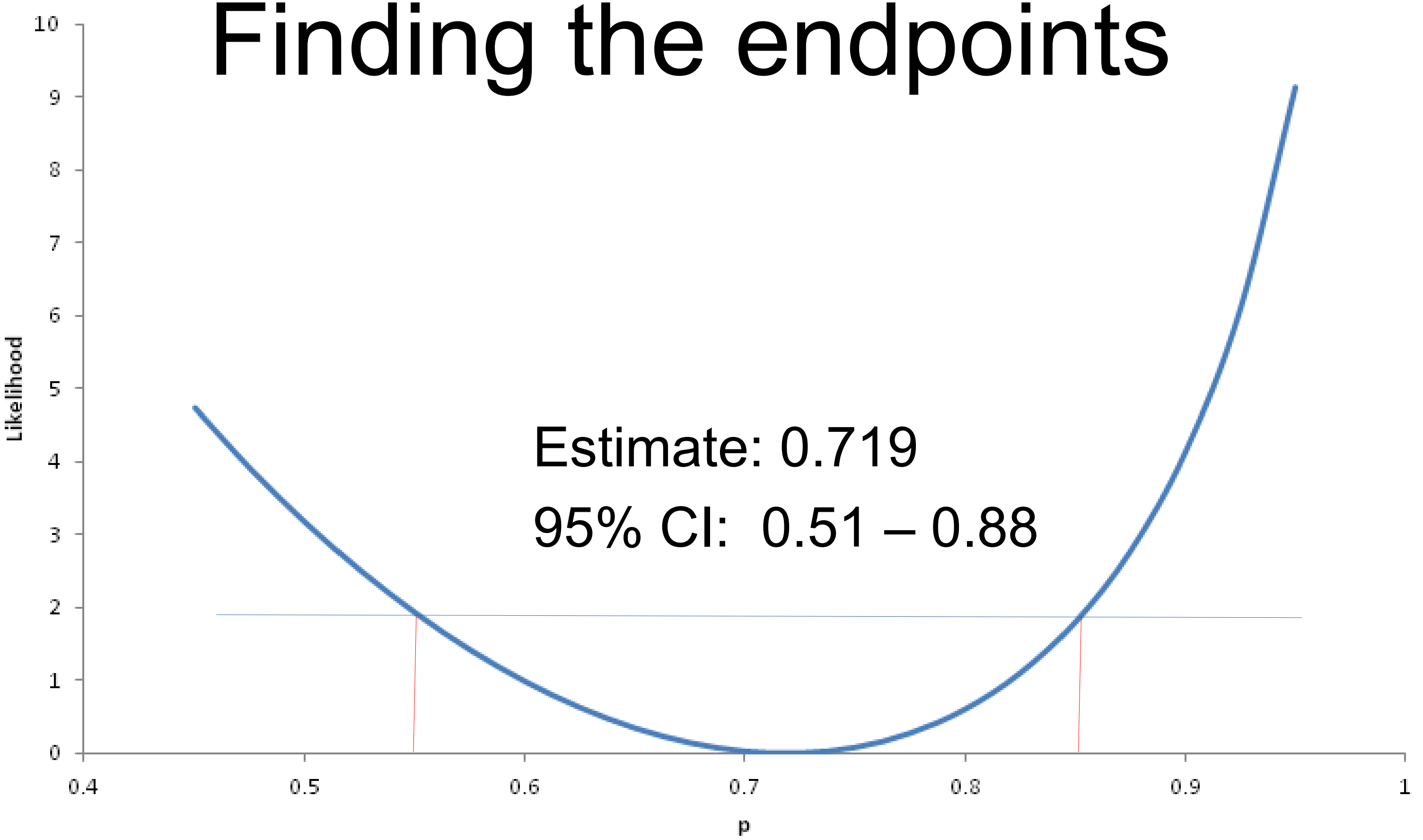
Subtracting the minimum -log likelihood



Using likelihood functions to obtain confidence intervals

- Called **profile likelihood** intervals
- Twice the negative log-likelihoods are Chi-square distributed
- Identify upper and lower limits based on the critical value from a Chi-square distribution
  - $\chi^2_{df=1,\alpha=0.05}/2 = 1.92$
  - Thus... values of p with log-likelihoods within 1.92 units of the maximum likelihood are within the interval
  - And...there is a 95% chance that possible values of p that fall within this interval could be the population parameter

Finding the endpoints



Advantage: profile likelihood CI's are not symmetrical around p

Adding data

- What if instead of 23 out of 32 wasps choosing mated butterflies, we had gotten 46 out of 64?
- No obvious place for n to be used in our 95% CI's, but n does affect the likelihood function
- Compare the two...

Effect of increased n on likelihood-based confidence intervals