

Likelihood

Outline for today

- What is probability
- What is likelihood
- Maximum likelihood estimation
- Example: estimate a proportion
- Likelihood-based confidence intervals
- Example: estimating speciation and extinction rates
- Log-likelihood ratio test

What is probability

The *probability* of an event is the proportion of times that the event would occur if we repeated a random trial over and over again under the same conditions.

A *probability distribution* is a list of all mutually exclusive outcomes of a random trial and their probabilities of occurrence.

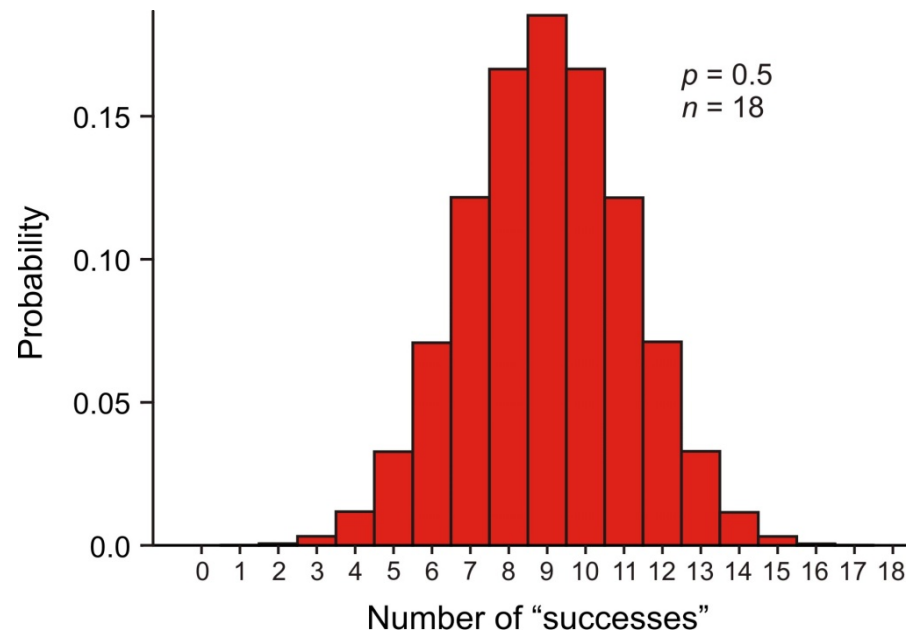
Example: binomial distribution

The *binomial distribution* is the probability distribution of the number of “successes” in n independent trials, when the probability of success p is the same in each trial.

$$\Pr[Y \text{ successes}] = \binom{n}{Y} p^Y (1-p)^{n-Y}$$

$\binom{n}{Y}$ counts up the different ways of getting Y successes and $n - Y$ failures (e.g., S,S,F; S,F,S; F,S,S)

Graph on right shows $\Pr[0]$, $\Pr[1]$, ... when $p = 0.50$ and $n = 18$



What is conditional probability

The *conditional probability* of an event is the probability of that event occurring given that a condition is met.

“|” symbol used to indicate “given that”

The probability that the second child born to a couple is a girl given that their first child was a girl

$\text{Pr}[\text{second child is girl} \mid \text{first child is girl}]$

$\text{Pr}[\text{we see an elephant today} \mid \text{we are in Serengeti}]$

$\text{Pr}[\text{we see an elephant today} \mid \text{we are in Manhattan}]$

What is likelihood

Likelihood is a conditional probability.

The likelihood of a population parameter equaling a specific value, given the data, is the probability of obtaining the observed data *given* that the population parameter equals the specific value.

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

Useful factoid:

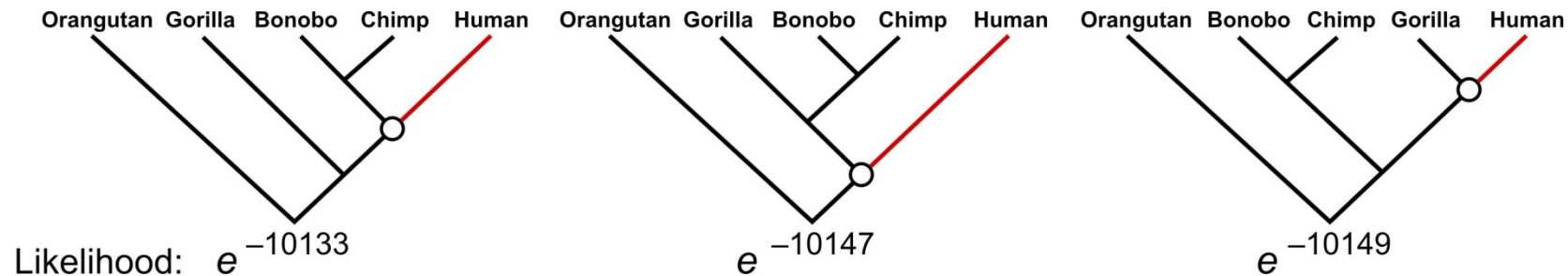
The extent to which data supports one parameter value or hypothesis against another is equal to the ratio of their likelihoods (difference in their log-likelihoods) (“Law of Likelihood”)

Method invented by R. A. Fisher when a 3rd-year undergraduate.

Likelihood is used a lot in phylogeny estimation

Three proposed trees of ancestor–descendant relationships between humans and the other great apes. The human branch and our shared ancestor with the other apes is highlighted. Numbers at the bottom are the likelihoods of each proposed tree based on gene sequence data (Rannala and Yang 1996). The likelihood of the left-most tree is the highest.

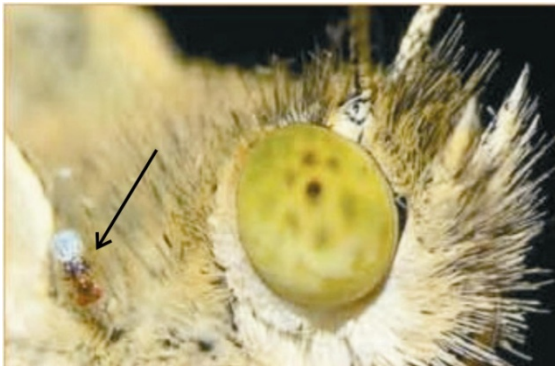
$$L[\text{tree} \mid \text{gene sequences}] = \Pr[\text{gene sequences} \mid \text{tree}]$$



What matters is not the likelihood of each tree as such, but the likelihood of each tree relative to the others.

Example 1: Estimate a binomial proportion p

Data: The tiny wasp, *Trichogramma brassicae*, rides on female cabbage white butterflies, *Pieris brassicae*. When a butterfly lays her eggs on a cabbage, the wasp climbs down and parasitizes the freshly laid eggs. Fatouros et al. (2005) carried out trials to determine whether the wasps can distinguish mated female butterflies from unmated females. In each trial a single wasp was presented with two female cabbage white butterflies, one a virgin female, the other recently mated. $Y = 23$ of 32 wasps tested chose the mated female. What is the proportion p of wasps in the population choosing the mated female?



$Y = 23, n = 32$. Estimate p .

Likelihood function for the binomial proportion p

$$L[p \mid Y \text{ chose mated female}] = \Pr[Y \text{ chose mated female} \mid p]$$
$$= \binom{n}{Y} p^Y (1-p)^{n-Y}.$$

Data: $Y = 23, n = 32$

$$L[p \mid 23 \text{ chose mated}] = \binom{32}{23} p^{23} (1-p)^9.$$

For example, the likelihood of $p = 0.5$, given the data, is

$$L[p = 0.5 \mid 23 \text{ chose mated}] = \binom{32}{23} (0.5)^{23} (1-0.5)^9$$
$$= 0.00653.$$

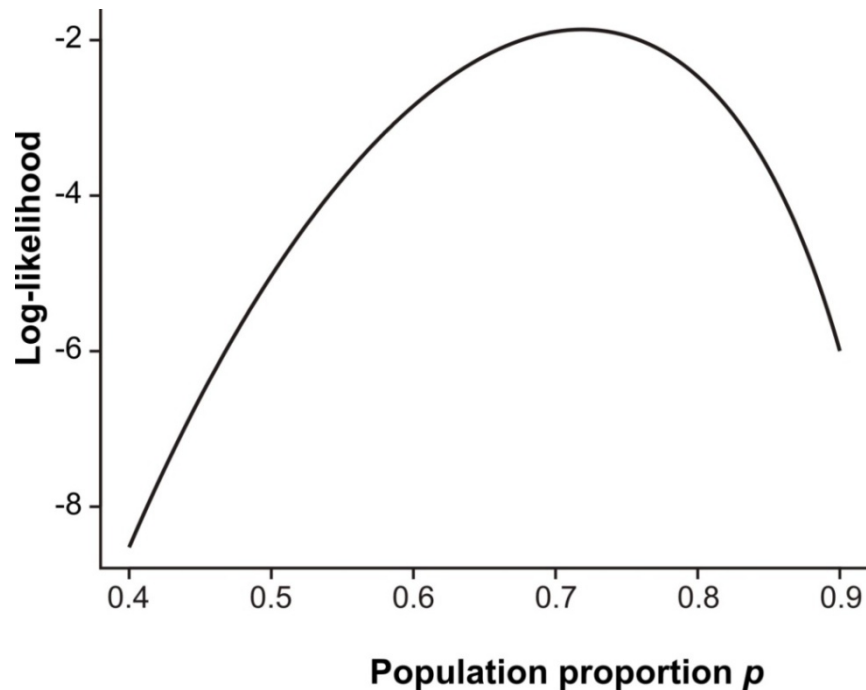
```
In R:  
dbinom(23, 32, prob=0.5)  
[1] 0.00653062
```


Easier to work with log-likelihoods

$$\ln L[0.5 | 23 \text{ choose mated}] = \ln \left[\binom{32}{23} \right] + 23 \ln[0.5] + 9 \ln[1 - 0.5]$$
$$= -5.03125.$$

```
dbinom(23, 32, prob=0.5, log=TRUE)
[1] -5.031253
```

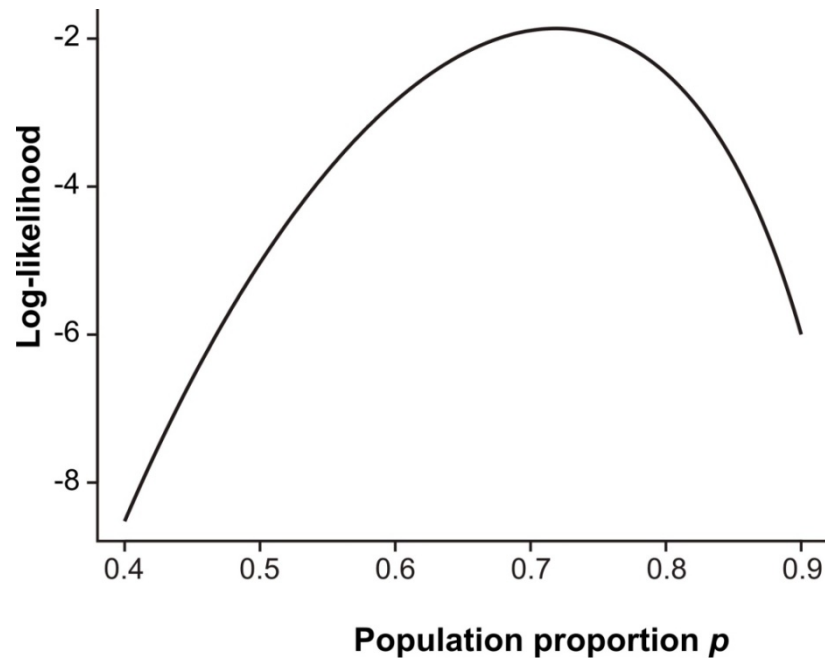
Repeat for many values of p to get the log-likelihood curve:



Likelihood works backward from probability

Typically we use probability to predict unknown data outcomes based on *known parameters*.

Here we use likelihood to estimate unknown parameters based on *known data*.



The likelihood function is not a probability distribution

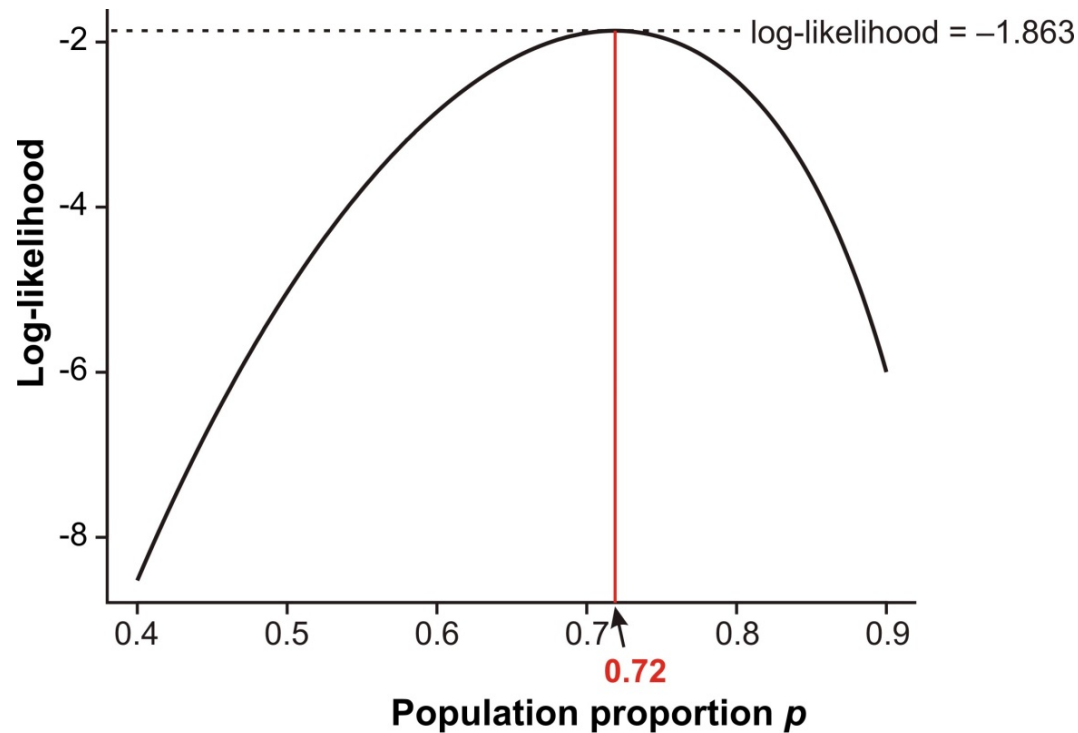
The population proportion p is the variable of the function, but it is not a random variable (its value is not determined by random trial).

The likelihood ratio (difference of log-likelihood) measures relative support for alternative parameter values



Maximum likelihood estimate

The *maximum likelihood estimate* of a parameter is the parameter value having the highest likelihood (and log-likelihood), given the data. This is the parameter value most strongly supported by the data.

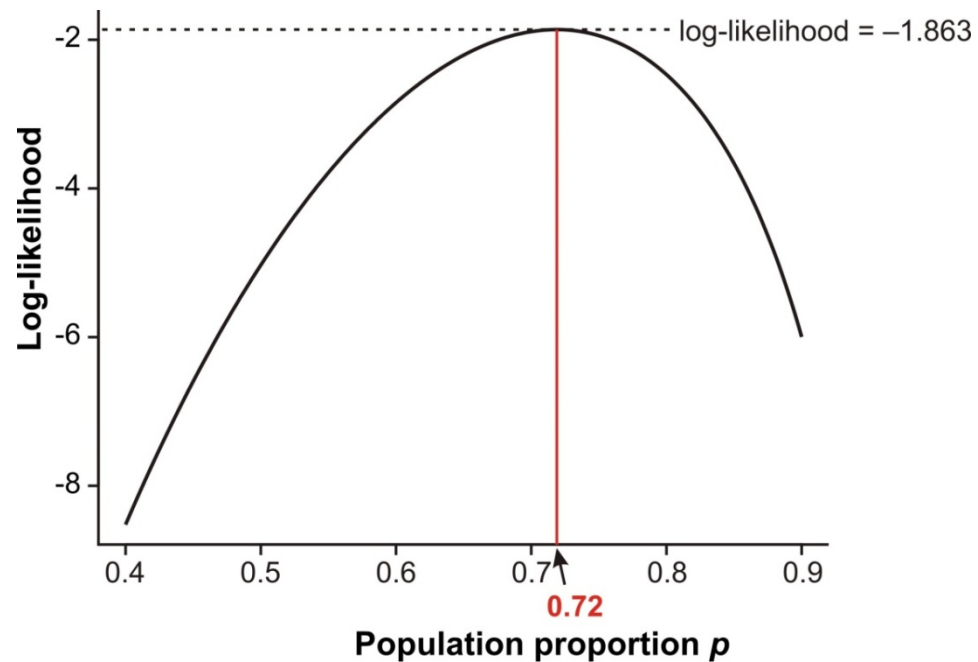


Maximum likelihood estimate

The ML estimate could have been obtained more easily as

$$p = Y/n = 2332/3300 = 0.71$$

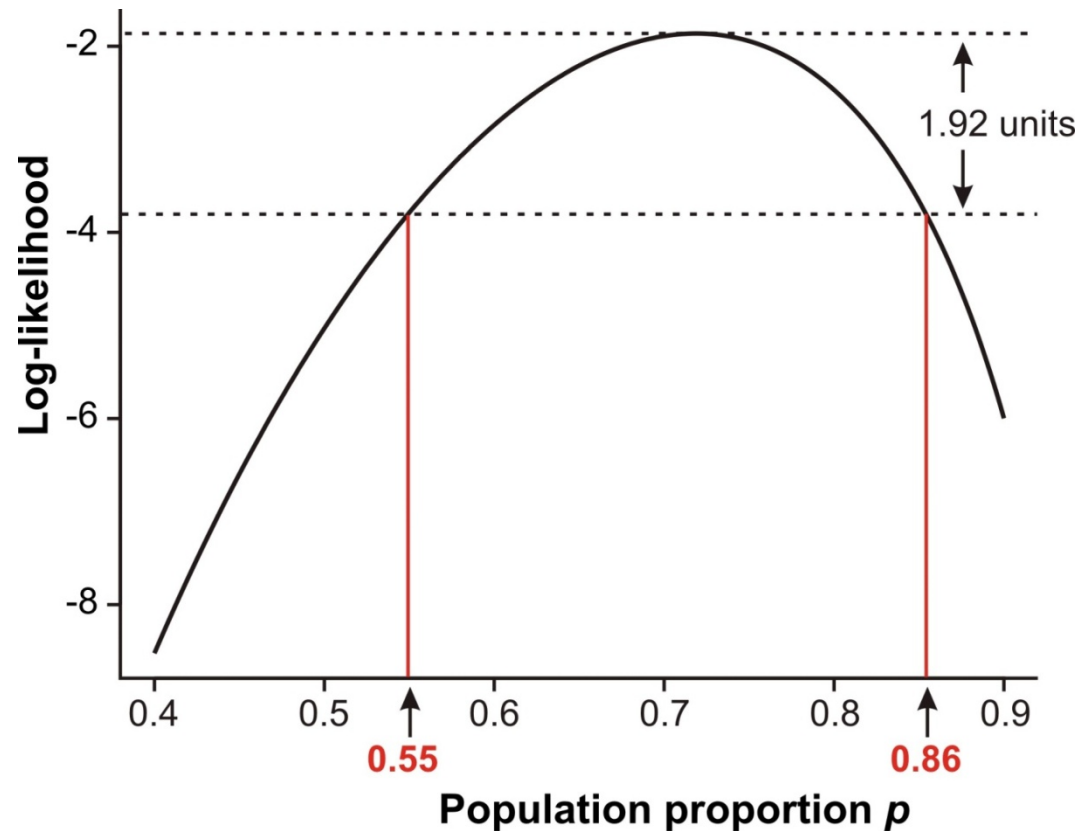
The conventional formula for estimating a proportion yields the ML estimate. Most estimates you are familiar with are ML estimates.



Likelihood-based confidence intervals

When estimating a single parameter, approximate 95% confidence intervals are obtained as the values corresponding to 1.92 log-likelihood units below the maximum.

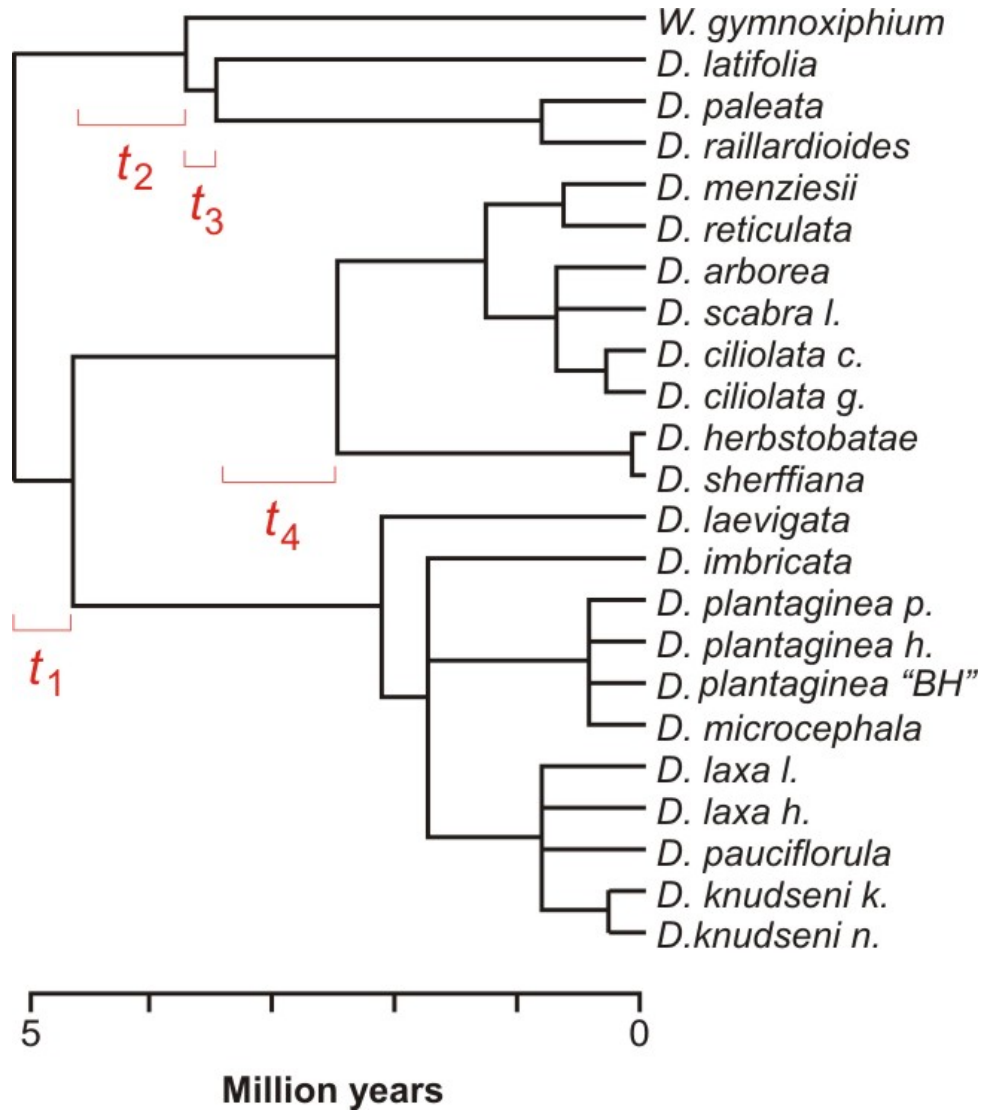
So the 95% CI for p in the wasp example is
 $0.55 \leq p \leq 0.86$



(1.92 is $\chi_1^2/2$)

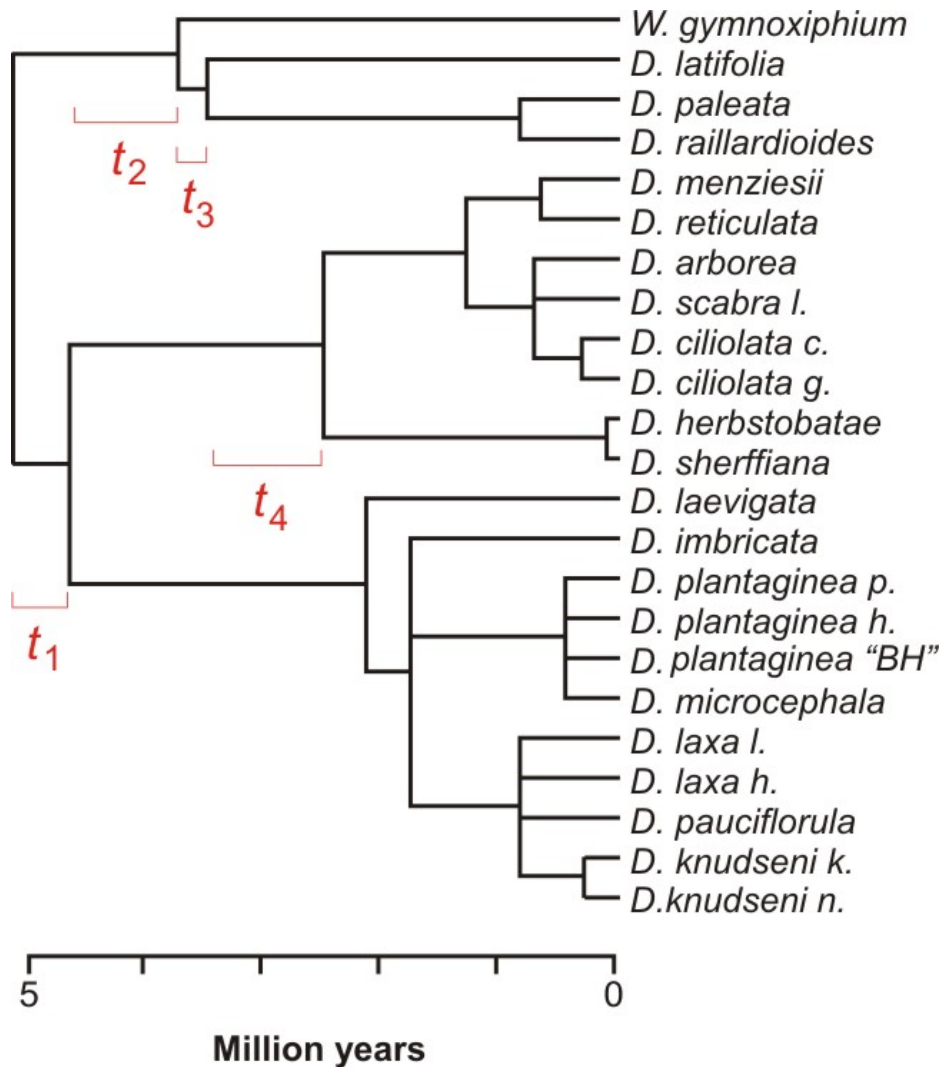
[break]

Example 2: Estimating speciation and extinction rates



You don't need to be a mathematician to use likelihood

What you need is the probability distribution



Nee et al (1994) found the probability distribution for the waiting time between branching events on the “reconstructed” phylogeny of extant species, assuming a constant birth-death process.

From (15) we can derive the probability density of t , the waiting time for a birth:

$$n(\lambda - \mu)e^{-n(\lambda - \mu)t} \times \frac{(1 - \frac{\mu}{\lambda} \exp(-(\lambda - \mu)(T - t_n - t)))^{n-1}}{(1 - \frac{\mu}{\lambda} \exp(-(\lambda - \mu)(T - t_n)))^n}, \quad (17)$$

You don't need to be a mathematician to use likelihood

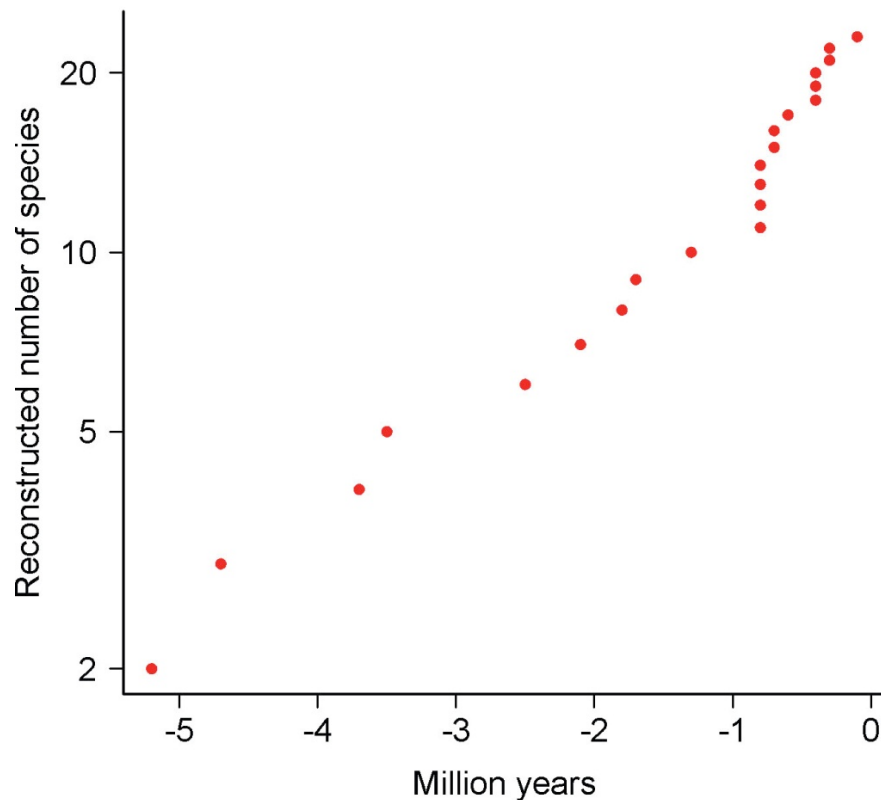
Use the formula for the probability distribution to calculate the likelihood, given the data

Branching times:

5.2, 4.7, 3.7, 3.5, 2.5, 2.1, 1.8, 1.7, 1.3, 0.8, 0.8, 0.8, 0.8, 0.7, 0.7, 0.6, 0.4, 0.4, 0.4, 0.3, 0.3, 0.1

Compute the waiting times t :

0.5, 1.0, 0.2, 1.0, 0.4, 0.3, 0.1, 0.4, 0.5, 0.0, 0.0, 0.0, 0.1, 0.0, 0.1, 0.2, 0.0, 0.0, 0.1, 0.0, 0.2



$$L[\hat{\lambda}, \hat{\mu} | t] = P[t | \hat{\lambda}, \hat{\mu}]$$

Grid search over possible parameter values

a **r** **loglike**

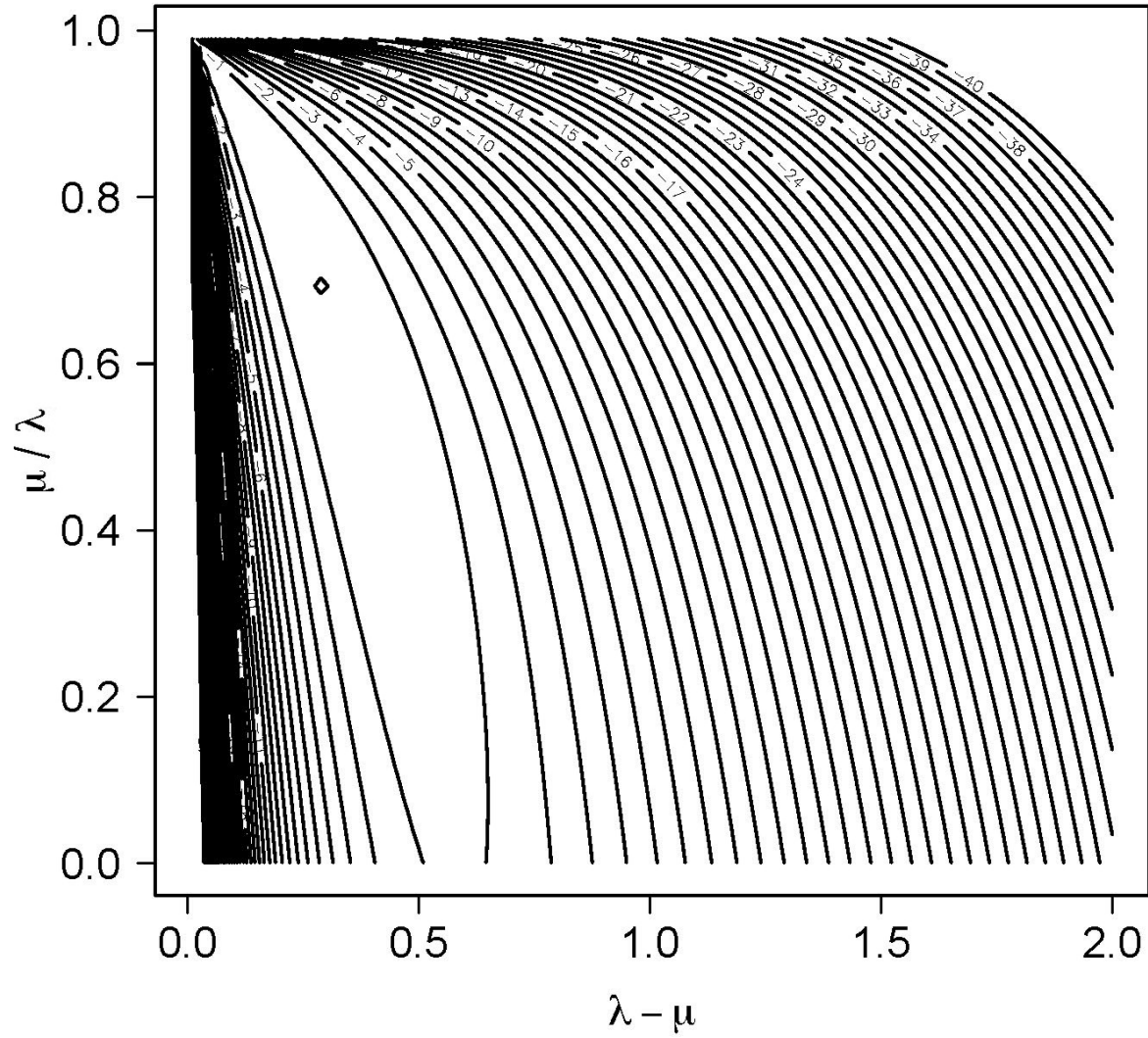
where $a = \lambda - \mu$ and $r = \mu / \lambda$

($\lambda \sim$ 'speciation rate'; $\mu \sim$ 'extinction rate')

0.1	0.1	-1.937178
0.1	0.2	-0.212507
0.1	0.3	1.662159
0.1	0.4	3.712741
0.1	0.5	5.969948
0.1	0.6	8.46619
0.1	0.7	11.21831
0.1	0.8	14.13799
0.1	0.9	16.39656
0.2	0.1	8.510571
0.2	0.2	9.701548
0.2	0.3	10.94532
0.2	0.4	12.23591
0.2	0.5	13.5552
0.2	0.6	14.85593
0.2	0.7	16.01336
0.2	0.8	16.65578
0.2	0.9	15.30007
0.3	0.1	13.02322
0.3	0.2	13.81377

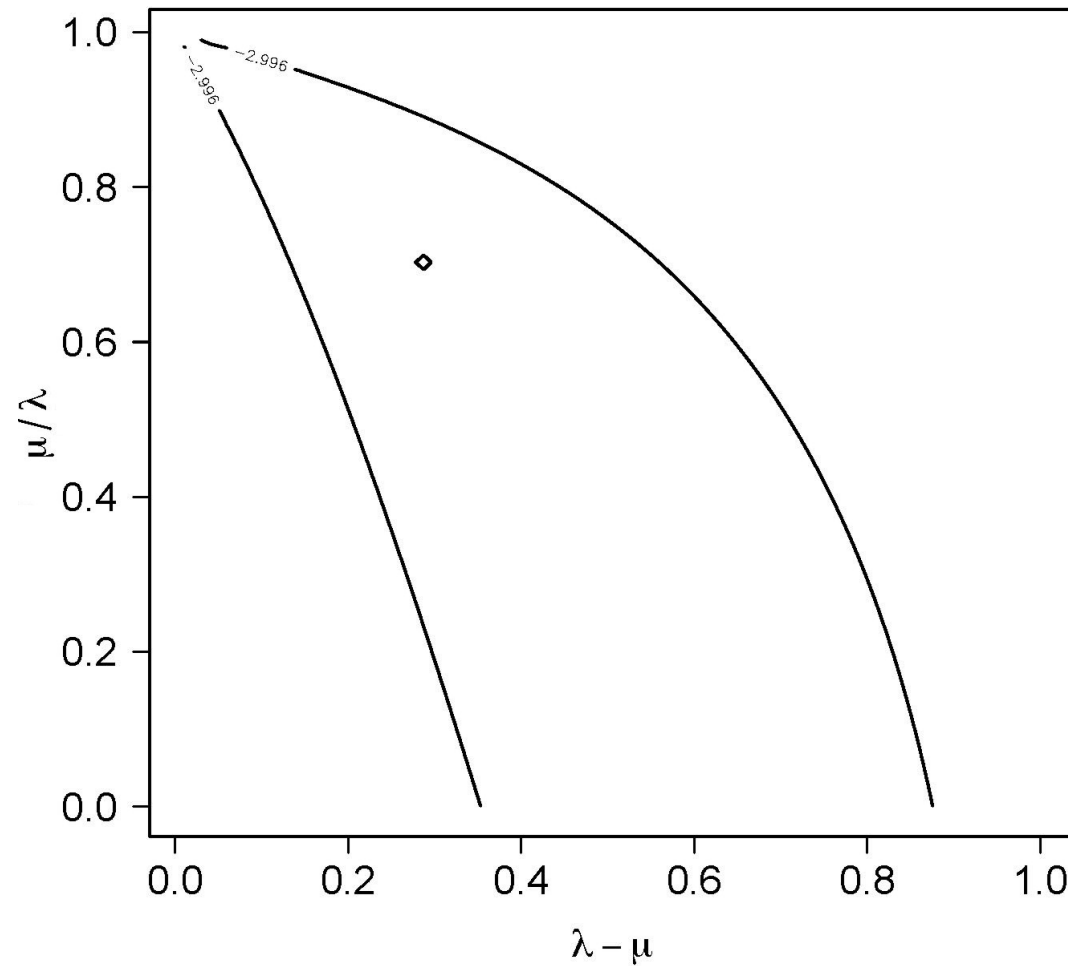
.....

Log-likelihood surface



Likelihood-based confidence region

When estimating two parameters, an approximate 95% confidence region is obtained by the values corresponding to 2.996 log-likelihood units below the maximum



(2.996 is $\chi_2^2/2$)

Example 2: Conclusion: estimate speciation, extinction in silverswords

- Maximum likelihood the only way to estimate λ, μ
- The approach is the same as with a single parameter
- Needed to find a formula for the probability distribution
- Use R to calculate log-likelihood for a range of possible values for λ, μ
- Contour plot to get likelihood surface and confidence region
- Very few species in silversword example, so estimates not well bounded
- Also, the χ^2 approximation dubious with so few species, need to check using simulation

Log-likelihood ratio test

Hypothesis test comparing the likelihoods of two models fit to the data.

“Full model” vs “Reduced model”, just as in linear models.

For example:

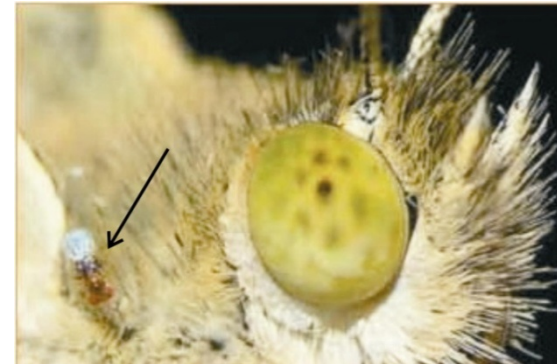
H_0 : Wasps choose mated and unmated females with equal probability ($p = 0.5$)

(= “Reduced model”)

H_A : Wasps prefer one type of female over the other ($p \neq 0.5$)

(= “Full model”)

Data: $Y = 23, n = 32$.



Log-likelihood ratio test

$$G = 2 \ln \left(\frac{L[\text{"Full model"} | \text{data}]}{L[\text{"Reduced model"} | \text{data}]} \right)$$

Both models are fitted using maximum likelihood.

The log-likelihood ratio statistic G is approximately χ^2 distributed with degrees of freedom equal to the difference between the “Full” and “Reduced” models in the number of parameters estimated from the data.

Very general method – applies to all probability distributions.

The approximation improves with increasing sample size.

Log-likelihood ratio test

Wasp example

$$G = 2 \ln \left(\frac{L[p = \hat{p} = 0.72 \mid 23 \text{ of } 32 \text{ chose mated female}]}{L[p = p_0 = 0.50 \mid 23 \text{ of } 32 \text{ chose mated female}]} \right)$$

From calculations using formulae shown earlier,

$$L[0.72 \mid 23 \text{ chose mated female}] = 0.1553$$

$$L[0.50 \mid 23 \text{ chose mated female}] = 0.00653$$

$$G = 2 \ln(0.1553 / 0.00653) = 6.336$$

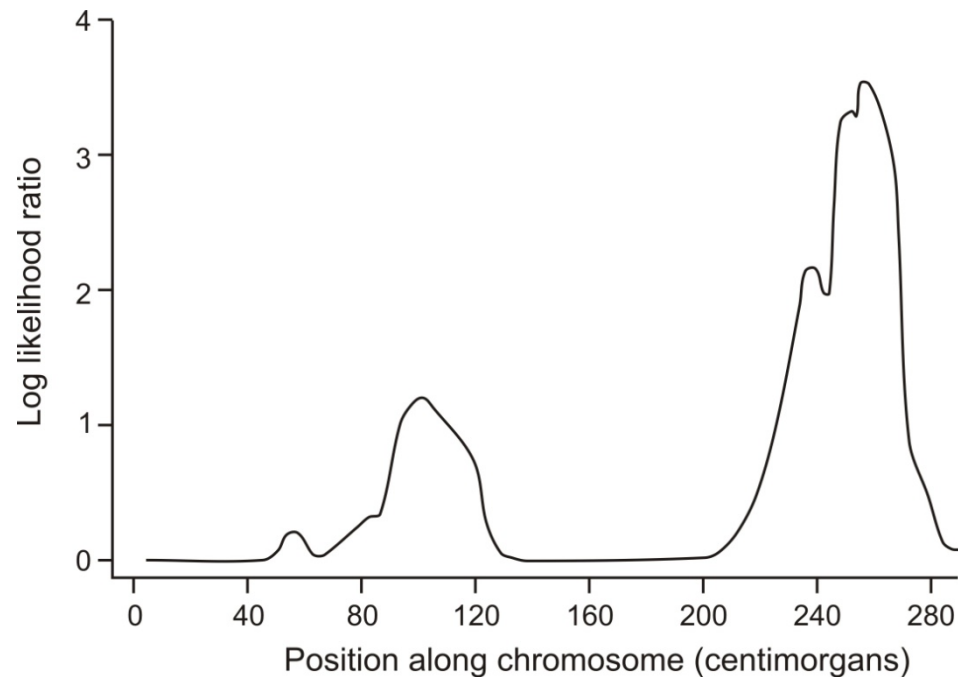
$$\chi_1^2 = 3.841$$

Therefore reject H_0 .

Log-likelihood ratios are used a lot in gene mapping

The data are the observed marker states of healthy and diseased individuals. At each marker along the chromosome two models are fitted to the data. The “full model” estimates the proportion of diseased individuals separately for each genotype at the marker. The “reduced model” fits an equal proportion of diseased individuals at all genotypes at the marker. Each likelihood is the probability of obtaining the observed data if the given model is correct. The log of the ratio of these two likelihoods (“full” divided by “reduced”) reflects the strength of evidence that a gene is located at a marker site.¹

Evidence for a gene affecting schizo-affective disorder on human chromosome 1.



¹ The arbitrary convention in gene mapping studies is to use the base-10 logarithm rather than the natural log. The resulting quantity is called the “LOD score”.

Discussion paper for next week:

Verhoeven et al. (2005) Controlling false discovery rate when multiple testing.

Download from “**assignments**” tab on course web site.

Presenters: Kim & Sarah

Moderators: Alice & Ian