

## Likelihoods (Part II)

Rebecca C. Steorts (slide adaption from Maria Tacket) and material from Chapter 2 of Roback and Legler text.

# Announcements

1. Office hours (mine) today will be from 3:00 – 4:00 PM in my office.
2. Homework 3 has been posted.
3. Any typos from last lecture have been updated, so please check for these.
4. Homework 3 has been posted (and the exercise has been clarified that corresponds with this.) Thanks to Zoe regarding catching this!
5. Homework 1 has been returned.

# Comments on Homework 1

1. For Exercise 2 part b, some students had a misunderstanding that if a predictor is binary, the linearity assumption is violated.
2. For Exercise 3 part a, some students did not fully understand what does “holding a variable constant” or “adjusting for a variable” mean when interpreting results from a multiple linear regression.
3. Question 6: Issue regarding the concept of interpreting the coefficients in the presence of an interaction term was not clear to all students.

Please attend lab on Friday if you would like to go over some of these concepts.

## Computing set up

```
knitr::opts_chunk$set(echo = T,  
                       results = "hide")  
  
library(tidyverse)  
library(tidymodels)  
library(GGally)  
library(knitr)  
library(patchwork)  
library(viridis)  
library(ggfortify)  
  
ggplot2::theme_set(ggplot2::theme_bw(base_size = 16))  
colors <- tibble::tibble(green = "#B5BA72")
```

# Using Likelihoods

- ▶ Describe the concept of a likelihood
- ▶ Construct the likelihood for a simple model
- ▶ Define the Maximum Likelihood Estimate (MLE) and use it to answer an analysis question
- ▶ Identify three ways to calculate or approximate the MLE and apply these methods to find the MLE for a simple model
- ▶ Use likelihoods to compare models

# What is the likelihood?

A **likelihood** is a function that tells us how likely we are to observe our data for a given parameter value (or values).

- ▶ Unlike Ordinary Least Squares (OLS), they do not require the responses be independent, identically distributed, and normal (iidN)
- ▶ They are not the same as probability functions

# Probability function vs. likelihood

- ▶ **Probability function:** Fixed parameter value(s) + input possible outcomes  $\Rightarrow$  probability of seeing the different outcomes given the parameter value(s)
- ▶ **Likelihood:** Fixed data + input possible parameter values  $\Rightarrow$  probability of seeing the fixed data for each parameter value

## Data: Fouls in college basketball games

The data set `04-refs.csv` includes 30 randomly selected NCAA men's basketball games played in the 2009 - 2010 season.<sup>1</sup>

We will focus on the variables `foul1`, `foul2`, and `foul3`, which indicate which team had a foul called them for the 1st, 2nd, and 3rd fouls, respectively.

- ▶ H: Foul was called on the home team
- ▶ V: Foul was called on the visiting team

We are focusing on the first three fouls for this analysis, but this could easily be extended to include all fouls in a game.

---

<sup>1</sup>The data set was derived from `basektball0910.csv` used in BMLR Section 11.2



## Fouls in college basketball games

```
refs <- read_csv("data/04-refs.csv")  
refs |> slice(1:5) |> kable()
```

We will treat the games as independent in this analysis.

# Different likelihood models

## **Model 1 (Unconditional Model):**

- ▶ What is the probability the referees call a foul on the home team, assuming foul calls within a game are independent?

# Different likelihood models

## **Model 1 (Unconditional Model):**

- ▶ What is the probability the referees call a foul on the home team, assuming foul calls within a game are independent?

## **Model 2 (Conditional Model):**

- ▶ Is there a tendency for the referees to call more fouls on the visiting team or home team?
- ▶ Is there a tendency for referees to call a foul on the team that already has more fouls?

**Ultimately we want to decide which model is better.**

# Exploratory data analysis

```
refs |>  
count(foul1, foul2, foul3) |> kable()
```

There are

- ▶ 46 total fouls on the home team
- ▶ 44 total fouls on the visiting team

## Model 1: Unconditional model

What is the probability the referees call a foul on the home team, assuming foul calls within a game are independent?

# Likelihood

Let  $p_H$  be the probability the referees call a foul on the home team.

**The likelihood for a single observation**

$$Lik(p_H) = p_H^{y_i}(1 - p_H)^{n_i - y_i}$$

Where  $y_i$  is the number of fouls called on the home team. (In this example, we know  $n_i = 3$  for all observations.)

# Likelihood

Let  $p_H$  be the probability the referees call a foul on the home team.

## **The likelihood for a single observation**

$$Lik(p_H) = p_H^{y_i}(1 - p_H)^{n_i - y_i}$$

Where  $y_i$  is the number of fouls called on the home team. (In this example, we know  $n_i = 3$  for all observations.)

## **Example**

For a single game where the first three fouls are  $H, H, V$ , then

$$Lik(p_H) = p_H^2(1 - p_H)^{3-2} = p_H^2(1 - p_H)$$

## Model 1: Likelihood contribution

Foul 1	Foul 2	Foul 3	n	Likelihood contribution
H	H	H	3	$p_H^3$
H	H	V	2	$p_H^2(1 - p_H)$
H	V	H	3	$p_H^2(1 - p_H)$
H	V	V	7	<b>A</b>
V	H	H	7	<b>B</b>
V	H	V	1	$p_H(1 - p_H)^2$
V	V	H	5	$p_H(1 - p_H)^2$
V	V	V	2	$(1 - p_H)^3$

Fill in **A** and **B**.



## Model 1: Solution Likelihood contribution

Foul 1	Foul 2	Foul 3	n	Likelihood contribution
H	H	H	3	$p_H^3$
H	H	V	2	$p_H^2(1 - p_H)$
H	V	H	3	$p_H^2(1 - p_H)$
H	V	V	7	$p_H(1 - p_H)^2$
V	H	H	7	$p_H^2(1 - p_H)$
V	H	V	1	$p_H(1 - p_H)^2$
V	V	H	5	$p_H(1 - p_H)^2$
V	V	V	2	$(1 - p_H)^3$

## Model 1: Likelihood function

Because the observations (the games) are independent, the **likelihood** is

$$Lik(p_H) = \prod_{i=1}^n p_H^{y_i} (1 - p_H)^{3-y_i}$$

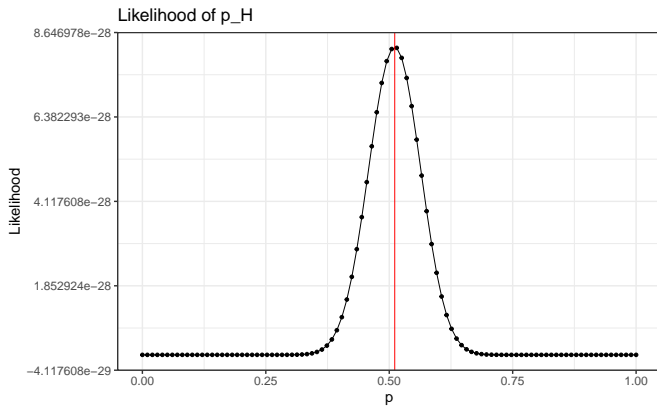
We will use this function to find the **maximum likelihood estimate (MLE)**. The MLE is the value between 0 and 1 where we are most likely to see the observed data.

# Finding the maximum likelihood estimate

There are three primary ways to find the MLE:

- ▶ Approximate using a graph
- ▶ Numerical approximation
- ▶ Using calculus

# Approximate MLE from a graph



## Visualizing the likelihood

What is your best guess for the MLE,  $\hat{p}_H$ ?

- a. 0.489
- b. 0.500
- c. 0.511
- d. 0.556

## Find the MLE using numerical approximation

Specify a finite set of possible values the for  $p_H$  and calculate the likelihood for each value

```
# write an R function for the likelihood
```

```
ref_lik <- function(ph) {  
  ph^46 *(1 - ph)^44  
}
```

```
# search possible values for p and return max
```

```
nGrid = 1000  
ph <- seq(0, 1, length = nGrid)  
lik <- ref_lik(ph)  
ph[lik == max(lik)]
```

```
# use the optimize function to find the MLE
```

```
optimize(ref_lik, interval = c(0,1), maximum = TRUE)
```

# Find the MLE using numerical approximation

Exercise:

1. Re-write the code to find the MLE using the log-likelihood.
2. The MLE should be the same.
3. Remember why we work with the log-likelihood (convenience and numerical stability).

Video on connections between likelihood and log-likelihood:

<https://www.youtube.com/watch?v=8nogLkirA3I>

**While it's correct to work with either one, it's best practice to work with the log-likelihood, so please do this in your homework assignment.**

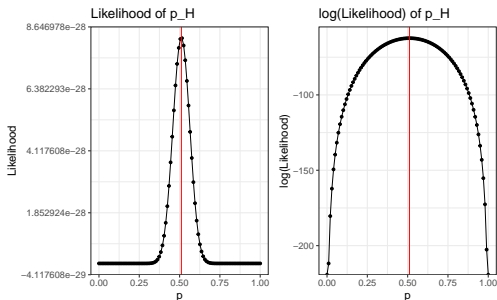
## Find MLE using calculus

- ▶ Find the MLE by taking the first derivative of the likelihood function.
- ▶ This can be tricky because of the Product Rule, so we can maximize the **log(Likelihood)** instead. The same value maximizes the likelihood and  $\log(\text{Likelihood})$



# Find MLE using calculus

- ▶ Find the MLE by taking the first derivative of the likelihood function.
- ▶ This can be tricky because of the Product Rule, so we can maximize the  **$\log(\text{Likelihood})$**  instead. The same value maximizes the likelihood and  $\log(\text{Likelihood})$



Find MLE using calculus

$$Lik(p_H) = \prod_{i=1}^n p_H^{y_i} (1 - p_H)^{3-y_i}$$

## Find MLE using calculus

$$Lik(p_H) = \prod_{i=1}^n p_H^{y_i} (1 - p_H)^{3-y_i}$$

$$\log(Lik(p_H)) = \sum_{i=1}^n y_i \log(p_H) + (3 - y_i) \log(1 - p_H)$$

$$= 46 \log(p_H) + 44 \log(1 - p_H)$$

Find MLE using calculus

$$\frac{d}{dp_H} \log(Lik(p_H)) = \frac{46}{p_H} - \frac{44}{1 - p_H} = 0$$

## Find MLE using calculus

$$\frac{d}{dp_H} \log(\text{Lik}(p_H)) = \frac{46}{p_H} - \frac{44}{1 - p_H} = 0$$

$$\Rightarrow \frac{46}{p_H} = \frac{44}{1 - p_H}$$

## Find MLE using calculus

$$\frac{d}{dp_H} \log(Lik(p_H)) = \frac{46}{p_H} - \frac{44}{1 - p_H} = 0$$

$$\Rightarrow \frac{46}{p_H} = \frac{44}{1 - p_H}$$

$$\Rightarrow 46(1 - p_H) = 44p_H$$

## Find MLE using calculus

$$\frac{d}{dp_H} \log(\text{Lik}(p_H)) = \frac{46}{p_H} - \frac{44}{1 - p_H} = 0$$

$$\Rightarrow \frac{46}{p_H} = \frac{44}{1 - p_H}$$

$$\Rightarrow 46(1 - p_H) = 44p_H$$

$$\Rightarrow 46 = 90p_H$$

## Find MLE using calculus

$$\frac{d}{dp_H} \log(\text{Lik}(p_H)) = \frac{46}{p_H} - \frac{44}{1 - p_H} = 0$$

$$\Rightarrow \frac{46}{p_H} = \frac{44}{1 - p_H}$$

$$\Rightarrow 46(1 - p_H) = 44p_H$$

$$\Rightarrow 46 = 90p_H$$

$$\hat{p}_H = \frac{46}{90} = 0.511$$



## Find MLE using calculus

$$\frac{d}{dp_H} \log(Lik(p_H)) = \frac{46}{p_H} - \frac{44}{1 - p_H} = 0$$

$$\Rightarrow \frac{46}{p_H} = \frac{44}{1 - p_H}$$

$$\Rightarrow 46(1 - p_H) = 44p_H$$

$$\Rightarrow 46 = 90p_H$$

$$\hat{p}_H = \frac{46}{90} = 0.511$$

## Model 2: Conditional model

Is there a tendency for referees to call more fouls on the visiting team or home team?

Is there a tendency for referees to call a foul on the team that already has more fouls?

## Model 2: Conditional model

Now let's assume fouls are not independent within each game. We will specify this dependence using conditional probabilities.

- ▶ **Conditional probability:**  $P(A|B)$  = Probability of  $A$  given  $B$  has occurred

## Model 2: Conditional model

Now let's assume fouls are not independent within each game. We will specify this dependence using conditional probabilities.

- ▶ **Conditional probability:**  $P(A|B)$  = Probability of  $A$  given  $B$  has occurred

Define new parameters:

- ▶  $p_{H|N}$ : Probability referees call foul on home team given there are equal numbers of fouls on the home and visiting teams

## Model 2: Conditional model

Now let's assume fouls are not independent within each game. We will specify this dependence using conditional probabilities.

- ▶ **Conditional probability:**  $P(A|B)$  = Probability of  $A$  given  $B$  has occurred

Define new parameters:

- ▶  $p_{H|N}$ : Probability referees call foul on home team given there are equal numbers of fouls on the home and visiting teams
- ▶  $p_{H|HBias}$ : Probability referees call foul on home team given there are more prior fouls on the home team

## Model 2: Conditional model

Now let's assume fouls are not independent within each game. We will specify this dependence using conditional probabilities.

- ▶ **Conditional probability:**  $P(A|B)$  = Probability of  $A$  given  $B$  has occurred

Define new parameters:

- ▶  $p_{H|N}$ : Probability referees call foul on home team given there are equal numbers of fouls on the home and visiting teams
- ▶  $p_{H|HBias}$ : Probability referees call foul on home team given there are more prior fouls on the home team
- ▶  $p_{H|VBias}$ : Probability referees call foul on home team given there are more prior fouls on the visiting team

## Model 2: Likelihood contributions

Foul 1	Foul 2	Foul 3	n	Likelihood contribution
H	H	H	3	$(p_{H N})(p_{H HBias})(p_{H HBias}) =$ $(p_{H N})(p_{H HBias})^2$
H	H	V	2	$(p_{H N})(p_{H HBias})(1 - p_{H HBias})$
H	V	H	3	$(p_{H N})(1 - p_{H HBias})(p_{H N})$
H	V	V	7	<b>A</b>
V	H	H	7	<b>B</b>
V	H	V	1	$(1 - p_{H N})(p_{H VBias})(1 - p_{H N}) =$ $(1 - p_{H N})^2(p_{H VBias})$
V	V	H	5	$(1 - p_{H N})(1 - p_{H VBias})(p_{H VBias})$
V	V	V	2	$(1 - p_{H N})(1 - p_{H VBias})(1 - p_{H VBias})$ $= (1 - p_{H N})(1 - p_{H VBias})^2$

Fill in **A** and **B**.

## Likelihood function

$$Lik(p_{H|N}, p_{H|HBias}, p_{H|VBias}) = [(p_{H|N})^{25}(1 - p_{H|N})^{23}(p_{H|HBias})^8 \\ (1 - p_{H|HBias})^{12}(p_{H|VBias})^{13}(1 - p_{H|VBias})^9]$$

**(Note: The exponents sum to 90, the total number of fouls in the data)**



## Likelihood function

$$Lik(p_{H|N}, p_{H|HBias}, p_{H|VBias}) = [(p_{H|N})^{25}(1 - p_{H|N})^{23}(p_{H|HBias})^8 \\ (1 - p_{H|HBias})^{12}(p_{H|VBias})^{13}(1 - p_{H|VBias})^9]$$

**(Note: The exponents sum to 90, the total number of fouls in the data)**

$$\log(Lik(p_{H|N}, p_{H|HBias}, p_{H|VBias})) = 25 \log(p_{H|N}) + 23 \log(1 - p_{H|N}) \\ + 8 \log(p_{H|HBias}) + 12 \log(1 - p_{H|HBias}) \\ + 13 \log(p_{H|VBias}) + 9 \log(1 - p_{H|VBias})$$

## Exercise

On your own, fill in the table and verify that the likelihood is indeed correct.

# Model Comparisons

- ▶ Nested Models
- ▶ Non-nested Models

## Comparing Nested Models

## Nested Models

Nested models: Models such that the parameters of the reduced model are a subset of the parameters for a larger model

Example:

$$\text{Model A: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$$\text{Model B: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$$

Model A is nested in Model B. We could use likelihoods to test whether it is useful to add  $x_3$  and  $x_4$  to the model.

$$H_0 : \beta_3 = \beta_4 = 0$$

$$H_a : \text{at least one } \beta_j \text{ is not equal to } 0$$

## Nested Models

Another way to think about nested models: Parameters in larger model can be equated to get the simpler model or if some parameters can be set to constants

Example:

Model 1:  $p_H$

Model 2:  $p_{H|N}, p_{H|HBias}, p_{H|VBias}$

Model 1 is nested in Model 2. The parameters  $p_{H|N}$ ,  $p_{H|HBias}$ , and  $p_{H|VBias}$  can be set equal to  $p_H$  to get Model 1.

$$H_0 : p_{H|N} = p_{H|HBias} = p_{H|VBias} = p_H$$

$H_a$  : At least one of  $p_{H|N}, p_{H|HBias}, p_{H|VBias}$  differs from the others

# Steps to compare models

1. Find the MLEs for each model.
2. Plug the MLEs into the log-likelihood function for each model to get the maximum value of the log-likelihood for each model.
3. Find the difference in the maximum log-likelihoods
4. Use the Likelihood Ratio Test to determine if the difference is statistically significant

## Step 1

Model 1:  $\hat{p}_H = 46/90 = 0.511$

```
loglik1 <- function(ph){  
  log(ph^46 * (1 - ph)^44)  
}  
loglik1(46/90)
```

Solution = -62.36



## Step 2

Model 2:

- ▶  $\hat{p}_{H|N} = 25/48 = 0.521$
- ▶  $\hat{p}_{H|HBias} = 8/20 = 0.4$
- ▶  $\hat{p}_{H|VBias} = 13/22 = 0.591$

```
loglik2 <- function(phn, phh, phv) {  
  log(phn^25 * (1 - phn)^23 * phh^8 *  
      (1 - phh)^12 * phv^13 * (1 - phv)^9)  
}  
(loglik2(25/48, 8/20, 13/22))
```

Solution = -61.57

## Step 3

Find the difference in the log-likelihoods

```
(diff <- loglik2(25/48, 8/20, 13/22) - loglik1(46/90))
```

Diff = 0.7878

Is the difference in the maximum log-likelihoods statistically significant?

# Likelihood Ratio Test

Test statistic

$$\begin{aligned} LRT &= 2[\max\{\log(Lik(\text{larger model}))\} - \max\{\log(Lik(\text{reduced model}))\}] \\ &= 2\log\left(\frac{\max\{(Lik(\text{larger model}))\}}{\max\{(Lik(\text{reduced model}))\}}\right) \end{aligned}$$

LRT follows a  $\chi^2$  distribution where the degrees of freedom equal the difference in the number of parameters between the two models

## Step 4

```
(LRT <- 2 * (loglik2(25/48, 8/20, 13/22) - loglik1(46/90)))
```

$LRT = 1.576$

The test statistic follows a  $\chi^2$  distribution with 2 degrees of freedom. Therefore, the p-value is  $P(\chi^2 > LRT)$ .

```
pchisq(LRT, 2, lower.tail = FALSE)
```

p-value = 0.4548

The p-value is very large, so we fail to reject  $H_0$ . We do not have convincing evidence that the conditional model is an improvement over the unconditional model.

Therefore, we can stick with the **unconditional** model.

## Comparing non-nested models

## Comparing non-nested models

We can also consider AIC and BIC. We cannot consider a LRT for a non-nested model.

## Comparing non-nested models

$$\text{AIC} = -2(\text{max log-likelihood}) + 2p$$

```
(Model1_AIC <- 2 * loglik1(46/90) + 2 * 1)
```

$$\text{AIC}_1 = -122.72$$

```
(Model2_AIC <- 2 * loglik2(25/48, 8/20, 13/22) + 2 * 3)
```

$$\text{AIC}_2 = -117.15$$

## Comparing non-nested models

```
(Model1_BIC <- 2 * loglik1(46/90) + 1 * log(30))
```

BIC\_1 = -121.32

```
(Model2_BIC <- 2 * loglik2(25/48, 8/20, 13/22) + 3 * log(30))
```

BIC\_2 = -112.94

Which model would you choose and why based upon the AIC and BIC values?



## Looking ahead

- ▶ Likelihoods help us answer the question of how likely we are to observe the data given different parameters
- ▶ In this example, we did not consider covariates, so in practice the parameters we want to estimate will look more similar to this

$$p_H = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

- ▶ Finding the MLE becomes much more complex (in practice) and numerical methods may be required.

# References