University of Nevada, Reno

Lecture 9: Maximum Likelihood

Perry Williams, PhD

NRES 779

Bayesian Hierarchical Modeling in Natural Resources

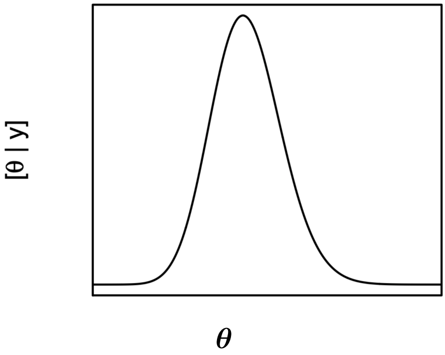
Why Likelihood

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

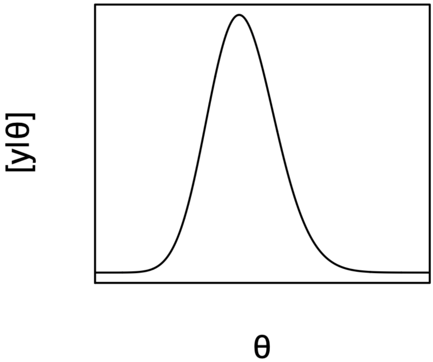
Learning Objectives (Lecture and Lab)

- 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.

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



Likelihood inference is based on $[y|\theta]$



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.

Fitting Models to Data

How do we fit models with lots of parameters to lots of data points?
In the example, we had:

- a single parameter, $\boldsymbol{\theta}$,
- one set of observations,
- 4 successes on ten draws,
- 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.,

$$\boldsymbol{\mu}_i = f(\boldsymbol{\theta}, x_i)$$
$$L(\boldsymbol{\mu}_i | y_i) \propto [y_i | \boldsymbol{\mu}_i, \sigma^2]$$

The total likelihood is the product of the individual likelihoods, assuming **conditional independence**

$$L(\boldsymbol{\theta} | y_i) = c \prod_{i=1}^{n} [y_i | f(\boldsymbol{\theta}, x_i), \sigma^2]$$

Conditional Independence

- The data are independent *conditional on the value of the model's parameters*. What this means is that the residuals (i.e., $\epsilon_i = y_i - f(\boldsymbol{\theta}, x_i)$ 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:

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

Exponential Distribution

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

- 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: $\lambda =$ the average rate of occurrence events per time or space
- Moments: mean $= \frac{1}{\lambda}$, variance $= \frac{1}{\lambda^2}$

Exponential Distribution

- 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 scaler or vector.
  - `rexp(x=y_i,rate=lambda)` returns a vector of length n of random draws from a exponential distribution.
  - 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

$$
\begin{aligned}
y_i &= 10 days \\
L(\lambda|y_i) &= [y_i|\lambda] \\
[y_i|\lambda] &= \lambda e^{-\lambda y_i}
\end{aligned}
$$

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

$$\log([y_i|\lambda]) = \log(\lambda) - \lambda y_i$$

$$\frac{d \log([y_i|\lambda])}{d\lambda} = \frac{1}{\lambda} - y_i$$

Set above equation to 0, and solve for λ.

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

Maximum Likelihood Estimate of λ, two data points

$$\mathbf{y} = [y_1, y_2] = 10\textit{days}, 18\textit{days}$$
$$L(\lambda|\mathbf{y}) = [y_1|\lambda][y_2|\lambda]$$
$$L(\lambda|\mathbf{y}) = \prod_{i=1}^{n=2}[y_i|\lambda]$$
$$[\mathbf{y}|\lambda] = \lambda e^{-\lambda y_1}\lambda e^{-\lambda y_2}$$
$$[\mathbf{y}|\lambda] = \prod_{i=1}^{n=2}\lambda e^{-\lambda y_i}$$

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

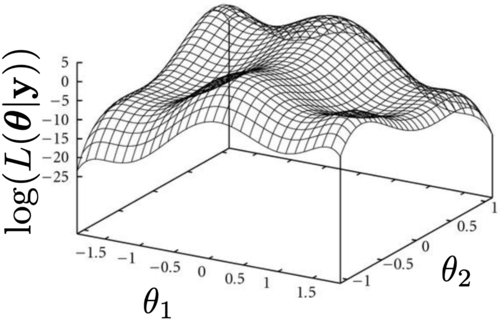
$$\log([\mathbf{y}|\lambda]) = 2\log(\lambda) - \lambda(y_1 + y_2)$$

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

Set above equation to 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 is 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 $= 1$?