Maximum likelihood estimation

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 $\pi = P(Heads)$

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

Warm up: Class Activity, Part I

https://sta214-f22.github.io/class_activities/ca_lecture_6.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 $-1, -1, 0, 1, 0, -1$.

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

Class Activity

Old 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
}
```

> How should I modify this code to compute the new likelihood?

Class Activity

```r
# List the values for pi hat
pi_hat <- seq(from = 0, to = 1, by = 0.05)

# 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] <- (2*pi_hat[i])^3 *
      (pi_hat[i])^2 * (1 - 3 * pi_hat[i])
}
```

What is our estimate $\widehat{\pi}_0$?

So far

- Our R code suggests that $\widehat{\pi}_i$ maximizes the likelihood
- BUT, we haven't considered all possible values of $\widehat{\pi}_i$
- We could consider more values, but we can't compute a likelihood for every possible $\widehat{\pi}$, even in R
- Luckily, we don't have to

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

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

Maximum likelihood estimation with calculus

Step 1: Write down the likelihood

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

Step 2: Take the log

$\log L(\widehat{\pi}) =$

- 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

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
- $\dfrac{d}{dx}(f(x)+g(x)) = \dfrac{d}{dx}f(x) + \dfrac{d}{dx}g(x)$

Maximum likelihood estimation with calculus

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

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

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

Maximum likelihood estimation with calculus

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

- 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, Part II

https://sta214-f22.github.io/class_activities/ca_lecture_6.html

Class activity, Part II

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

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