Likelihood

ESS 575 Models for Ecological Data

N. Thompson Hobbs

February 12, 2019

Why likelihood?

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

Learning objectives for lecture and lab

- ▶ Understand the concept of likelihood and its relationship to the probability of the data conditional on the parameters.
- ▶ Describe a likelihood profile and how it differs from the plot of a probability density function.
- ▶ Be able to use single and multiple observations to obtain maximum likelihood estimates of single and multiple parameters.
- ▶ Be able to combine data with prior information in the likelihood framework.

$[y \mid \boldsymbol{\theta}]$

Prevalence is a term used in disease ecology to indicate the proportion of a population that is infected. Prevalence of chronic wasting disease in male mule deer on winter range north of Fort Collins average 12%. A sample of 24 male deer includes 4 infected individuals. What is the probability of obtaining these data if the estimate of prevalence is true?
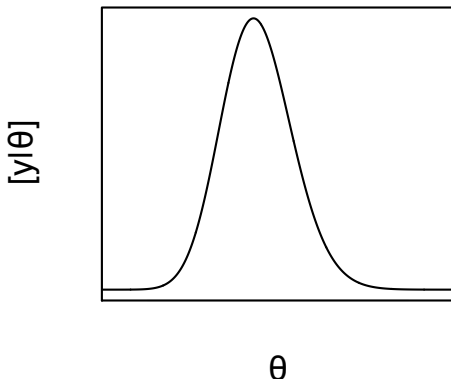
$[\theta \mid y]$

We obtain a sample of 24 male deer includes 4 infected individuals from deer winter range north of Fort Collins. In light of these data, what is the probability that the true value of prevalence, ψ, is found in $q_L \leq \psi \leq q_U$?

Bayesian inference is based on $[\boldsymbol{\theta}|y]$

$[\theta\,|\,y]$

θ
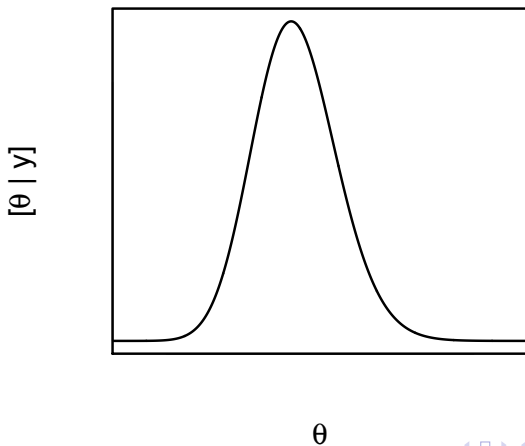
Inference from likelihood is based on $[y \mid \theta]$

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

The key idea in likelihood

- In a probability mass or probability density function, the parameter θ is constant (known) and the data y are random variables. The function sums or integrates to 1 over its support.

- In a likelihood function, the data are constant (known) and the parameter is unknown but fixed. We use $[y \mid \theta]$ to assess the likelihood of different values of θ in light of the data. In this case, the function does not sum or integrate to one over all possible values of the parameter.

$$\underbrace{L(\theta|y)}_{\text{likelihood function}} \quad \propto \quad \underbrace{[y|\theta]}_{\text{PDF or PMF}} \tag{1}$$

Likelihood is *proportional* to probability or probability density.

Discuss notation

$$L(\boldsymbol{\theta} \mid y) \propto [y \mid \boldsymbol{\theta}] \tag{2}$$
$$L(\boldsymbol{\theta} \mid y) = c[y \mid \boldsymbol{\theta}] \tag{3}$$
$$L(\boldsymbol{\theta} \mid y) = [y \mid \boldsymbol{\theta}] \tag{4}$$

Intuition for likelihood

Cut to beans in cans.

The parameter is fixed.

We *know* the parameter value, $\theta = .5$. We make three draws. What are the possible outcomes for the number of white beans? What probability mass function would you use to model theses data? What is the probability of each outcome? What is the sum of the individual probabilities?

The parameter is fixed and the data vary.

We make three draws.

| θ	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 fixed and the parameter varies

We make a single draw of three beans and obtain two whites. We seek evidence in the data for three hypothesized parameter values, $\theta = \frac{1}{6}$, $\theta = \frac{1}{2}$, and $\theta = \frac{5}{6}$. Compute the likelihoods for these parameter values and sum them.

The data are fixed and the parameter varies.

We have three hypothesized parameter values. Data in hand are 2 whites on three draws. What is the likelihood of each parameter value?

Hypothesis, θ_i	$[y \mid \theta_i]$
$\theta_1 = 5/6$	.347
$\theta_2 = 1/2$	.375
$\theta_3 = 1/6$	.069
$\sum_{i=1}^{3}[y \mid \theta_i]$	.791

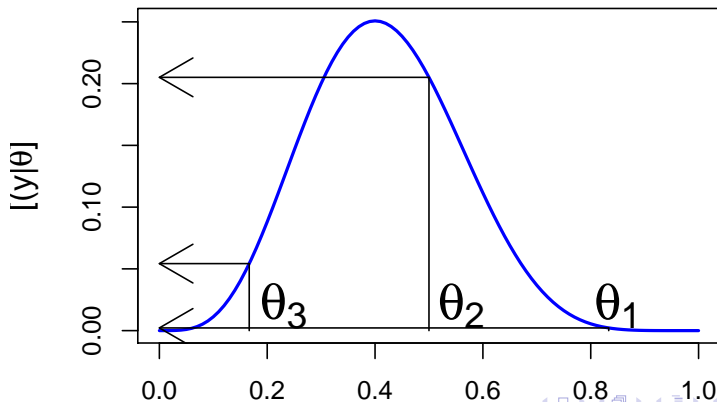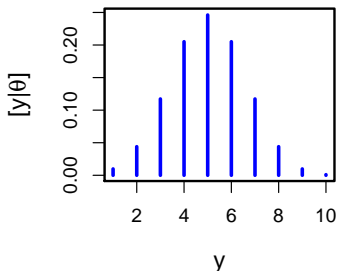## A likelihood profile:

Pr(2 white on 3 draws|θ)
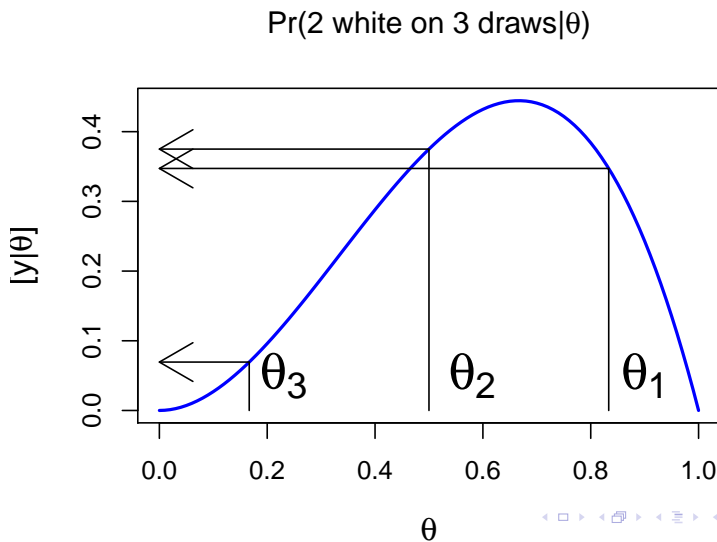
A likelihood profile:

We now draw 10 beans and obtain 4 whites.

Pr(4 white on 10 draws|θ)

Likelihood vs Probability:
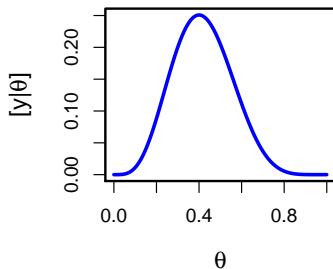
The *points* on each graph are probabilities. The graph on the left is a true PMF. The graph on the right is not.

Exercise

We fix $\theta = .5$ and vary the data in the probability mass function. We fix the data at two whites on three draws and vary θ in the likelihood profile. There is a single point on each graph with the exact same value. What is that point?

How do we fit models with multiple parameters to multiple data points ?

In the example we had a single parameter, θ, one set of observation, 4 successes on ten 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.,

$$
\begin{align}
\mu_i &= g(\boldsymbol{\theta}, x_i) \tag{5} \\
L(\mu_i | y_i) &\propto \underbrace{[y_i | \mu_i, \boldsymbol{\sigma}^2]}_{\text{PDF or PMF}} \tag{6}
\end{align}
$$

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

$$
L(\boldsymbol{\theta} \mid y) = c \prod_{i=1}^{n} [y_i \mid g(\boldsymbol{\theta}, x_i), \boldsymbol{\sigma}^2] \tag{7}
$$

What does conditionally independent mean?

▶ The data are independent *conditional on the value of the model's parameters*. What this means is that the *residuals*, i.e., $\varepsilon_i = y_i - g(\boldsymbol{\theta}, x_i)$ do not show any trend. They should be centered on 0 throughout the range of fitted values.

▶ The residuals should not be correlated with each other; that is, they must not be *autocorrelated*. More about this later in the semester when we will learn how to model spatial and temporal structure in the residuals.

Log likelihoods:

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

$$\log(L(\boldsymbol{\theta}|y)) = \log(c) + \sum_{i=1}^{n} \log([y_i \mid g(\boldsymbol{\theta}, x_i), \boldsymbol{\sigma}^2]) \qquad (8)$$

The exponential distribution

$$y_i \sim \text{exponential}(\lambda)$$
$$P(y_i \mid \lambda) = \lambda e^{-\lambda y_i}$$

- Data: y_i, "Waiting times" or interval of space for an event to happen in a Poisson process. The number of events per interval is given by the Poisson, the time between events is given by the exponential.
- How do the data arise? Times between sightings of a species, lifespan, random samples of anything that decreases exponentially with time or distance, distances between mutations on a strand of DNA.
- Parameter: λ = the average rate of occurrence events per time or space

Moments: mean= $1/\lambda$, variance= $1/\lambda^2$

- R functions
  - `dexp(x=y_i, rate=lambda)` returns the probability of y_i conditional on the value given for rate. y can be a scalar or vector.
  - `rexp(n=, rate=lamb)` returns a vector of length n of random draws from a exponential distribution with shape parameter `rate = `.
  - Also see `qexp()` and `pexp()` in R help.

Maximum likelihood estimate of λ

We are studying metapopulations of frogs in small ponds. We
assume extinctions occur independently influenced by a suite of
variables (i.e., pond size and juxtaposition, and the other usual
suspects). We start with a sample of ponds containing frogs and
monitor them daily. When the frogs are not longer found in the
pond, we note the time this occurs. We want to estimate the
average time required for a pond to go extinct.

Maximum likelihood estimate of λ, one data point:

$y_i = 10 \, \text{days}$
$L(\lambda | y_i) = [y_i | \lambda]$
$[y_i | \lambda] = \lambda \, e^{-\lambda \, y_i}$

Take the log of each side to make
the expression easier to differentiate:
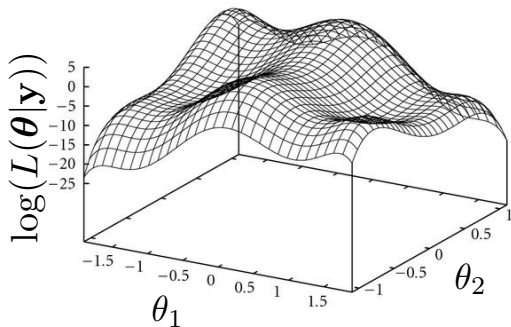
$\log([y|\lambda]) = \log(\lambda) - \lambda \, y_i$
$\frac{d \log([y|\lambda])}{d\lambda} = \frac{1}{\lambda} - y_i$

Set $= 0$ and solve for λ :

$\lambda_{\text{mle}} = \frac{1}{y_i} = \frac{1}{10}$

Maximum likelihood estimate of λ, two data points:

$\mathbf{y} = [y_i, y_2] = 10\,\text{days}, 18\,\text{days}$
$L(\lambda|\mathbf{y}) = [y_1|\lambda][y_2|\lambda]$
$[\mathbf{y}|\lambda] = \lambda\, e^{-\lambda\, y_1} \lambda\, e^{-\lambda\, y_2}$

Take the log of each side to make
the expression easier to differentiate:

Generalizing: $\lambda_{\text{mle}} = \frac{n}{\sum_{i=1}^{n} y_i}$

$\log([\mathbf{y}|\lambda]) = 2\log(\lambda) - \lambda\,(y_1 + y_2)$
$\frac{d[\log([\mathbf{y}|\lambda])]}{d\lambda} = \frac{2}{\lambda} - (y_1 + y_2)$

Set $= 0$ and solve for λ:

$\lambda_{\text{mle}} = \frac{2}{y_1 + y_2} = \frac{2}{10 + 18}$

Maximum likelihood by numerical methods

Main points

- ▶ Likelihood allows us to evaluate the relative strength of evidence for one parameter or model relative to another.

- ▶ 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 data are variable and the parameter are fixed in probability mass functions and probability density functions. These functions sum or integrate to one over the support of the random variable, y.

Looking ahead: the relationship between likelihood and Bayes

What must be done to assure that the area under
the curve = 1 ?