

# Likelihood

## Models for Socio-Environmental Data

Chris Che-Castaldo, Mary B. Collins, N. Thompson Hobbs

December 09, 2020



# Why Likelihood?

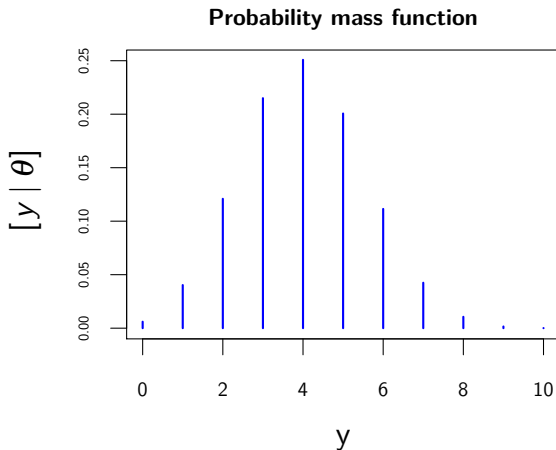
- Maximum likelihood is an important statistical approach.
- Likelihood is a component of all Bayesian models.

# Learning objectives for lecture and lab

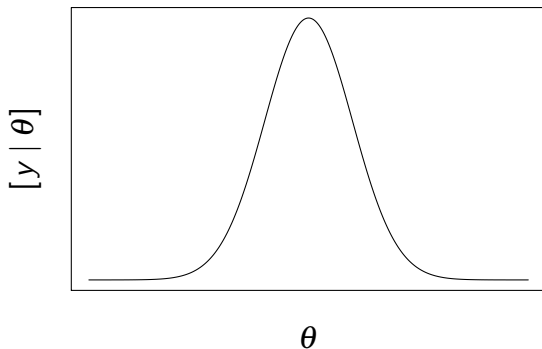
- Understand the concept of likelihood and its relationship to the probability of data conditional on parameters.
- Describe a likelihood profile and how it differs from a probability density function.
- Compose likelihoods for multiple parameters and multiple observations.

## Course progression so far...

$$[y]$$
$$[y \mid \theta]$$
$$y \sim \text{binomial}(\theta, 10)$$



## Inference from likelihood is based on $[y | \theta]$



- Likelihood allows us to compare alternative parameter values by calculating the probability (or probability density) of the data conditional on parameters  $[y | \theta]$ .
- As you will see, all evidence based on likelihood is relative.

# The key idea in likelihood

- In a probability mass or density function, the parameter  $\theta$  is constant (fixed) and the outcomes  $\mathbf{y}$  vary (these outcomes represent data we may observe). The function sums or integrates to 1 over its support.
- In a likelihood function, the data  $\mathbf{y}$  are constant (fixed) and the parameter  $\theta$  varies. We use  $[y | \theta]$  to assess the likelihood of different values of  $\theta$  in light of the data. In this case, the function does not sum or integrate to one over all possible values of the parameter.

$$\underbrace{L(\theta | y)}_{\text{likelihood function}} \propto \underbrace{[y | \theta]}_{\text{PDF or PMF}}$$

Likelihood is *proportional* to probability or probability density

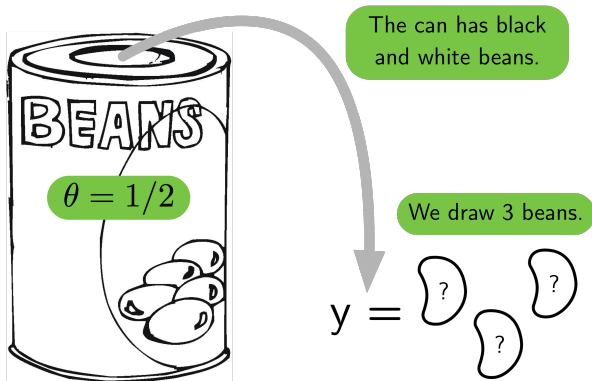
## Discuss notation

$$L(\theta | y) \propto [y | \theta] \quad (1)$$

$$L(\theta | y) = c[y | \theta] \quad (2)$$

$$L(\theta | y) = [y | \theta] \quad (3)$$

# The parameter is known and the data are random



- What are the possible outcomes?
- What probability mass function would you use to model these data?
- What is the probability of each outcome?
- What is the sum of the individual probabilities?



## The parameter is known and the data are random

We draw three beans from a can with equal numbers of white and black beans. The possible outcomes are:

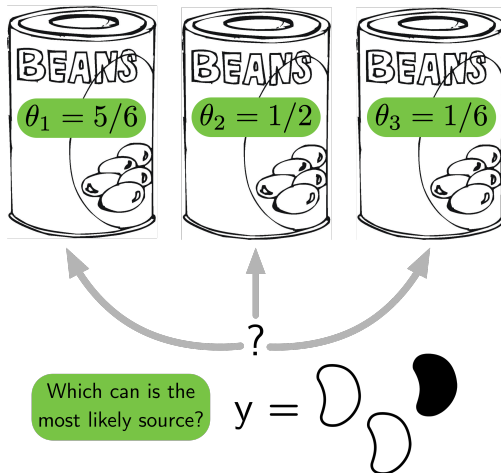
$\theta$	number of white beans, $y_i$	$[y_i \mid \theta = ?]$
	$\sum_{i=1}^4 [y_i \mid \theta = ?]$	?

## The parameter is known and the data are random

We draw three beans from a can with equal numbers of white and black beans. The possible outcomes are:

$\theta$	number of white beans, $y_i$	$[y_i \mid \theta = .5]$
.5	0	.125
.5	1	.375
.5	2	.375
.5	3	.125
	$\sum_{i=1}^4 [y_i \mid \theta = .5]$	1

The data are known and the parameter is random



- What is the likelihood of each parameter value?

## The data are known and the parameter is random

We have three hypothesized parameter values,  $(5/6, 1/2, 1/6)$ . Data in hand are 2 white beans on three draws. The likelihood of each parameter value is:

hypothesis, $\theta_i$	number of white beans, $y$	$[y = ? \mid \theta_i]$
	$\sum_{i=1}^3 [y = ? \mid \theta_i]$	?

## The data are known and the parameter is random

We have three hypothesized parameter values,  $(5/6, 1/2, 1/6)$ . Data in hand are 2 white beans on three draws. The likelihood of each parameter value is:

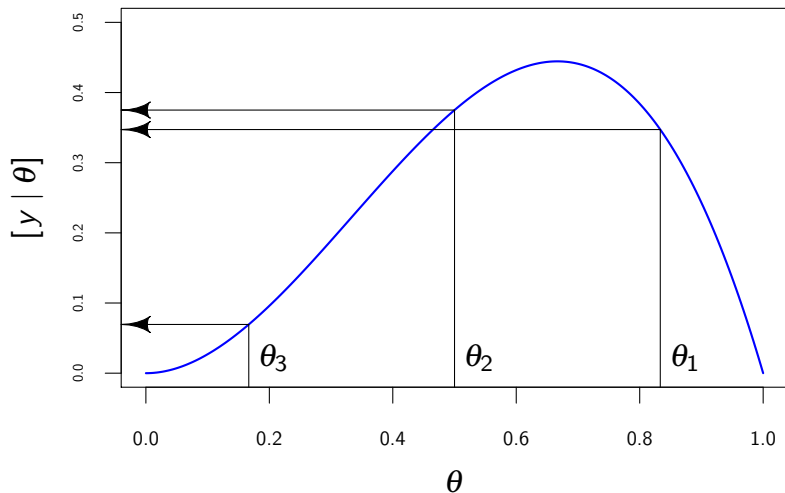
hypothesis, $\theta_i$	number of white beans, $y$	$[y = 2 \mid \theta_i]$
$\theta_1 = 5/6$	2	0.35
$\theta_2 = 1/2$	2	0.38
$\theta_3 = 1/6$	2	0.069
	$\sum_{i=1}^3 [y = 2 \mid \theta_i]$	0.79

## The data are known and the parameter is random (rescaled)

We have three hypothesized parameter values,  $(5/6, 1/2, 1/6)$ . Data in hand are 2 white beans on three draws. The likelihood of each parameter value is:

hypothesis, $\theta_i$	number of white beans, $y$	$[y = 2 \mid \theta_i]$
$\theta_1 = 5/6$	2	$0.35 / 0.38 = 0.92$
$\theta_2 = 1/2$	2	$0.38 / 0.38 = 1.0$
$\theta_3 = 1/6$	2	$0.069 / 0.38 = 0.18$
	$\sum_{i=1}^3 [y = 2 \mid \theta_i]$	2.1

## A likelihood profile: 2 white beans on 3 draws



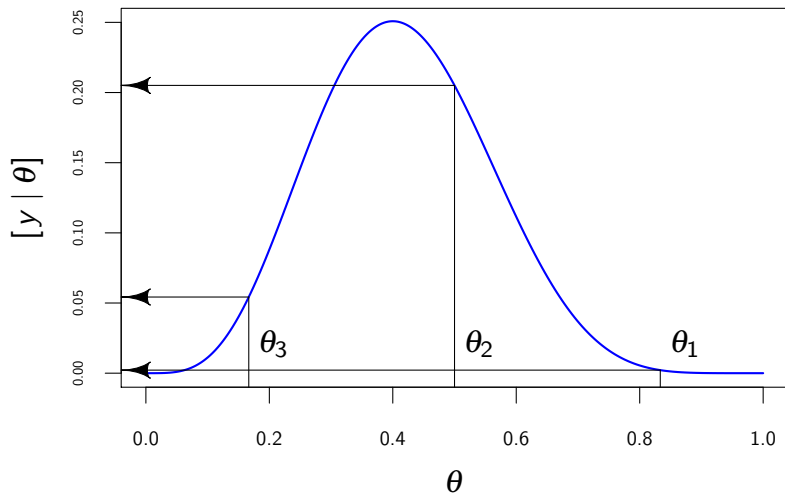
## Class exercise: Can of beans

An adventurous person takes a draw of 10 beans from one of the cans where the identity of the can is unknown. Of the 10 beans drawn from the mystery can, 4 are white.

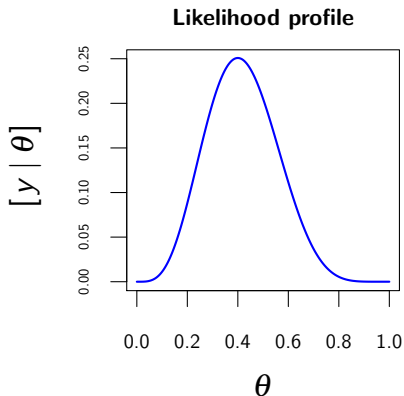
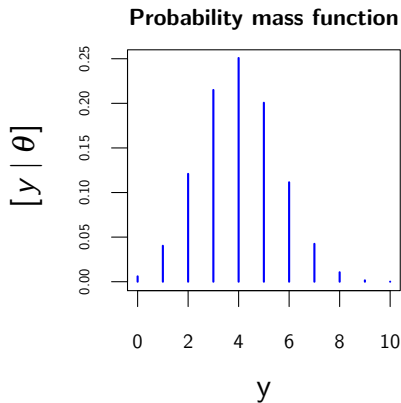
- Which of the three cans is the most likely to have produced this draw?
- How much more likely is this can than the other two?



## A likelihood profile: 4 white beans on 10 draws



## Likelihood vs Probability:



$$[y | \theta] = \binom{10}{y} \theta^y (1 - \theta)^{n-y}$$

## How do we fit models with multiple parameters?

- In the example we had a single parameter,  $\theta$ , one set of observation, 4 successes on 10 draws, and a binomial likelihood.
- However, we could have made the likelihood a function of the *predictions* of a model, and used any probability mass function or probability density function as a “wrapper” for the predictions, i.e.,

$$\begin{aligned}\mu_i &= g(\theta, x_i) \\ L(\theta, \sigma^2 \mid y_i) &\propto \underbrace{[y_i \mid \mu_i, \sigma^2]}_{\text{PDF or PMF}}\end{aligned}$$

# Likelihood Surfaces

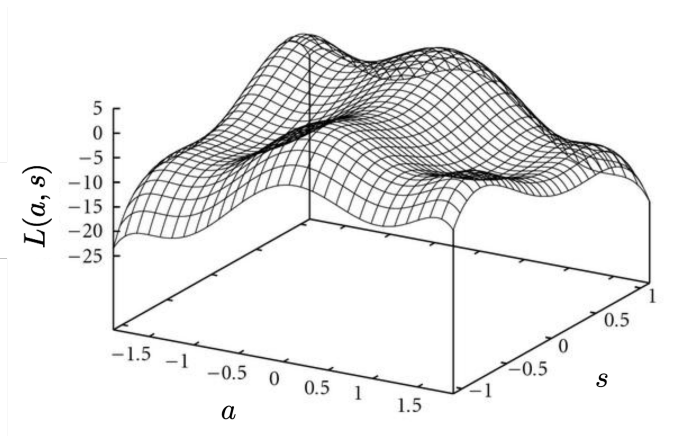


Figure courtesy of [Sergiy Nesterko](#).

## How do we fit models with multiple parameters and multiple data points?

The total likelihood is the product of the individual likelihoods, assuming the data are *conditionally independent*:

$$L(\boldsymbol{\mu}, \sigma^2 \mid \mathbf{y}) = c \prod_{i=1}^n [y_i \mid g(\boldsymbol{\theta}, x_i), \sigma^2]$$

## What does conditionally independent mean?

Independence is an assumption! Remember from the chain rule:

$$Pr(y_1, \dots, y_n) = Pr(y_1 | y_2 \dots y_n) Pr(y_2 | y_3 \dots y_n) \dots Pr(y_n).$$

However, by assuming that these random variables are independent, you can simplify the joint probability into:

$$Pr(y_1, \dots, y_n) = Pr(y_1) Pr(y_2) \dots Pr(y_n),$$

such that the total likelihood a product of the individual likelihoods.

## Log likelihoods:

We often use the sum of the log likelihoods to get the total log likelihood as a basis for fitting models:

$$\log(L(\boldsymbol{\theta}, \sigma^2 \mid y)) = \log(c) + \sum_{i=1}^n \log([y_i \mid g(\boldsymbol{\theta}, x_i), \sigma^2])$$

## Class Exercise: Likelihood profile of $\lambda$ for tadpole counts

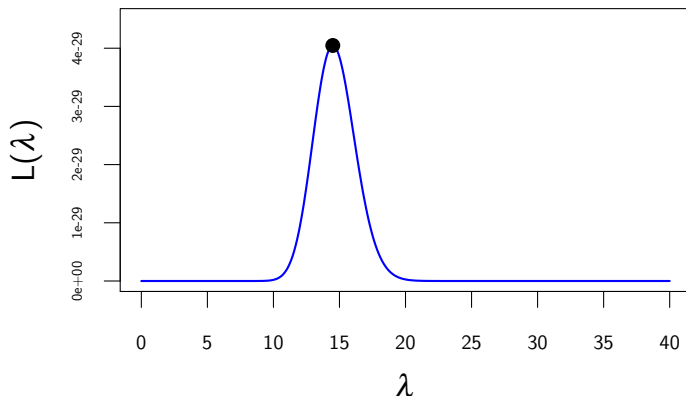
Assume we have observed the following counts of tadpoles in funnel traps from 6 different ponds: 7, 11, 54, 12, 2, 1 individuals.

Assuming these counts were generated using a Poisson process governed by  $\lambda$ , compute a likelihood profile using R.

Eyeball this profile to determine the MLE.



## Likelihood profile of $\lambda$ for tadpole counts

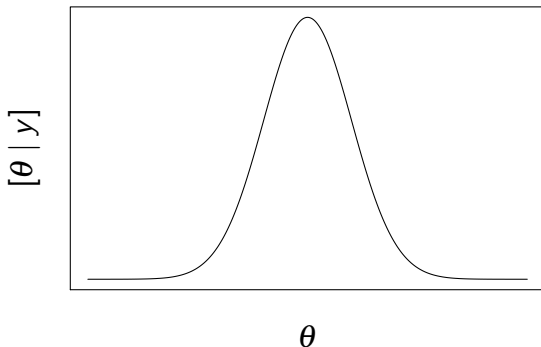


```
lambda <- seq(0, 40, length = 1000)
y <- NA
for(i in 1:length(lambda)) {y[i] <- prod(dpois(c(7, 11, 54, 12, 2, 1), lambda[i]))}
```

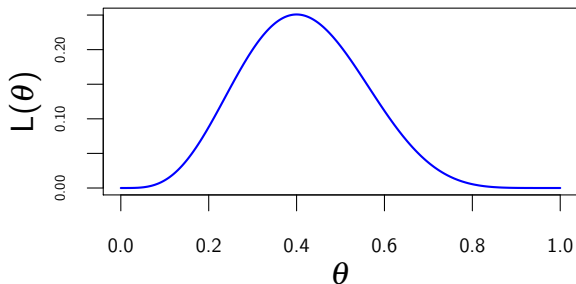
# Things to remember

- Likelihood allows us to evaluate the relative strength of evidence for one parameter or model relative to another.
- Likelihood is a component of all Bayesian models.
- The data are fixed and the parameters are variable in likelihood functions. These functions do not integrate or sum to one over the range of values of the parameter.
- The outcomes are variable and the parameters are fixed in probability mass functions and density functions. These functions sum or integrate to one over the support of the random variable.

Looking ahead: Bayesian inference is based on  $[\theta | y]$



## Looking ahead: Getting to $[\theta | y]$ from a likelihood



- What must be done to insure that the area under the curve = 1?
- If you do this, is this now a probability density function for  $\theta$ ?

## What does conditionally independent mean?

We evaluate the independence assumption by examining the residuals ( $\varepsilon$ ) from a model, where ( $\varepsilon_i = y_i - g(\theta, x_i)$ ).

The independence assumption holds if knowing a residual tells you nothing about the other residuals.

We assess this by ensuring that the residuals:

- do not show a trend, meaning they should be centered on 0 throughout the range of fitted values,
- and are not be autocorrelated.

## Likelihood ratio confidence intervals

Find the upper and lower bounds of an interval where all  $\lambda$  values within that interval are as consistent with the data as  $\lambda_{MLE}$ .

We compute the likelihood ratio statistic:

$$R = 2 \log \left( \frac{L(\lambda_{MLE} | y)}{L(\lambda_0 | y)} \right) \sim \chi_{k=1}^2$$

which is distributed  $\chi^2$  with 1 degree of freedom. Note that we fail to reject  $H_0$  that  $\lambda = \lambda_0$  if  $R < \chi_{k=1}^2(1 - \alpha)$ .

## Likelihood ratio confidence intervals

We determine the  $(1 - \alpha = 0.95)$  likelihood ratio confidence interval by finding the upper and lower bounds for all values of  $\lambda_0$  where we would fail to reject  $H_0$ .

$$2 \log \left( \frac{L(\lambda_{MLE} | y)}{L(\lambda | y)} \right) < \chi_{k=1}^2(0.95)$$

$$L(\lambda_{MLE}) - \frac{3.84}{2} < L(\lambda | y)$$

$$L(\lambda | y) > L(\lambda_{MLE}) - 1.92$$

# Likelihood ratio confidence intervals

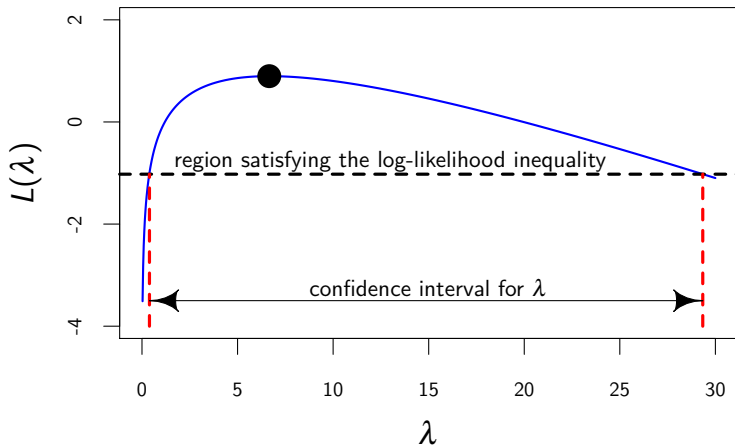
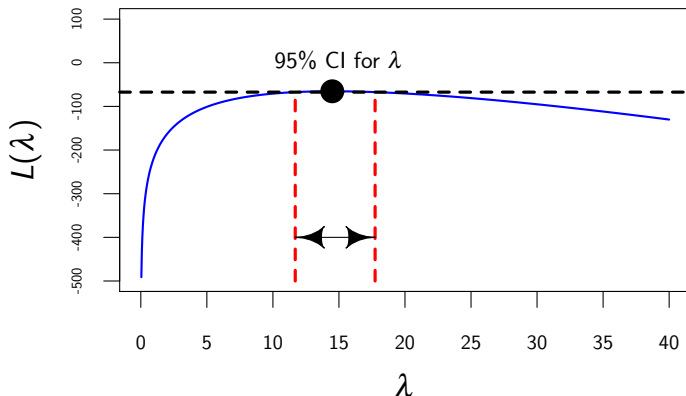


Figure courtesy of the [UNC Biology Department](#).



## Likelihood profile of $\lambda$ for tadpoles



```
lambda <- seq(0, 40, length = 1000)
y <- NA
for(i in 1:length(lambda)) {y[i] <- log(prod(dpois(c(7, 11, 54, 12, 2, 1), lambda[i])))}
```