Outline for today

- What is probability

- What is likelihood

- Maximum likelihood estimation

- Example 1: estimate a proportion

- Likelihood works backward from probability

- Likelihood-based confidence intervals

- Example 2: estimate survival curve

- Log-likelihood ratio test

- Example: test a proportion

What is probability

***Frequentist definition*:**

The *probability* of an event is the <u>proportion</u> of times that the event would occur if a random trial is repeated 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.

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

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

Graph shows Pr[0], Pr[1] , Pr[2], ...
for example when p = 0.50 and n = 18
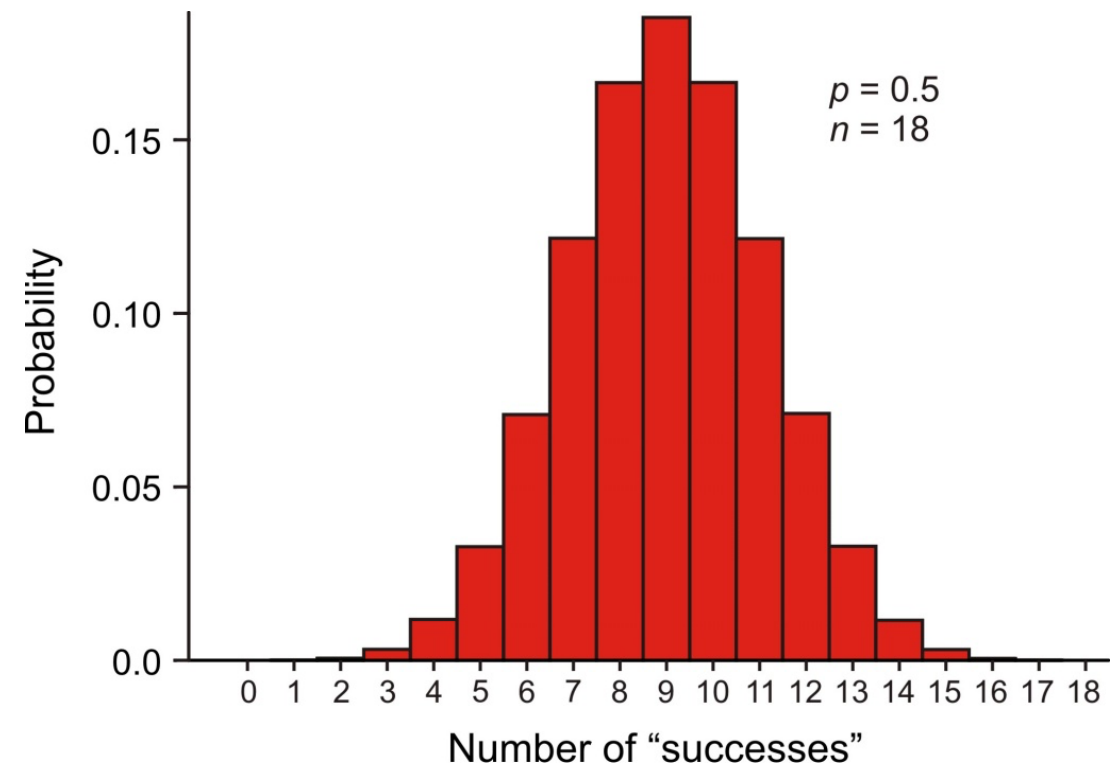
What is conditional probability

The *conditional probability* of an event is the probability of that event occurring given that a condition is met. "|" symbol used to indicate "given"

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

    Pr[second child is girl | first child is girl]

Other conditional probabilities:

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 | $p = 0.10$]

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[parameter = p | data] = Pr[data | parameter = p]

L[p = 0.10 | 12 successes out of 18 trials] = Pr[12 successes in 18 trials | p = 0.10]

Law of Likelihood:

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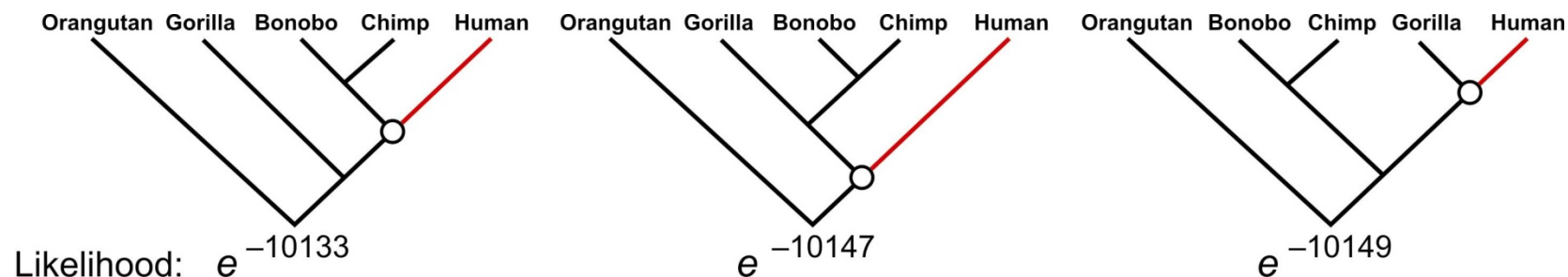
Method invented by R. A. Fisher when a 3rd-year undergraduate.

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 is highlighted. Numbers at the bottom are the likelihoods of each proposed tree based on gene sequence data and probability model of sequence evolution (Rannala and Yang 1996).
The likelihood of the left-most tree is the highest.

L[tree = i | gene sequences] = Pr[gene sequences | tree = i]

Orangutan Gorilla Bonobo Chimp Human   Orangutan Gorilla Bonobo Chimp Human   Orangutan Bonobo Chimp Gorilla Human

Likelihood: e^{-10133} e^{-10147} e^{-10149}

What matters is not the likelihood of each tree, but the likelihood of each tree <u>relative to the others</u>.

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.
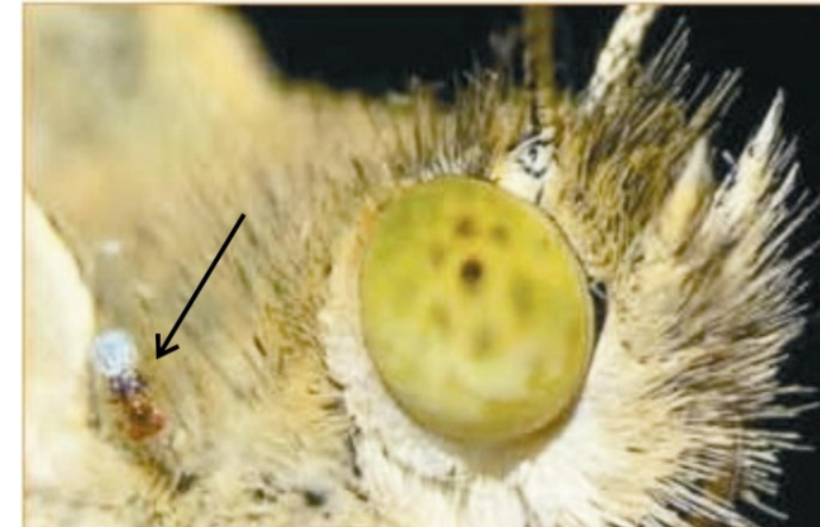
Fatouros et al. (2005) carried out trials to determine whether the wasps can distinguish mated female butterflies from unmated females. In each trial a single wasp was presented with two female cabbage white butterflies, one a virgin female, the other recently mated. Y = 23 of 32 wasps tested chose the mated female.

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

Y = 23 "successes", n = 32 trials. Use these data to estimate p.

Example 1: Estimate a binomial proportion *p*

Likelihood function for the binomial proportion *p*

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

$$L[p \mid Y \text{ chose mated female}] = \Pr[Y \text{ chose mated female} \mid p]$$

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

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(x = 23, size = 32, prob=0.5)
[1] 0.00653062
```
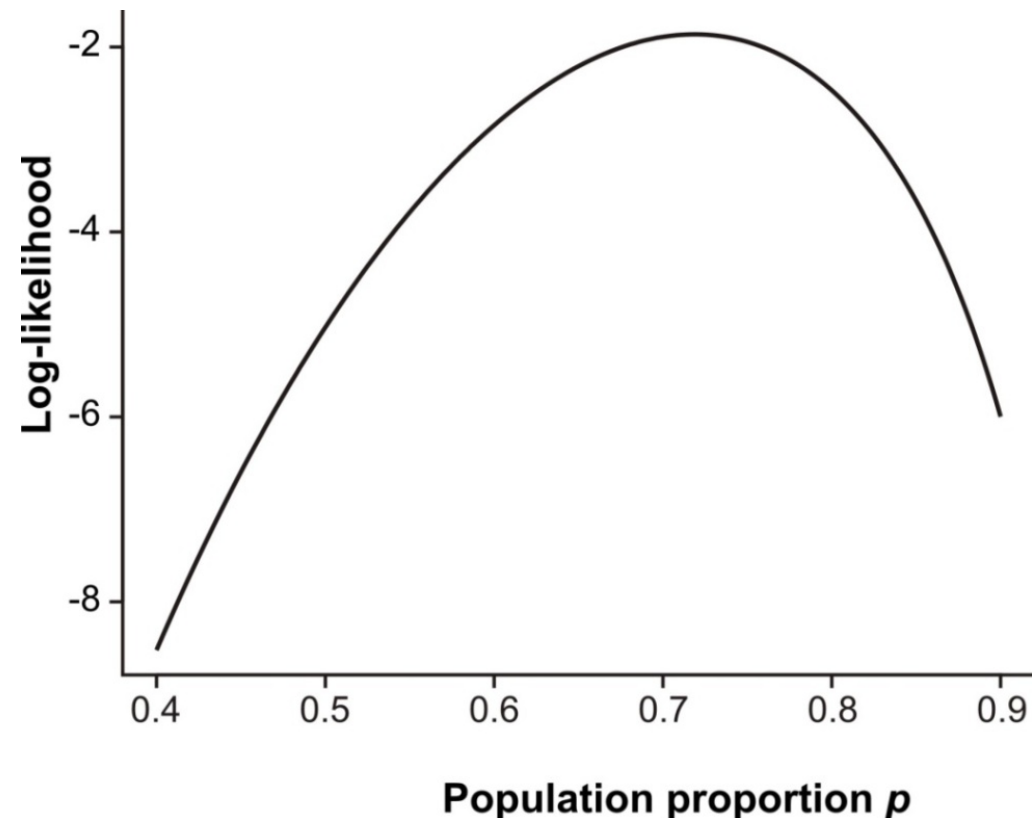
Example 1: Estimate a binomial proportion *p*

Easier to work with log-likelihoods

$$\ln L[0.5 \mid 23 \text{ chose mated}] = \ln\binom{32}{23} 23 \ln(0.5) \, 9 \ln(1-0.5) = -5.03125$$

in R:
```
dbinom(x = 23, size = 32, prob = 0.5, log = TRUE)
[1] -5.031253
```
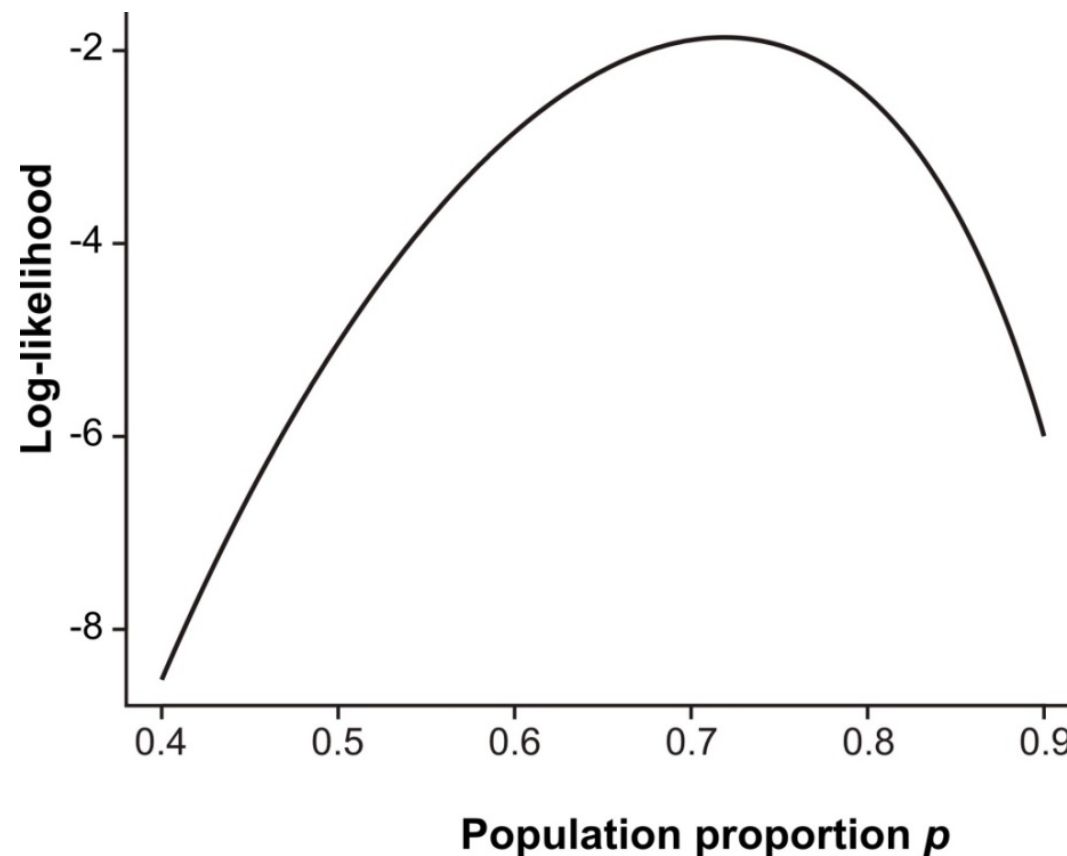
Repeat for many values of *p*
to get the log-likelihood curve:

Likelihood works backward from probability.

We use likelihood to estimate unknown parameters based on *known data*.
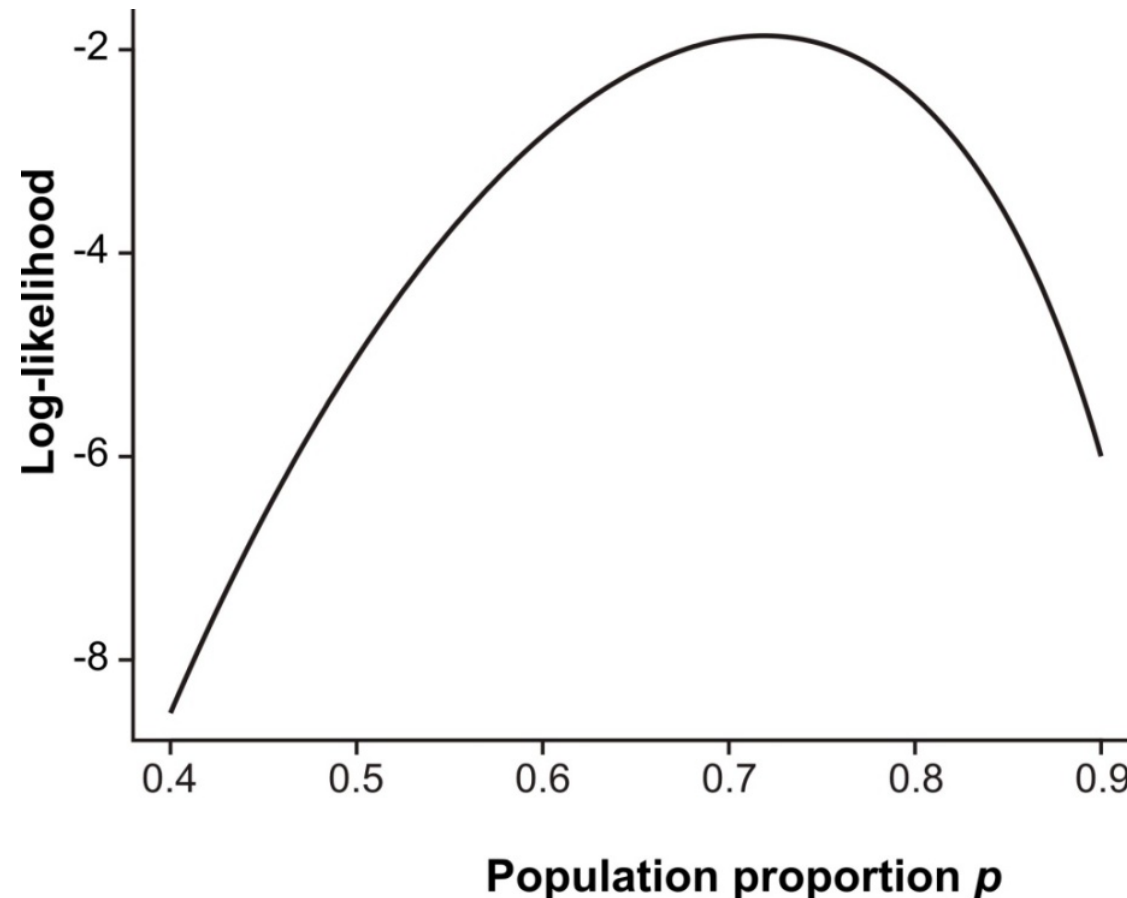
The parameters are treated as variables, the data are constant.

**Likelihood works backward from probability.**

But the likelihood function is not a probability distribution.

The population proportion p is the variable of the function, but it is not a <u>random</u> variable (its value is not determined as the outcome of a random trial).

Maximum likelihood estimate

The likelihood ratio (difference of log-likelihoods) measures relative support for alternative parameter values. The *maximum likelihood estimate* (MLE) 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.
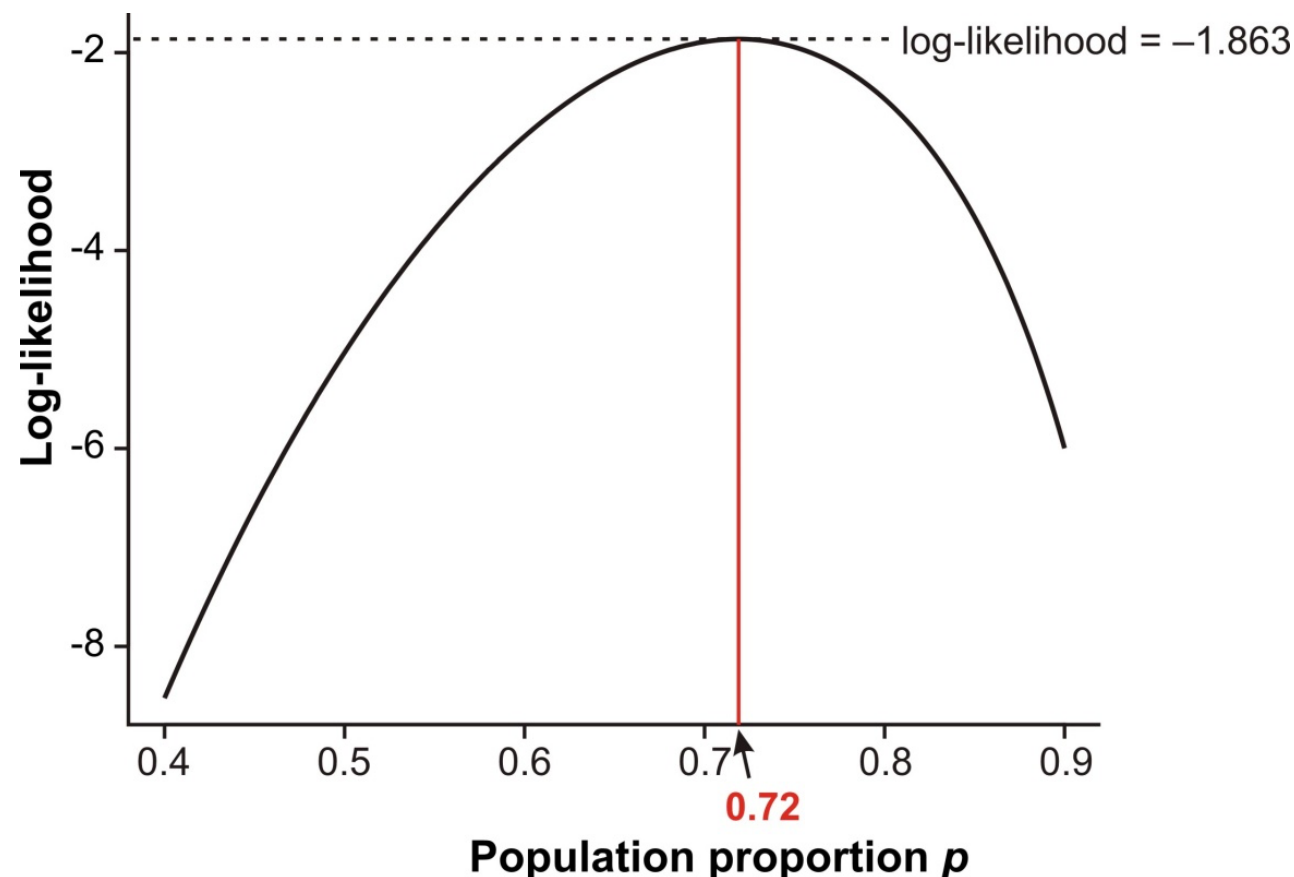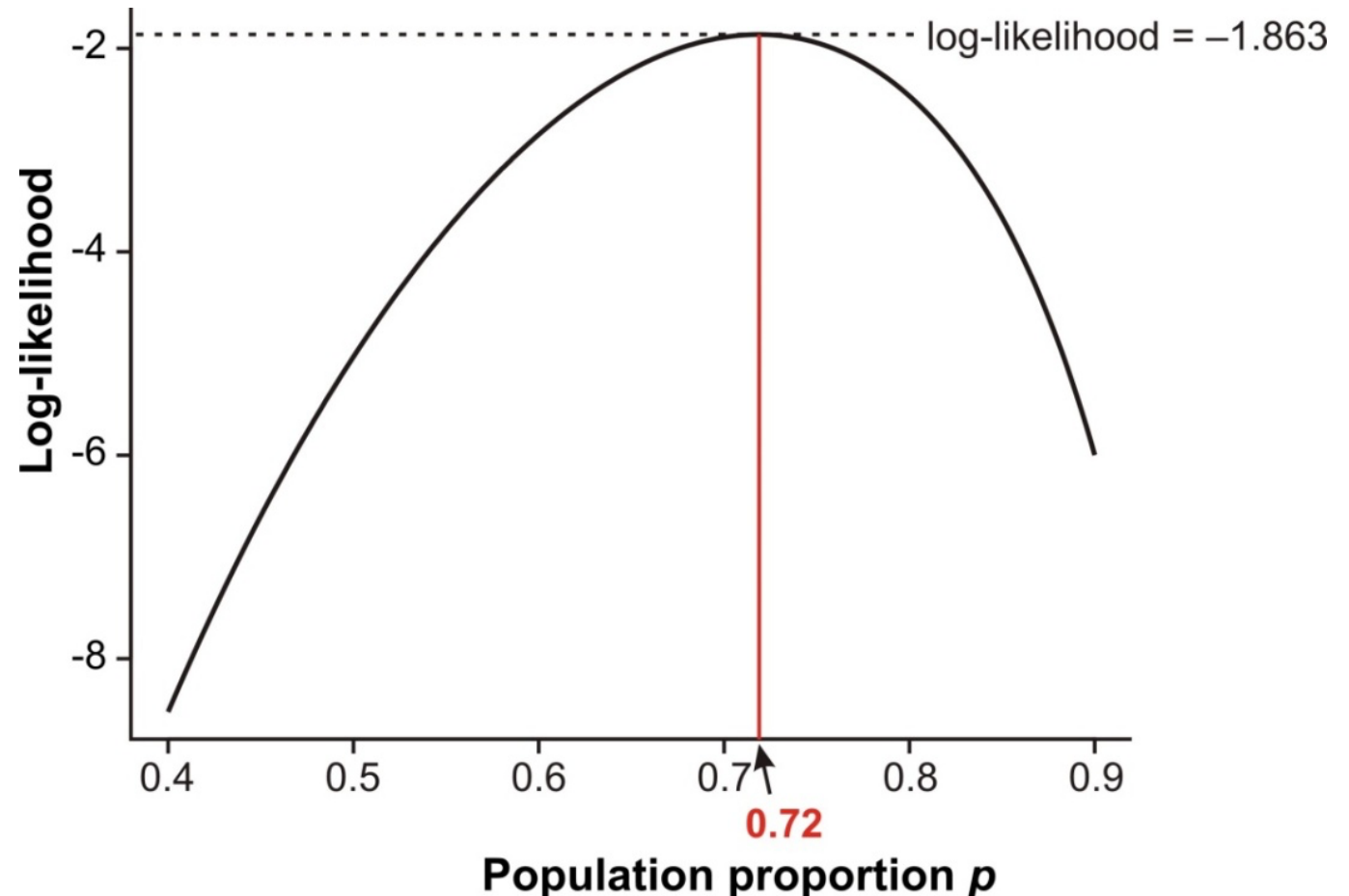
Maximum likelihood estimate

This ML estimate could instead have been obtained more easily as

$$\hat{p} = \frac{Y}{n} = \frac{23}{32} = 0.72$$

The conventional formula for estimating a proportion yields the ML estimate.

Most formulas you use to estimate parameters yield maximum likelihood estimates.
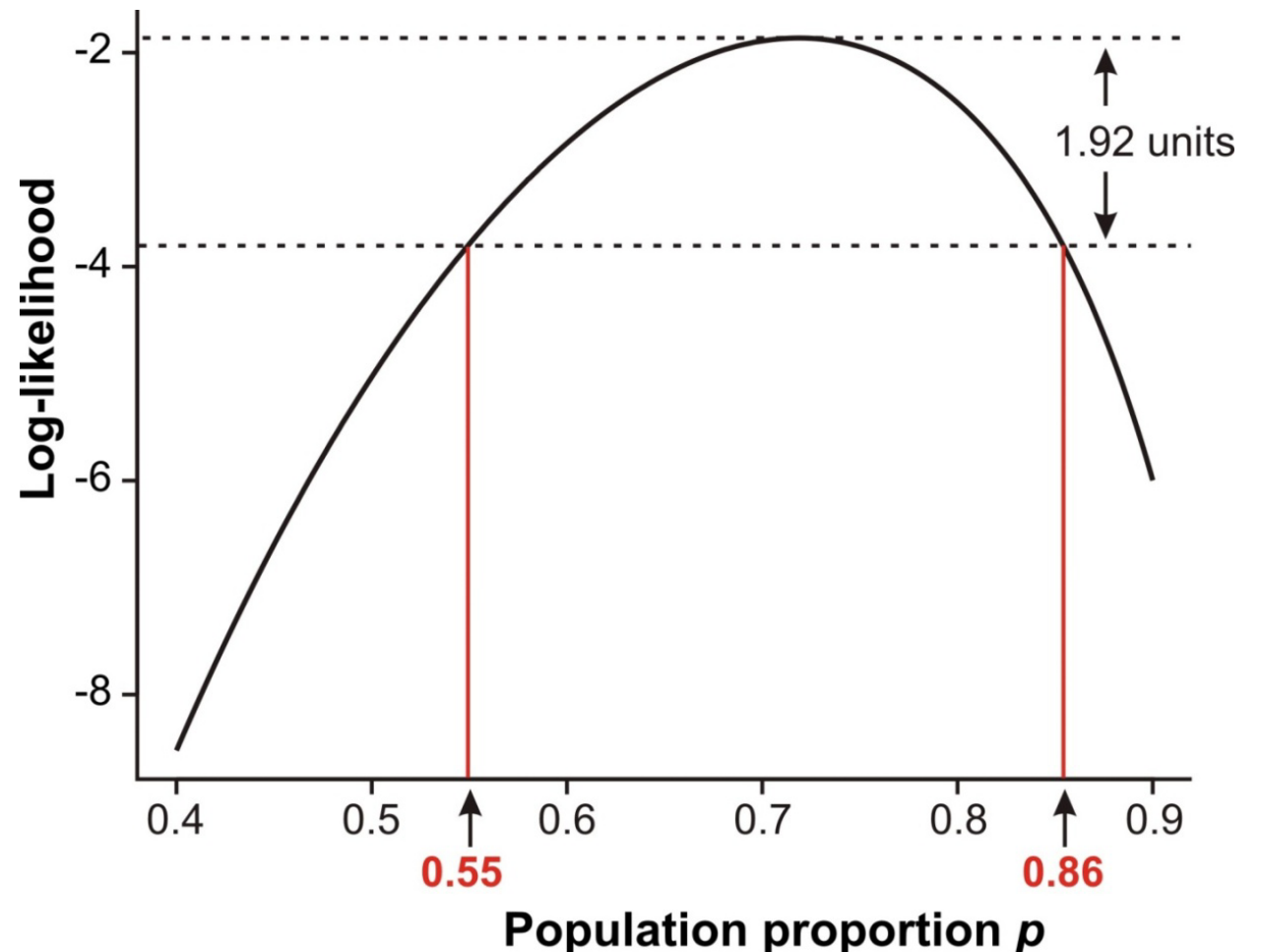
Likelihood-based confidence intervals

When estimating a single parameter, an approximate 95% confidence interval is obtained with the values corresponding to 1.92 log-likelihood units below the maximum.

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

$1.92 = \chi^2_{0.05,1}/2.$

The connection to the χ^2 distribution will become apparent later

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.

Example 2: Survival rates of characters in Game of Thrones

Costumes worn by Ygritte, Jon Snow and Tormund Giantsbane, Oslo 2014.

B. Skinstad. CC3.

https://commons.wikimedia.org/wiki/File:Game_of_Thrones_Oslo_exhibition_2014_-_Ygritte,_Jon_and_Tormund_costumes.jpg
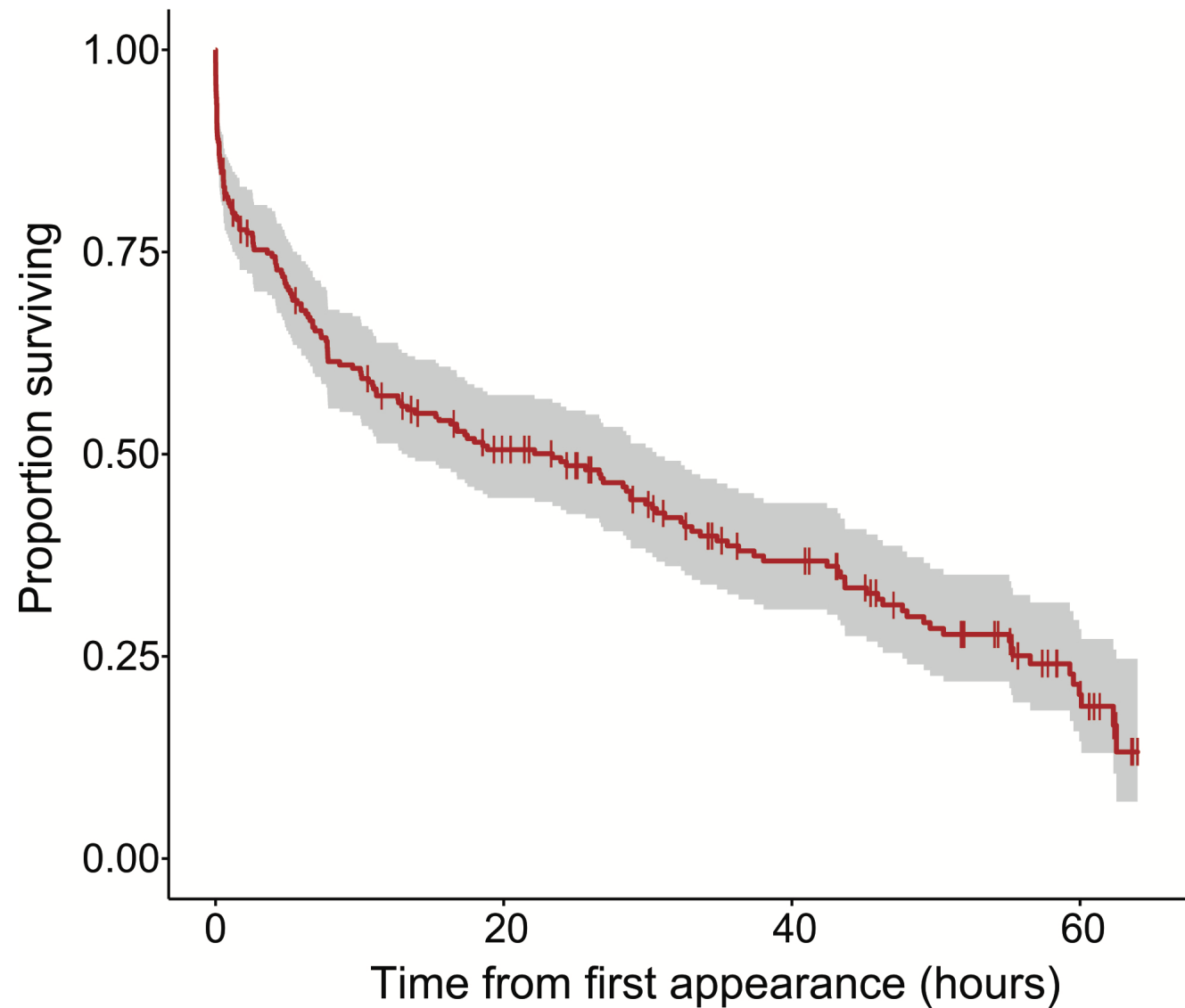
Example 2: Survival rates of characters in Game of Thrones

Death of male characters – causes

```
name         time.hrs                                              death_description
Waymar Royce   0.10                            Killed by a white walker in a forest
Gared Tuttle   0.11                       Decapitation by a white walker in a forest
Will           0.19              Decapitation by Ned Stark (legal execution) in a field
Jon Snow      63.99                                                             <NA>
Bran Stark    63.99                                                             <NA>
Robb Stark    24.34      Shot by arrows and stabbed in the abdomen in the wedding hall at The Twins
Rodrik Cassel 12.70                  Decapitated by Theon Greyjoy in Winterfell castle grounds
Rickon Stark  49.15 Shot with an arrow through the thorax by Ramsay Bolton in field near Winterfell
Eddard Stark   7.67             Decapitated by Ilyn Payne on the steps of the Great Sept of Baelor
Jory Cassel    4.16          Stabbed in the eye by Jamie Lannister in the streets of King's Landing
...
```

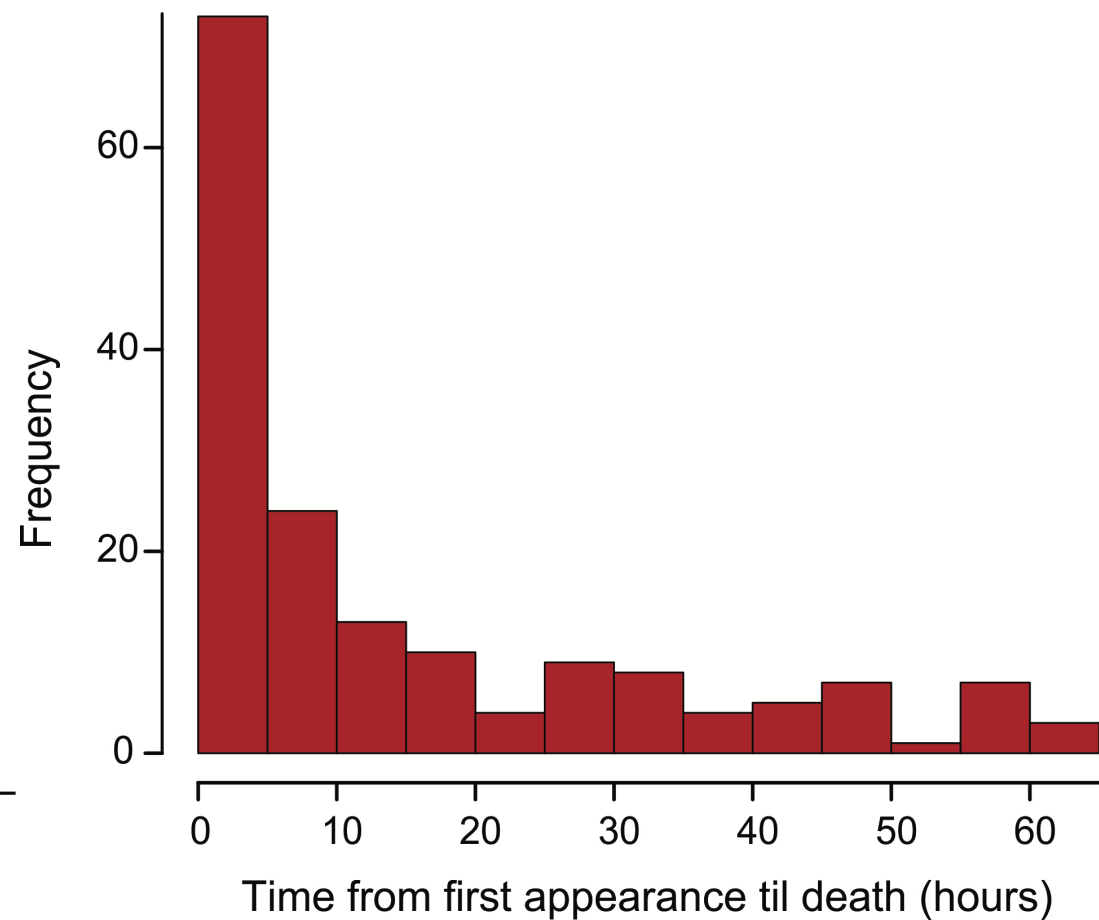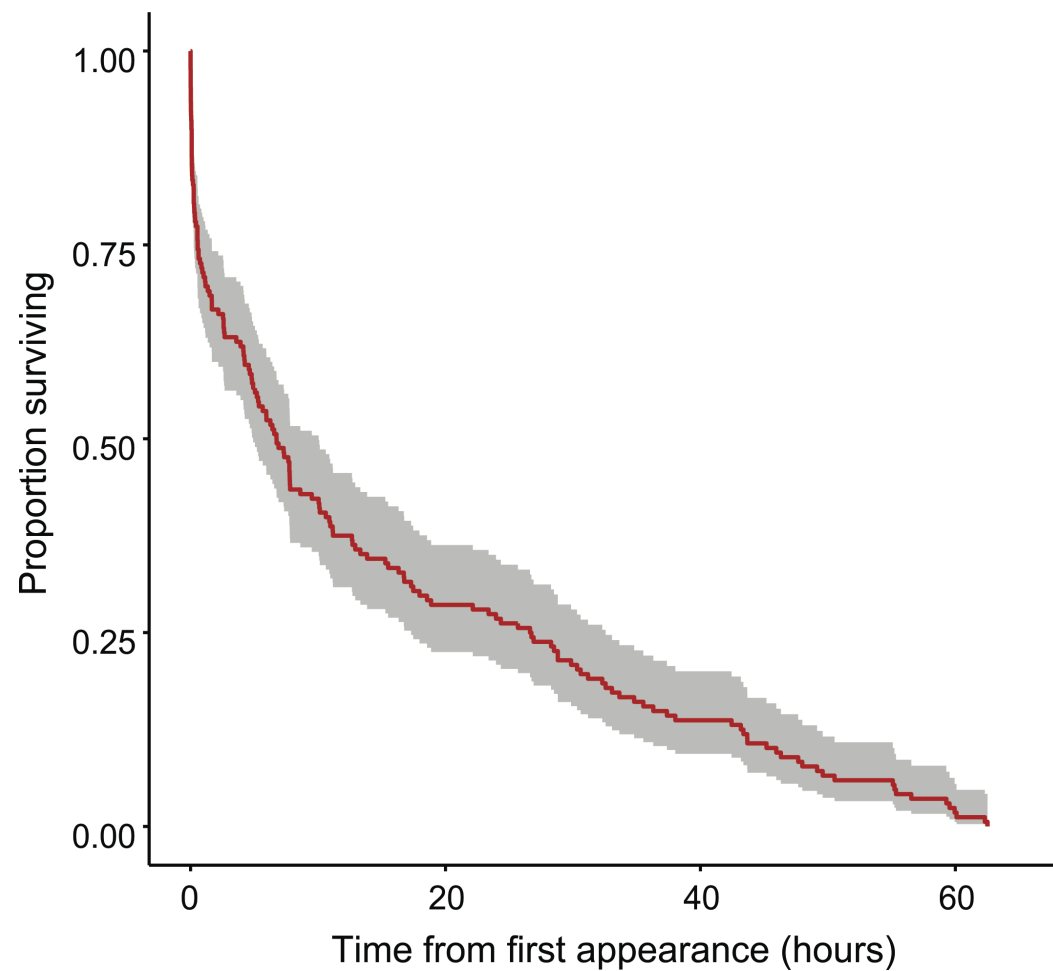Lystad, R. P. and B. T. Brown. 2018. *Injury Epidemiology* 5:

Example 2: Survival rates of characters in Game of Thrones

Survival curve, male characters all 8 seasons

Example 2: Survival rates of characters in Game of Thrones

Times until death of male characters known to have died by the end of 8 seasons

Example 2: Survival rates of characters in Game of Thrones

My goal was to fit a model to these data that would allow me to estimate death rates. I used the Weibull distribution, which is like the exponential distribution but allows fitting a wider diversity of shapes. It has two parameters, λ and k. If the shape parameter $\lambda = 1$, then the death rate is constant over time (i.e., has an exponential distribution).

No simple formula will calculate the ML estimates of λ. What to do?

I found a formula for the Weibull distribution online.

I used R to calculate log-likelihoods for a range of possible values for λ and k to find the maximum likelihood values.

Example 2: Survival rates of characters in Game of Thrones

Probability distribution for the time of death t

$$f(t) = \frac{\lambda}{k}\left(\frac{t}{k}\right)^{\lambda-1} e^{-\left(\frac{t}{k}\right)^{\lambda}}$$

Log likelihood calculation using observed times of death t_i

$$LL(\lambda, k) = \sum_i \log\left(\frac{\lambda}{k}\left(\frac{t_i}{k}\right)^{\lambda-1} e^{-\left(\frac{t_i}{k}\right)^{\lambda}}\right)$$

Using function in R

```
Loglike <- sum(dweibull(x = time.hrs, shape = λ, scale = k, log = TRUE))
```

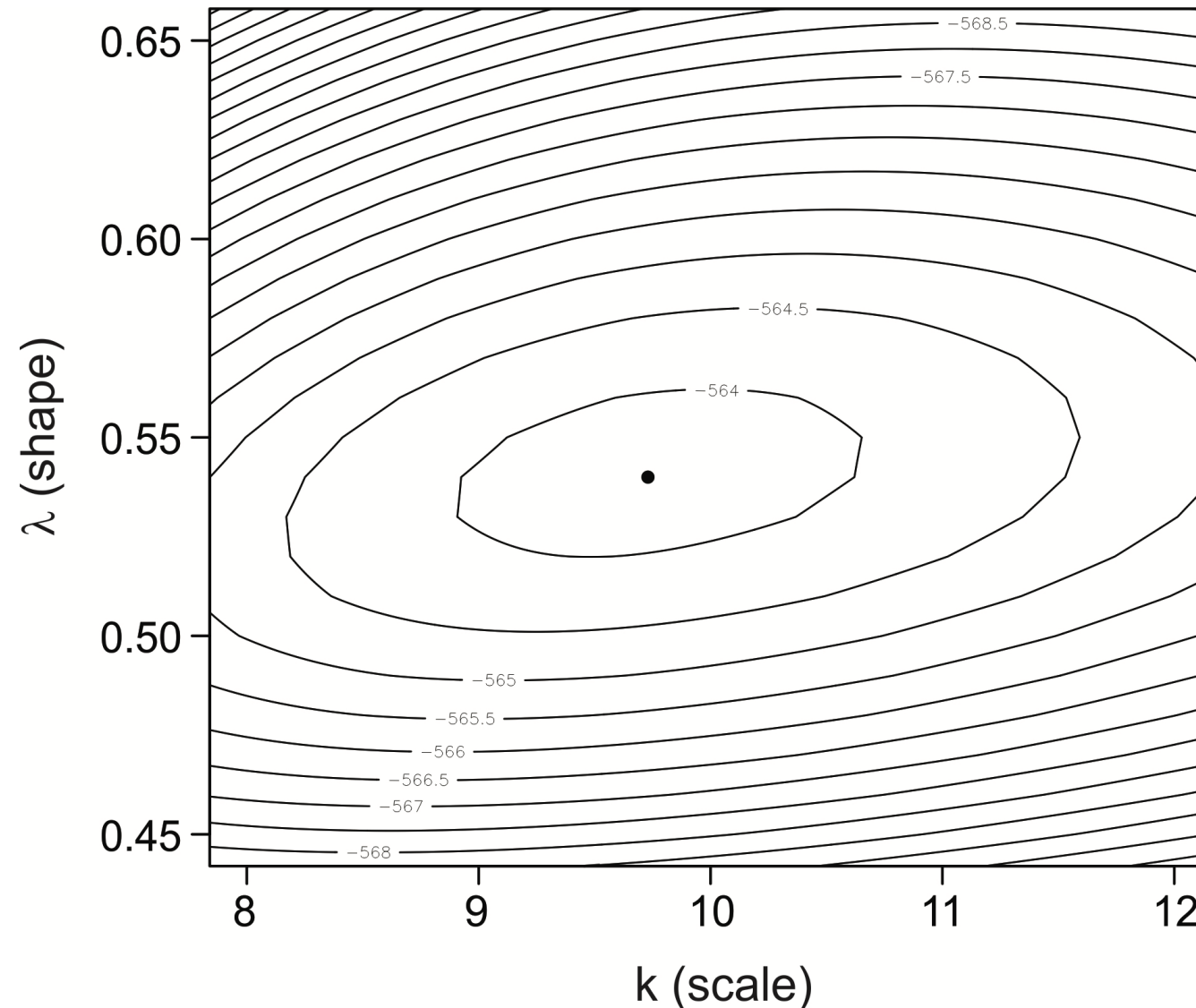# Example 2: Survival rates of characters in Game of Thrones

Since there are two parameters I searched over a grid of the two.

k	λ 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
6	-746.9967	-644.5018	-596.8230	-574.5869	-568.7294	-575.6231	-593.8734	-623.2852	-664.4679	-718.6863
6.1	-747.0068	-644.5142	-596.7977	-574.4575	-568.4026	-574.9750	-592.7432	-621.4655	-661.6929	-714.6161
6.2	-747.0172	-644.5281	-596.7769	-574.3381	-568.0948	-574.3598	-591.6659	-619.7272	-659.0390	-710.7217
6.3	-747.0279	-644.5435	-596.7604	-574.2283	-567.8052	-573.7758	-590.6388	-618.0662	-656.5002	-706.9939
6.4	-747.0388	-644.5603	-596.7481	-574.1276	-567.5329	-573.2215	-589.6595	-616.4787	-654.0705	-703.4243
6.5	-747.0499	-644.5784	-596.7397	-574.0355	-567.2770	-572.6955	-588.7258	-614.9611	-651.7446	-700.0049
6.6	-747.0612	-644.5978	-596.7349	-573.9518	-567.0368	-572.1964	-587.8354	-613.5101	-649.5175	-696.7283
6.7	-747.0727	-644.6183	-596.7338	-573.8759	-566.8116	-571.7230	-586.9863	-612.1224	-647.3844	-693.5875
6.8	-747.0844	-644.6399	-596.7359	-573.8075	-566.6005	-571.2740	-586.1766	-610.7952	-645.3408	-690.5759
6.9	-747.0963	-644.6626	-596.7413	-573.7462	-566.4031	-570.8484	-585.4044	-609.5256	-643.3826	-687.6875
7	-747.1083	-644.6863	-596.7497	-573.6918	-566.2186	-570.4452	-584.6682	-608.3110	-641.5057	-684.9163
7.1	-747.1205	-644.7110	-596.7610	-573.6438	-566.0465	-570.0632	-583.9662	-607.1489	-639.7065	-682.2571
7.2	-747.1329	-644.7366	-596.7751	-573.6021	-565.8863	-569.7017	-583.2969	-606.0370	-637.9815	-679.7045
7.3	-747.1453	-644.7630	-596.7918	-573.5663	-565.7374	-569.3596	-582.6590	-604.9730	-636.3273	-677.2538
7.4	-747.1579	-644.7903	-596.8110	-573.5362	-565.5993	-569.0361	-582.0510	-603.9548	-634.7408	-674.9005
7.5	-747.1706	-644.8183	-596.8327	-573.5116	-565.4715	-568.7306	-581.4717	-602.9805	-633.2191	-672.6402
7.6	-747.1835	-644.8471	-596.8566	-573.4921	-565.3537	-568.4421	-580.9198	-602.0482	-631.7593	-670.4689
7.7	-747.1964	-644.8766	-596.8827	-573.4776	-565.2454	-568.1699	-580.3943	-601.1562	-630.3589	-668.3827
7.8	-747.2094	-644.9068	-596.9109	-573.4678	-565.1463	-567.9135	-579.8939	-600.3026	-629.0152	-666.3779
7.9	-747.2225	-644.9375	-596.9411	-573.4626	-565.0558	-567.6721	-579.4178	-599.4860	-627.7260	-664.4512

Example 2: Survival rates of characters in Game of Thrones

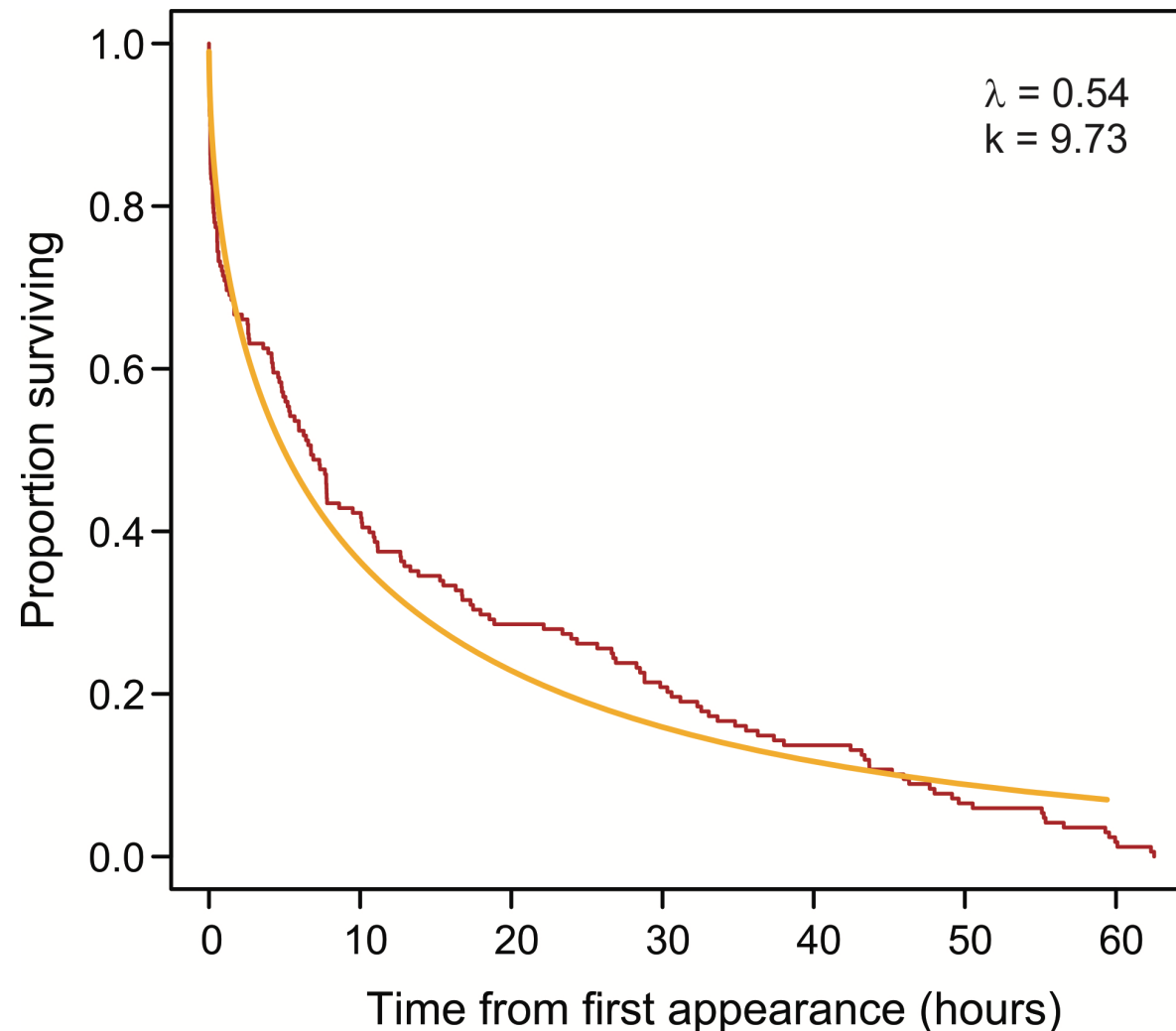The log-likelihood curve is now a surface

Found maximum likelihood estimates: $\hat{\lambda} = 0.54$, $\hat{k} = 9.73$

Example 2: Survival rates of characters in Game of Thrones

Fit to survival curve of male characters who died.

$\lambda < 1$ means that mortality not constant, it declines with time, as though defectives are eliminated quickly, and the remaining men live a bit longer.
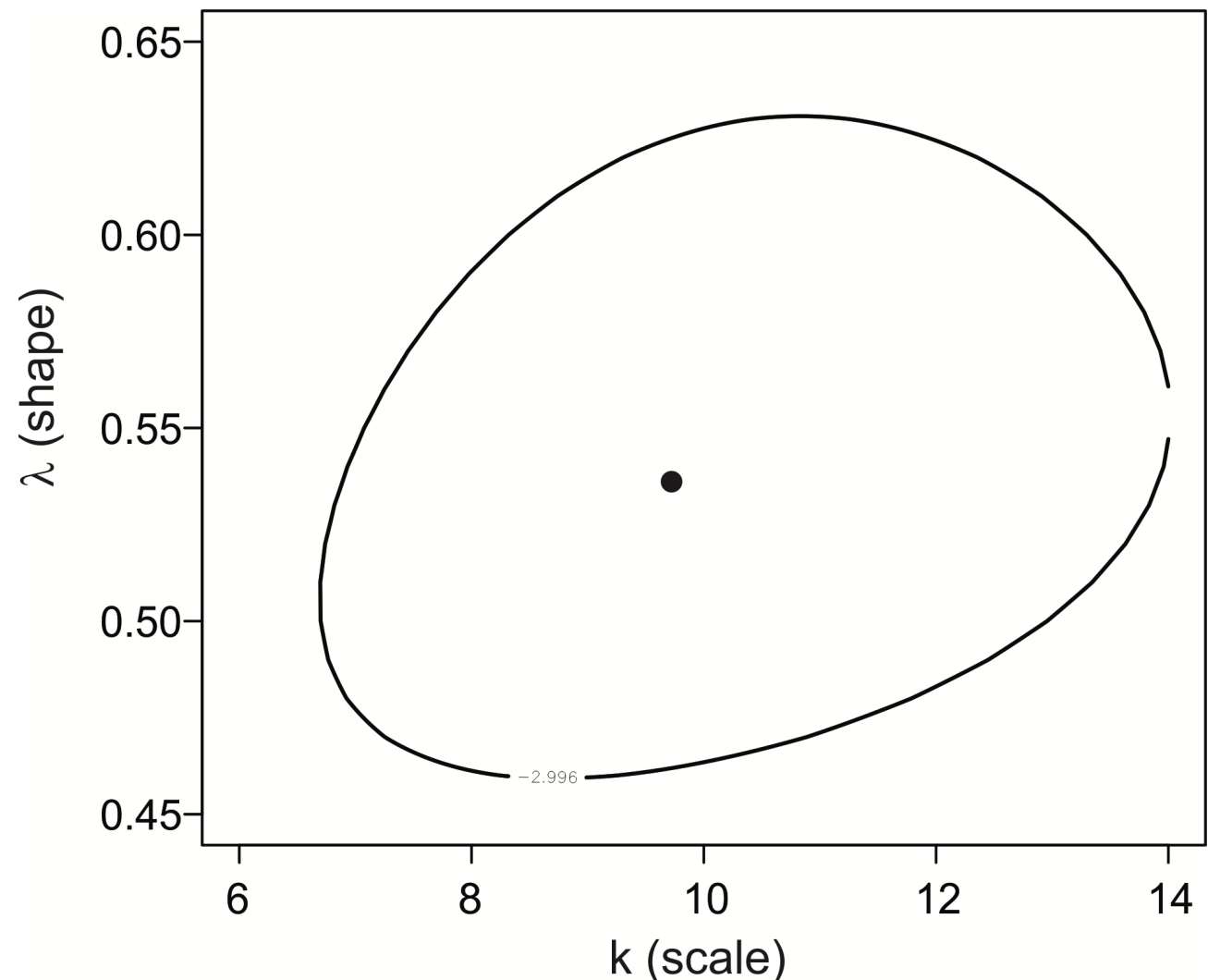

$\lambda = 0.54$
$k = 9.73$

Example 2: Survival rates of characters in Game of Thrones

Likelihood-based 95% confidence region

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

2.996 is $\chi^2_{0.05,2}/2$

You can see that the lower and upper limits of the 95% CI for λ is are 0.46 and 0.62.

Example 2: Survival rates of characters in Game of Thrones

We can also calculate *profile likelihood* function: at each possible value for λ, find the maximum likelihood estimates of k, and plot the value on the profile log-likelihood curve. This can be used to obtain a 95% confidence interval for λ: $0.47 \leq \lambda \leq 0.61$.

The likelihood-based 95% confidence interval is an approximation based on χ^2. It assumes that sample size is large.

R has many tools for calculating maximum likelihood estimates and confidence intervals.

The `survival` package in R can be used to fit models to survival curves that include censored data (data on individuals whose time of death is not known). These lead to similar estimates to the values estimated here.

Log-likelihood ratio test

Likelihood method to compare the fit of two models to data, *full* and *reduced*.

Models must be nested, i.e., one of the models (*reduced model*) must have a subset of the terms present in the other model (*full model*).

The LLRT allows a test of whether the "full model" fits the data statistically significantly better than a "reduced model".

Very general method – applies to any type of data, not necessarily normally distributed.

P-value is approximate, with approximation improving with increasing sample size.

Log-likelihood ratio test

$$G = 2 \ln \frac{L[\text{full model} \mid \text{data}]}{L[\text{reduced model} \mid \text{data}]}$$

G is the log-likelihood ratio test statistic.

Under H$_0$, *G* is approximately χ^2 distributed.

Degrees of freedom for the χ^2 are equal to the difference between the full model and the reduced model in the number of parameters estimated from the data.

Very general method – applies to any data.

The approximation to the χ^2 distribution improves with increasing sample size.
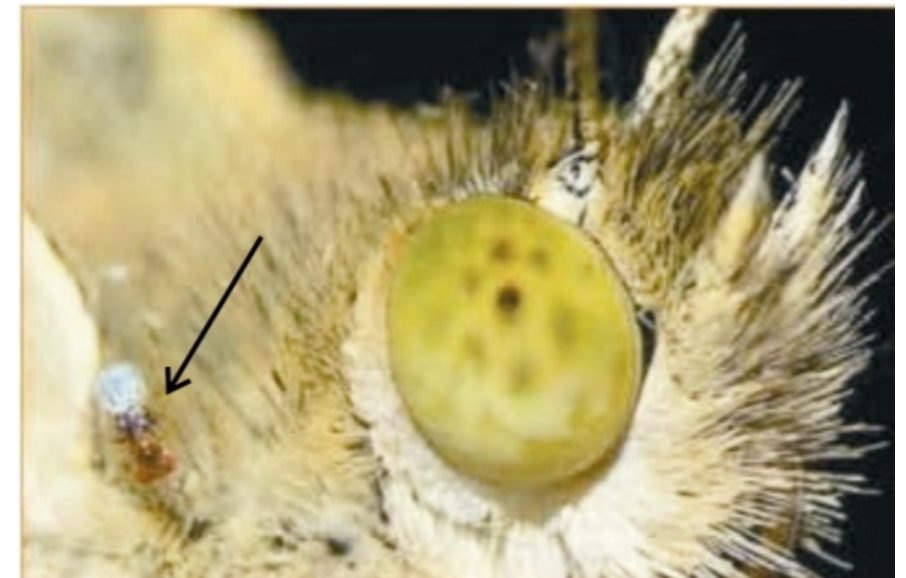
Log-likelihood ratio test

Example 1: 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. Results: 23 successes out of $n = 32$ trials.

Reduced model:

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

Full model:

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

To fit the full model, p is estimated from the data. In this sense, the full model has 1 more term than the reduced model.

Log-likelihood ratio test

$$G = 2 \ln \frac{L[\text{full model} \mid \text{data}]}{L[\text{reduced model} \mid \text{data}]}$$

Applied to the wasp example:

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

The full model uses the maximum likelihood estimate for the parameter (e.g., $\hat{p} = 0.72$ in the full model here).

Log-likelihood ratio test
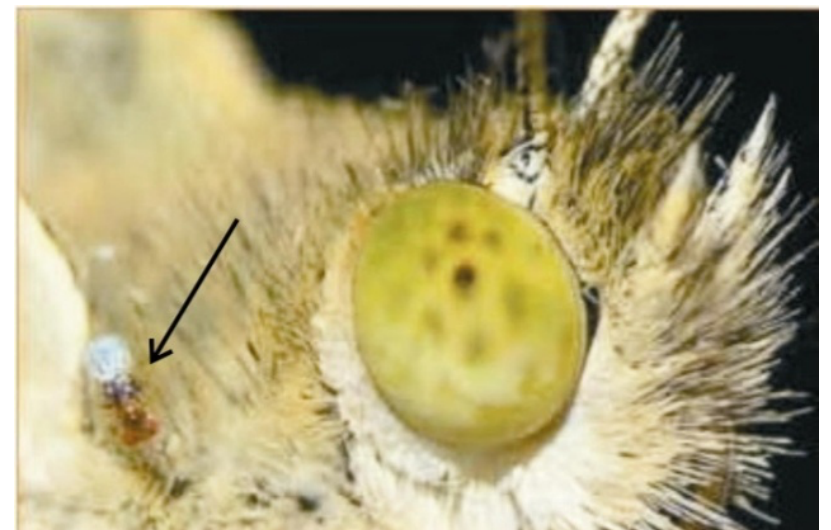
From calculations using formulae shown earlier,

$L[0.72 \mid 23 \text{ of } 32 \text{ chose mated female}] = 0.1553$

$L[0.50 \mid 23 \text{ of } 32 \text{ chose mated female}] = 0.00653$

$$G = 2 \ln \frac{0.1553}{0.00653} = 6.336$$

df = 1, so the critical value χ^2 = 3.841

Since 6.336 > 3.841, we reject H_0.

Wasp example: finale

The chemical that the wasps use to distinguish mated from unmated females is benzyl cyanide, which the male butterfly passes to the female during mating. The compound is an "anti-aphrodisiac", which makes the mated female less attractive to other male butterflies (Fatouros *et al*. 2005).
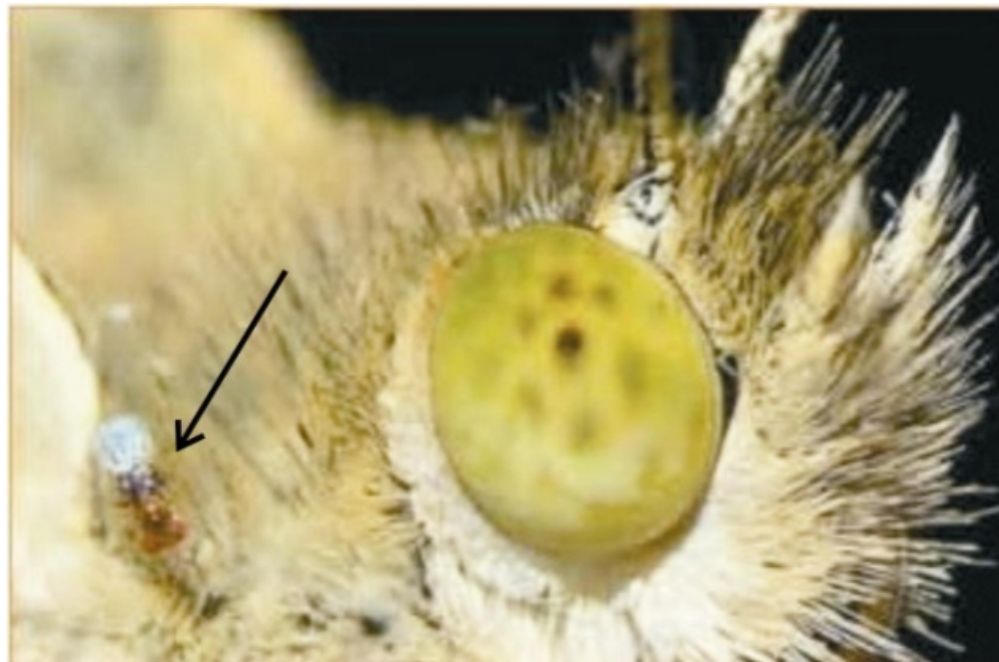
Log-likelihood ratios are used in gene mapping

At each marker along the chromosome two models are fitted to data on healthy and diseased individuals. The "reduced model" assumes that the frequency of healthy and diseased individuals is the same for every genotype. The "full model" assumes that genotype frequency at the marker differs between diseased and non-diseased individuals. The log of the ratio of the likelihoods of the two models (full model likelihood divided by reduced model likelihood) is called the LOD score, and is a measure of the strength of evidence for a causative mutation near that marker.

Evidence for a gene affecting schizo-affective disorder on human chromosome 1.
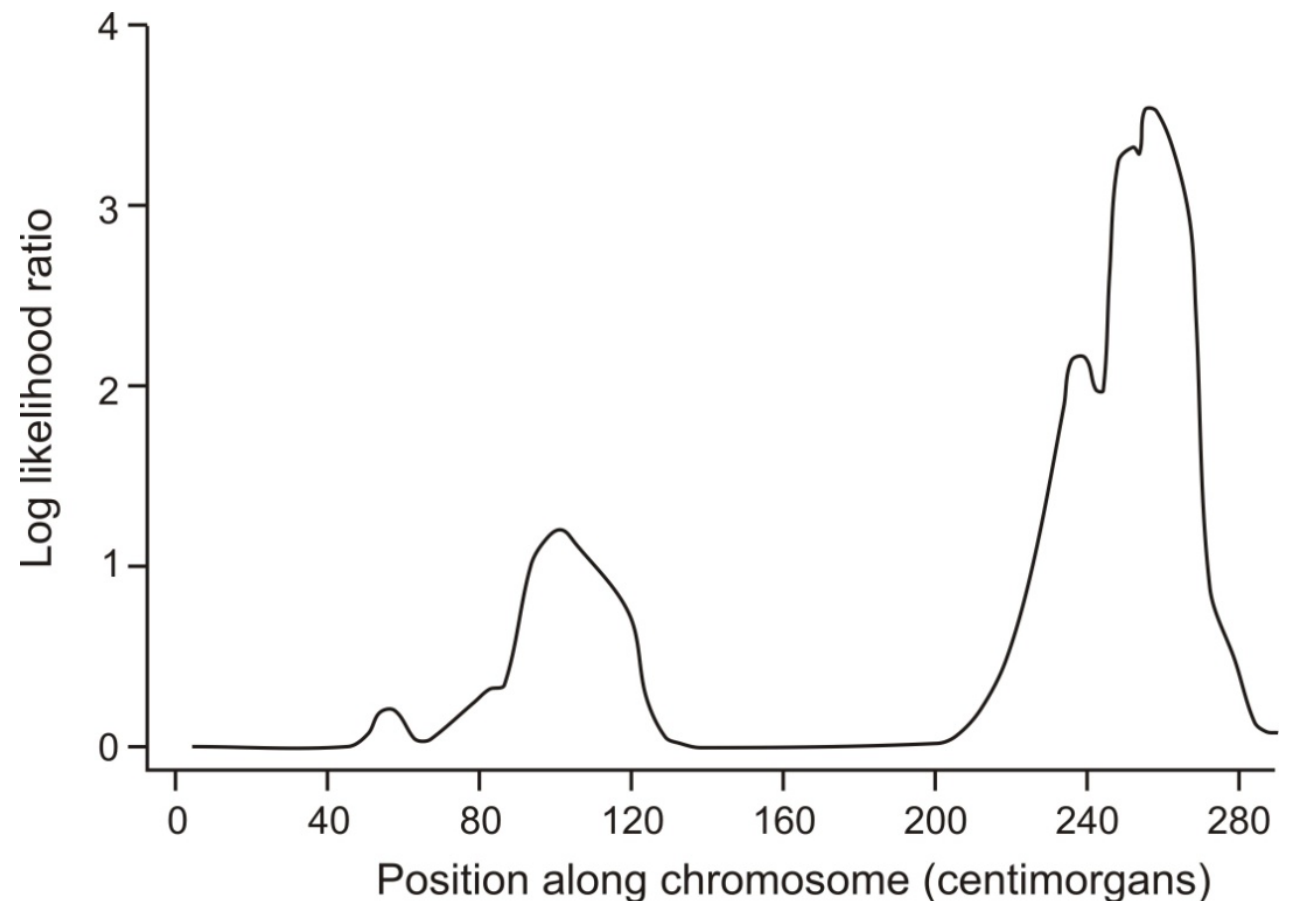
Next discussion paper:

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

Download from "**Handouts**" tab on course web site.

Presenters: Ralitza and Eric

Moderators: Robin & Jacob

Homework assignment 2

Obtain data set from supervisor/lab and fit a linear model in R.

Details on "Homework" page on course web site

Due Friday, March 14 at 5 pm.