Maximum likelihood estimation

- HW 2 released on course website, due next Friday
- Project 1, Part 1 (EDA) released on course website, due in 2 weeks
    - Project 1 will be released in several smaller pieces

Recap

Definition: The *likelihood* $L(Model) = P(Data|Model)$ of a model is the probability of the observed data, given that we assume a certain model and certain values for the parameters that define that model.

Coin example: flip a coin 5 times, with $\pi = P(Heads)$

- Model: $Y_i \sim Bernoulli(\pi)$, and $\widehat{\pi} = 0.9$
- Data: $y_1, \ldots, y_5 = T, T, T, T, H$
- Likelihood: $L(\widehat{\pi}) = P(y_1, \ldots y_5|\pi = 0.9) = 0.00009$

Recap

Maximum likelihood estimation: pick the parameter estimate that maximizes the likelihood.

Coin example: flip a coin 5 times, with $\pi = P(Heads)$

$\hat{\pi} = P(Heads)$

- Observed data: T, T, T, T, H
- Likelihood: $L(\widehat{\pi}) = \underbrace{(1 - \widehat{\pi})^4}_{4\ T}(\widehat{\pi})$ 1 H
- Choose $\widehat{\pi}$ to maximize $L(\widehat{\pi})$

Computing likelihood in R

Observed data: T, T, T, T, H

- We are going to consider several different potential values for $\widehat{\pi}$:

$$0, 0.1, 0.2, 0.3, \ldots, 0.9, 1$$

- For each potential value, we will compute the likelihood:

$$L(\widehat{\pi}) = (1 - \widehat{\pi})^4 (\widehat{\pi})$$

- We then see which value has the highest likelihood.
- Is this all possible values? No, but let's start here.

R code

0.1 0.1

0 0.1 0.2 . . . 0.9 1

"sequence"

```
# List the values for pi hat
pi_hat <- seq(from = 0, to = 1, by = 0.1)

# Create a space to store the likelihoods
likelihood <- rep(0,length(pi_hat))
```

11

0 0 . . . 0 0 (11 times

Fill in likelihood for each value of

π_i in 0, 0.1, 0.2, . . . , 1

1: length (pi_hat)

1:11

[...]   position in a vector or matrix

pi_hat[1] = 0   pi_hat[2] = 0.1   pi_hat[3] = 0.2

R code

= 1 2 3 ··· 9 10 11

0 0.1 0.2 ··· 0.9 1   (11 values)

```
# List the values for pi hat
pi_hat <- seq(from = 0, to = 1, by = 0.1)

# Create a space to store the likelihoods
likelihood <- rep(0,length(pi_hat))

# Compute and store the likelihoods
for( i in 1:length(pi_hat) ){
  likelihood[i] <- pi_hat[i]*(1-pi_hat[i])^4
}
```

11

for loop

i = 1   likelihood[1] <- pi_hat[1] * (1- pi_hat[1])^4

i = 2   likelihood[2] <- pi_hat[2] * (1- pi_hat[2])^4

i = 3

⋮   ⋮

i = 11

$L(\hat{\pi}) = \hat{\pi}(1-\hat{\pi})^4$

pi-hat $[\ 0 \quad 0.1 \quad 0.2 \quad \ldots \quad 0.9 \quad 1\ ]$

likelihood $[\ \_\_ \quad \_\_ \quad \_\_ \quad \ldots \quad \_\_ \quad \_\_\ ]$

$0 * (1-0)^4$ $\quad$ $0.1(1-0.1)^4$ $\quad\quad$ $1(1-1)^4$

pi-hat[1] $\quad$ pi-hat[1] $\quad\quad$ pi-hat[11] $\quad$ pi-hat[11]

likelihood[i] = pi-hat[i] * (1- pi-hat[i])^4

R code

```
# List the values for pi hat
pi_hat <- seq(from = 0, to = 1, by = 0.1)

# Create a space to store the likelihoods
likelihood <- rep(0,length(pi_hat))

# Compute and store the likelihoods
for( i in 1:length(pi_hat) ){
  likelihood[i] <- pi_hat[i]*(1-pi_hat[i])^4
}
```

> Run this code in your R console. Which value of $\widehat{\pi}$ gives the highest likelihood?

Results

pi_hat	likelihood
0.0	0.00000
0.1	0.06561
0.2	0.08192
0.3	0.07203
0.4	0.05184
0.5	0.03125
0.6	0.01536
0.7	0.00567
0.8	0.00128
0.9	0.00009
1.0	0.00000

Issues: only considering a few values of π

$Y_1, \ldots, Y_n$ = observed data n = # observations

Maximum likelihood estimation with calculus

Suppose that $Y_i \sim Bernoulli(\pi)$. We observe n observations $Y_1, \ldots, Y_n$ and want to estimate π.

Step 1: Write down the likelihood

- Let $\widehat{\pi}$ be the estimate of π
- Let k be the number of times $Y_i = 1$ in the data

$\Rightarrow$ $n-k$ = # obs. where $Y_i = 0$

$$L(\widehat{\pi}) = \underbrace{\hat{\pi}^k}_{k \text{ obs. with } Y=1} \underbrace{(1-\hat{\pi})^{n-k}}_{n-k \text{ obs. w/ } Y=0}$$

$\Rightarrow \hat{\pi} \cdot \hat{\pi} \cdot \ldots \cdot \hat{\pi}$ (k times) $(1-\hat{\pi})(1-\hat{\pi}) \cdots (1-\hat{\pi})$ $(n-k)$

E.g. coin flip T, T, T, T, H

$n = 5$ $k = 1$ $n-k = 4$

$\hat{\pi}^1 (1-\hat{\pi})^4 = \hat{\pi}(1-\hat{\pi})^4$

Maximum likelihood estimation with calculus

$\log(x^y) = y\log(x)$

$\log(x \cdot y) = \log(x) + \log(y)$

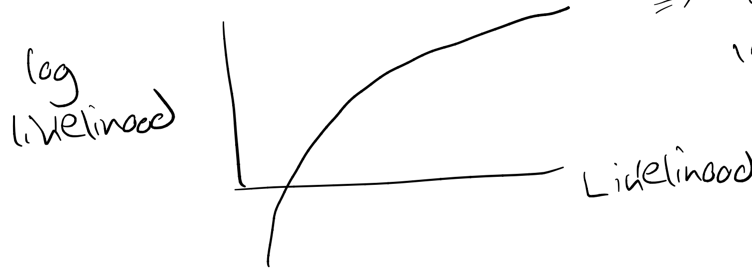
Step 1: Write down the likelihood

$$L(\widehat{\pi}) = \widehat{\pi}^k(1-\widehat{\pi})^{n-k}$$

Step 2: Take the log

$$\underbrace{\log L(\widehat{\pi})}_{\ell(\widehat{\pi})} = k\log(\hat{\pi}) + (n-k)\log(1-\hat{\pi})$$

- An advantage of taking the log is that it turns multiplication into addition, and exponents into multiplication
- This makes maximization easier
- Maximizing the log likelihood is the same as maximizing the likelihood



$\Rightarrow$ maximizing log likelihood $\Leftrightarrow$ maximizing likelihood

Maximum likelihood estimation with calculus

Step 2: log likelihood

$$\log L(\widehat{\pi}) = k\log(\widehat{\pi}) + (n-k)\log(1-\widehat{\pi})$$

- We want to find the value of $\widehat{\pi}$ that maximizes this function

> How do we find where maxima/minima occur for a function?

Maximum likelihood estimation with calculus

Step 2: log likelihood

$$\log L(\widehat{\pi}) = k \log(\widehat{\pi}) + (n-k)\log(1-\widehat{\pi})$$

- We want to find the value of $\widehat{\pi}$ that maximizes this function

> How do we find where maxima/minima occur for a function?

Take the first derivative and set equal to 0!

Maximum likelihood estimation with calculus

Want to differentiate

$$\log L(\widehat{\pi}) = k \log(\widehat{\pi}) + (n-k)\log(1-\widehat{\pi})$$

Remember some rules for differentiation:

- $\dfrac{d}{dx}\log x = \dfrac{1}{x}$
- $\dfrac{d}{dx} cf(x) = c\dfrac{d}{dx} f(x)$ for constant c
- $\dfrac{d}{dx}(f(x)+g(x)) = \dfrac{d}{dx} f(x) + \dfrac{d}{dx} g(x)$

we want

$$\frac{d\,\ell(\hat{\pi})}{d\hat{\pi}}$$

and so n, k are constants wrt $\hat{\pi}$

Maximum likelihood estimation with calculus

Step 3: take the first derivative, and set = 0

$$\log L(\widehat{\pi}) = k \log(\widehat{\pi}) + (n-k)\log(1-\widehat{\pi})$$

$$\frac{d}{d\widehat{\pi}} \log L(\widehat{\pi}) = \frac{\partial}{\partial \hat{\pi}}\left(k \log(\hat{\pi})\right) + \frac{\partial}{\partial \hat{\pi}}\left((n-k)\log(1-\hat{\pi})\right)$$

$$= \frac{k}{\hat{\pi}} + \frac{(n-k)}{1-\hat{\pi}}(-1)$$

$$= \frac{k}{\hat{\pi}} - \frac{(n-k)}{1-\hat{\pi}} \overset{\text{set}}{=} 0$$

$$\Rightarrow \frac{k}{\hat{\pi}} = \frac{n-k}{1-\hat{\pi}} \quad \Rightarrow \frac{1-\hat{\pi}}{\hat{\pi}} = \frac{n-k}{k}$$

$$\Rightarrow \frac{1}{\hat{\pi}} = \frac{n}{k} \quad \Rightarrow \boxed{\hat{\pi} = \frac{k}{n}}$$

\# 1s

\# observations

Maximum likelihood estimation with calculus

So our maximum likelihood estimate is $\widehat{\pi} = \dfrac{k}{n}$, the sample proportion

$k = 1$

$n = 5$

$\Rightarrow \hat{\pi} = \frac{1}{5} = 0.2$

- Our data: T, T, T, T, H
- This implies that $\widehat{\pi} = \dfrac{1}{5} = 0.2$
- This matches what we saw in R

Class activity

https://sta214-s23.github.io/class_activities/ca_lecture_5.html

Class activity

- $P(Y_i = 0) = \pi_0$
- $P(Y_i = -1) = 2\pi_0$
- $P(Y_i = 1) = 1 - 3\pi_0$

Observe data $-1, -1, 0, 1, 0, -1$.

$L(\widehat{\pi}_0) = \ ?$

$$L(\hat{\pi}_0) = P(-1)P(-1)P(0)P(1)P(0)P(-1)$$

$$= (2\hat{\pi}_0)(2\hat{\pi}_0)(\hat{\pi}_0)(1-3\hat{\pi}_0)(\hat{\pi}_0)(2\hat{\pi}_0)$$

$$= (2\hat{\pi}_0)^3 \cdot \hat{\pi}_0^2 (1-3\hat{\pi}_0)$$

Class activity

$$\log L(\widehat{\pi}_0) = 3\log(2) + 3\log(\widehat{\pi}_0) + 2\log(\widehat{\pi}_0) + \log(1 - 3\widehat{\pi}_0)$$

$$\frac{d}{d\widehat{\pi}_0}\log L(\widehat{\pi}_0) =$$

$$\frac{3}{\hat{\pi}_0} + \frac{2}{\hat{\pi}_0} - \frac{3}{1-3\hat{\pi}_0} \overset{\text{set}}{=} 0$$

$$\Rightarrow \frac{5}{\hat{\pi}_0} = \frac{3}{1-3\hat{\pi}_0}$$

$$\Rightarrow \frac{1-3\hat{\pi}_0}{\hat{\pi}_0} = \frac{3}{5}$$

$$\Rightarrow \frac{1}{\hat{\pi}_0} - 3 = \frac{3}{5}$$

$$\Rightarrow \frac{1}{\hat{\pi}_0} = \frac{3}{5} + 3 = \frac{18}{5}$$

$$\Rightarrow \hat{\pi}_0 = \frac{5}{18}$$