

## Set08 - Likelihood

STAT 401 (Engineering) - Iowa State University

February 2, 2017

# Statistical modeling

## Definition

A **statistical model** is a pair  $(\mathcal{S}, \mathcal{P})$  where  $\mathcal{S}$  is the set of possible observations, i.e. the sample space, and  $\mathcal{P}$  is a set of probability distributions on  $\mathcal{S}$ .

Typically, we will assume the data have a specific form, say  $p(y|\theta)$ , but the parameter (vector)  $\theta$  is unknown. Thus  $p(y|\theta)$  for all allowable values for  $\theta$  provide the set  $\mathcal{P}$ .

# Binomial model

Suppose our data are

- the number of success  $y$
- out of some number of attempts  $n$
- where each attempt is independent
- given a common probability of success  $\theta$ .

Then a reasonable statistical model is

$$Y \sim \text{Bin}(n, \theta)$$

since for any  $0 < \theta < 1$  this model provides positive probability over the entire sample space, i.e. all possible observations.

Formally,

- $\mathcal{S} = \{0, 1, 2, \dots, n\}$
- $\mathcal{P} = \{\text{Bin}(n, \theta) : 0 < \theta < 1\}$ .

# Normal model

Suppose our data are

- a set of real numbers, i.e. between  $-\infty$  and  $\infty$ ,
- the population mean is  $\mu$  and population variance is  $\sigma^2$ ,
- the probability density function is reasonably approximated by a bell-shaped curve,
- and each observation is independent of the others.

Then a reasonable statistical model is

$$Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$$

since for  $-\infty < \mu < \infty, 0 < \sigma^2 < \infty$  this model provides positive density over the entire sample space, i.e. all possible observations.

Formally,

- $\mathcal{S} = \{y : y \in \mathbb{R}^n\}$
- $\mathcal{P} = \{N(\mu, \sigma^2) : -\infty < \mu < \infty, 0 < \sigma^2 < \infty\}$  where  $\theta = (\mu, \sigma^2)$ .

# Likelihood

## Definition

The **likelihood function**, or simply **likelihood**, is the joint probability mass/density function for fixed data when viewed as a function of the parameter vector  $\theta$ . Generally, we will write the joint probability mass or density function as  $p(y|\theta)$  and thus the likelihood is

$$L(\theta) = p(y|\theta)$$

but where  $y$  is fixed and known, i.e. it is your data.

The **log-likelihood** is the (natural) logarithm of the likelihood, i.e.

$$\ell(\theta) = \log L(\theta).$$

The likelihood describes the relative support in the data for different values for your parameter, i.e. the larger the likelihood is the more consistent that parameter value is with the data.

# Binomial likelihood

Suppose  $Y \sim \text{Bin}(n, \theta)$ , then

$$p(y|\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}.$$

where  $\theta$  is considered fixed (but often unknown) and the argument to this function is  $y$ .

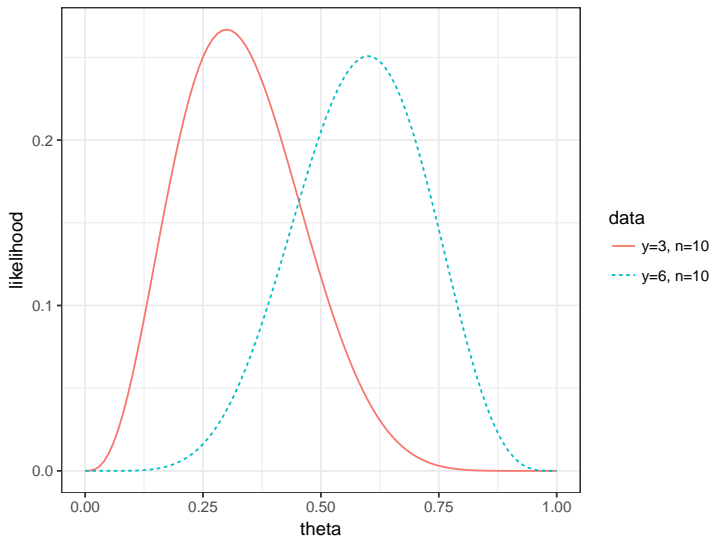
Thus the likelihood is

$$L(\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

where  $y$  is considered fixed and known and the argument to this function is  $\theta$ .

Not that I write  $L(\theta)$  without any condition, e.g. on  $y$ , so that you don't confuse this with a probability mass (or density) function.

# Binomial likelihood



# Likelihood for independent observations

Suppose  $Y_i$  are independent with marginal probability mass/density function  $p(y_i|\theta)$ .

The joint distribution for  $y = (y_1, \dots, y_n)$  is

$$p(y|\theta) = \prod_{i=1}^n p(y_i|\theta).$$

The likelihood for  $\theta$  is

$$L(\theta) = \prod_{i=1}^n p(y_i|\theta)$$

where we are thinking about this as a function of  $\theta$  for fixed  $y$ .



## Normal model

Suppose  $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ , then

$$p(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}$$

and

$$\begin{aligned} p(y|\mu, \sigma^2) &= \prod_{i=1}^n p(y_i|\mu, \sigma^2) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i-\mu)^2} \end{aligned}$$

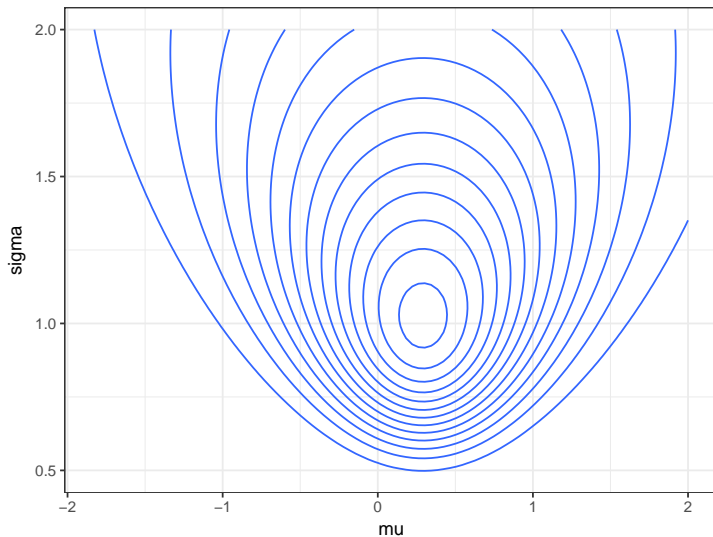
where  $\mu$  and  $\sigma^2$  are fixed (but often unknown) and the argument to this function is  $y = (y_1, \dots, y_n)$ .

The likelihood is

$$L(\mu, \sigma) = p(y|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i-\mu)^2}$$

where  $y$  is fixed and known and  $\mu$  and  $\sigma^2$  are the arguments to this function.

# Normal likelihood



# Maximum likelihood estimator

## Definition

The **maximum likelihood estimator (MLE)**,  $\hat{\theta}_{MLE}$  is the parameter value  $\theta$  that maximizes the likelihood function, i.e.

$$\hat{\theta}_{MLE} = \operatorname{argmax}_{\theta} L(\theta).$$

When the data are discrete, the MLE is parameter value that maximizes the probability of the observed data.

# Binomial MLE

If  $Y \sim \text{Bin}(n, \theta)$ , then

$$L(\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}.$$

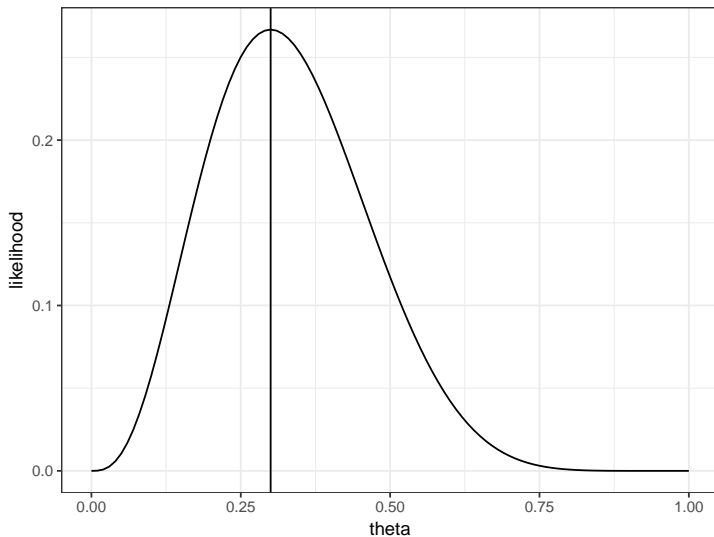
To find the MLE,

1. Take the derivative of  $\ell(\theta)$  with respect to  $\theta$ .
2. Set it equal to zero and solve for  $\theta$ .

$$\begin{aligned}\ell(\theta) &= \log \binom{n}{y} + y \log(\theta) + (n - y) \log(1 - \theta) \\ \frac{d}{d\theta} \ell(\theta) &= \frac{y}{\theta} - \frac{n-y}{1-\theta} \stackrel{\text{set}}{=} 0 \implies \\ \hat{\theta}_{MLE} &= y/n\end{aligned}$$

Take the second derivative of  $\ell(\theta)$  with respect to  $\theta$  and check to make sure it is negative.

# Binomial MLE



# Numerical maximization

```
likelihood <- function(theta) {  
  dbinom(3, size = 10, prob = theta)  
}  
  
optim(0.5, likelihood, lower = 0, upper = 1, method='L-BFGS-B',  
      control = list(fnscale = -1)) # maximize  
  
$par  
[1] 0.3000006  
  
$value  
[1] 0.2668279  
  
$counts  
function gradient  
      11      11  
  
$convergence  
[1] 0  
  
$message  
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```