

Review of Maximum Likelihood Estimation:

Flip a possibly unfair coin n times.

Let p be the probability of getting “Heads” on any one flip. $(0 \leq p \leq 1)$

Record 1’s and 0’s for H’s and T’s, respectively.

We now have a random sample

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

Review of Maximum Likelihood Estimation:

p is unknown and we want to estimate it.

- If the observed data has a lot of 1's in it, a higher value of p , closer to 1 is more likely.
- If the observed data has a lot of 0's in it, a lower value of p , closer to 0 is more likely.
- If the observed data has roughly an equal number of 0's and 1's, a value of p closer to 0.5 is more likely.

The Bernoulli probability mass function is

$$f(x; p) = p^x(1 - p)^{1-x}$$

for $x=0,1$. It is zero otherwise.

The joint pmf for X_1, X_2, \dots, X_n is

$$f(\vec{x}; p) \stackrel{\text{iid}}{=} \prod_{i=1}^n f(x_i; p)$$

$$= p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

for $x_i \in \{0, 1\}$.

$$f(\vec{x}; p) = P(\underbrace{X_1 = x_1, \dots, X_n = x_n}_{\text{This is a function of } p})$$

This is a function of p .

Find the value of p in $[0,1]$ that makes the probability of seeing $X_1 = x_1$, $X_2 = x_2, \dots X_n = x_n$ largest.

$$f(\vec{x}; p) = P(\underbrace{X_1 = x_1, \dots, X_n = x_n}_{\text{This is a function of } p})$$

This is a function of p .

Find the value of p in $[0,1]$ that makes the probability of seeing $X_1 = x_1$, $X_2 = x_2, \dots X_n = x_n$ “most likely”.

i.e. This is called the **maximum likelihood estimator** for p .

$$f(\vec{x}; p) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

Think about this as a function of p :

$$L(p) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

This is called a **likelihood function**.

$$L(p) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

It is easier to maximize the **log-likelihood**.

$$\ell(p) = \ln L(p)$$

$$= \left(\sum_{i=1}^n x_i \right) \ln p + \left(n - \sum_{i=1}^n x_i \right) \ln(1 - p)$$

$$\begin{aligned} \frac{d}{dp} \ell(p) &= \left(\sum_{i=1}^n x_i \right) \frac{1}{p} - \left(n - \sum_{i=1}^n x_i \right) \frac{1}{1-p} \\ &\stackrel{\text{set}}{=} 0 \end{aligned}$$

The MLE for p is

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

For continuous X_1, X_2, \dots, X_n , the joint pdf does not represent probability but the MLE is found in the same way.

Example:

Suppose that X_1, X_2, \dots, X_n is a random sample from the continuous Pareto distribution with pdf

$$f(x, \gamma) = \begin{cases} \frac{\gamma}{(1+x)^{\gamma+1}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

The joint pdf is

$$f(\vec{x}; \gamma) \stackrel{\text{iid}}{=} \prod_{i=1}^n f(x_i; \gamma)$$

$$\begin{aligned} &= \prod_{i=1}^n \frac{\gamma}{(1 + x_i)^{\gamma+1}} = \frac{\gamma^n}{\prod_{i=1}^n (1 + x_i)^{\gamma+1}} \\ &= \frac{\gamma^n}{\left[\prod_{i=1}^n (1 + x_i) \right]^{\gamma+1}} \end{aligned}$$

A likelihood is

$$L(\gamma) = \frac{\gamma^n}{\left[\prod_{i=1}^n (1 + x_i) \right]^{\gamma+1}}$$

$$\ell(\gamma) = \ln L(\gamma)$$

$$= n \ln \gamma - (\gamma + 1) \ln \left[\prod_{i=1}^n (1 + x_i) \right]$$

Equivalently,

$$\ell(\gamma) = n \ln \gamma - (\gamma + 1) \sum_{i=1}^n \ln(1 + x_i)$$

$$\ell'(\gamma) = \frac{n}{\gamma} - \sum_{i=1}^n \ln(1 + x_i) \stackrel{\text{set}}{=} 0$$

The MLE for γ is

$$\hat{\gamma} = \frac{n}{\sum_{i=1}^n \ln(1 + X_i)}$$