

Maximum Likelihood Estimation

Idea: Choose the value in the parameter space that makes the observed data “most likely”.

Example:

Suppose that we flip a biased coin which has the probability of getting “Heads” as either 0.2, 0.3, or 0.8.

Suppose that we flip the coin 20 times and see the results:

H, H, T, H, H, H, H, T, H, H, H, H, H, T,
H, H, H, H, H, H

Which of 0.2, 0.3, or 0.8 seems “most likely”?

What if we only flip the coin twice?

Model:

For $i=1,2$, let

$$X_i = \begin{cases} 1 & , \text{if we get "Heads" on the } i\text{th flip} \\ 0 & , \text{if we get "Tails" on the } i\text{th flip} \end{cases}$$

Let $p=P(\text{"Heads" on any one flip})$

Then

$$X_1, X_2 \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$$

where $p \in \{0.2, 0.3, 0.8\}$

Joint pmf:

$$\begin{aligned} f(x_1, x_2) &= P(X_1 = x_1, X_2 = x_2) \\ &= P(X_1 = x_1) \cdot P(X_2 = x_2) \quad (\text{indep}) \\ &= p^{x_1}(1-p)^{1-x_1} \mathbf{I}_{\{0,1\}}(x_1) \cdot p^{x_2}(1-p)^{1-x_2} \mathbf{I}_{\{0,1\}}(x_2) \end{aligned}$$

Tabulated values of the joint pmf:

		(x_1, x_2)			
		$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
p	0.2	0.64	0.16	0.16	0.04
	0.3	0.49	0.21	0.21	0.06
	0.8	0.04	0.16	0.16	0.64

		(x_1, x_2)			
		(0,0)	(0,1)	(1,0)	(1,1)
p	0.2	0.64	0.16	0.16	0.04
	0.3	0.49	0.21	0.21	0.06
	0.8	0.04	0.16	0.16	0.64

- When we observe the data to be (0,0)
i.e. (“Tails”, “Tails”),

the value of p that gives the highest
joint probability (0.64) is 0.2.

		(x_1, x_2)			
		(0,0)	(0,1)	(1,0)	(1,1)
p	0.2	0.64	0.16	0.16	0.04
	0.3	0.49	0.21	0.21	0.06
	0.8	0.04	0.16	0.16	0.64

- When we observe the data to be (0,1) or (1,0) i.e. (“Tails”, “Heads”) or (“Heads”, “Tails”),

the value of p that gives the highest joint probability (0.21) is 0.3.

		(x_1, x_2)			
		$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
p	0.2	0.64	0.16	0.16	0.04
	0.3	0.49	0.21	0.21	0.06
	0.8	0.04	0.16	0.16	0.64

- When we observe the data to be $(1,1)$ i.e. (“Heads”, “Heads”),

the value of p that gives the highest joint probability (0.64) is 0.8.

- When the data is $(x_1, x_2) = (0, 0)$
 $p = 0.2$ is “most likely”
(gives highest probability of seeing the data)
-

- When the data is $(x_1, x_2) = (0, 1)$ or
 $(x_1, x_2) = (1, 0)$
 $p = 0.3$ is “most likely”
(gives highest probability of seeing the data)
-

- When the data is $(x_1, x_2) = (1, 1)$
 $p = 0.8$ is “most likely”
(gives highest probability of seeing the data)

The maximum likelihood estimator for p is:

$$\hat{p} = \begin{cases} 0.2 & , \text{if } (x_1, x_2) = (0, 0) \\ 0.3 & , \text{if } (x_1, x_2) = (0, 1) \text{ or } (1, 0) \\ 0.8 & , \text{if } (x_1, x_2) = (1, 1) \end{cases}$$