LECTURE: MAXIMUM LIKELIHOOD ESTIMATE

1. MLE

Setting: Suppose you're trying to find a "good" estimator $\hat{\theta}$ for an unknown parameter θ

Definition:

Let $Y_1, Y_2, \ldots, Y_n$ be iid samples from a population with pmf $p_\theta(y)$ or density $f_\theta(y)$

The **likelihood** of the sample is given by:

$$L(Y_1, Y_2, \ldots, Y_n|\theta) = p_\theta(Y_1)p_\theta(Y_2)\cdots p_\theta(Y_n) \qquad \text{(discrete)}$$
$$L(Y_1, Y_2, \ldots, Y_n|\theta) = f_\theta(Y_1)f_\theta(Y_2)\cdots f_\theta(Y_n) \qquad \text{(continuous)}$$

Here we're using the subscript θ to emphasize the dependence on θ

Intuitively, think of L as the joint pmf/density of $Y_1, \cdots, Y_n$

The maximum likelihood estimator (MLE) chooses the value of the parameter θ which maximizes the likelihood of our sample.

Definition:

The **maximum likelihood estimator** $\hat{\theta}_{\text{MLE}}$ is the value of θ which maximizes the likelihood $L(Y_1, Y_2, \ldots, Y_n|\theta)$

2. Uniform Example

Example 1:

Suppose we have a population with a uniform distribution on $[0, b]$

Take samples $Y_1, \ldots, Y_n$ from this distribution.

(a) Find the MLE for b

STEP 1: The density of a uniform distribution on $[0, b]$ is

$$f_\theta(y) = \begin{cases} \frac{1}{b} & 0 \leq y \leq b \\ 0 & \text{otherwise} \end{cases}$$

STEP 2: Thus the likelihood function is

$$L(Y_1, Y_2, \ldots, Y_n|\theta) = f_\theta(Y_1) f_\theta(Y_2) \cdots f_\theta(Y_n) \quad = \begin{cases} \frac{1}{b^n} & 0 \leq Y_i \leq b \text{ for all } i \\ 0 & \text{otherwise} \end{cases}$$

We don't want L to be 0, since that is not a maximum, hence we need $0 \leq Y_i \leq b$ for all i

STEP 3: In order to make $\frac{1}{b^n}$ as *large* as possible, we need b to be as *small* as possible, while keeping the requirement $0 \leq Y_i \leq b$ for all i

Question: What is the smallest b such that $[0, b]$ contains all the samples Y_i?

Answer: Let b be the largest sample point $\max(Y_1, Y_2, \cdots, Y_n)$

Hence our MLE is $\hat{b}_{\text{MLE}} = Y_{\max} = \max(Y_1, Y_2, \ldots, Y_n)$

Note: This is different from our method of moments estimator $2\bar{Y}$

(b) Is $\hat{b}_{\text{MLE}}$ an unbiased estimator for b?

We need to find $E\left(\hat{b}_{\text{MLE}}\right)$ which requires finding the density of $Y_{\max}$

This is a little tricky to compute, but much easier in terms of CDFs.

STEP 1:

Let $Y \sim \text{Uniform}[0, b]$ and let $F(y) = P(Y \leq y)$ be the CDF of Y

Let $Y_{\max} = \max(Y_1, Y_2, \ldots, Y_n)$ and let $F_{\max}$ be the CDF for $Y_{\max}$.

$$\begin{aligned}
F_{\max}(y) =& P(Y_{\max} \leq y) = P(\max(Y_1, Y_2, \ldots, Y_n) \leq y) = P(Y_i \leq y \text{ for all } i)\\
=& P(Y_1 \leq y)P(Y_2 \leq y) \cdots P(Y_n \leq y)\\
=& F(y)F(y) \cdots F(y)\\
=& (F(y))^n
\end{aligned}$$

STEP 2: Since the density of Y is $f(y) = \frac{1}{b}$ for $0 \leq y \leq b$ and the CDF is the integral of the density, we get

$$F(y) = \begin{cases} 0 & y < 0 \\ \frac{y}{b} & 0 \leq y \leq b \\ 1 & y > 1 \end{cases}$$

$$\text{Therefore } F_{\max}(y) = (F(y))^n = \left(\frac{y}{b}\right)^n \text{ if } 0 \leq y \leq b$$

STEP 4: Since the density is the derivative of the cdf we get

$$f_{\max}(y) = \frac{d}{dy}F_{\max}(y) = ny^{n-1}\frac{1}{b^n}$$

The density is 0 outside the interval $[0, b]$, so with the appropriate limits the density becomes

$$f_{\max}(y) = \begin{cases} ny^{n-1}\frac{1}{b^n} & 0 \leq y \leq b \\ 0 & \text{otherwise} \end{cases}$$

STEP 5: Having the density, we can now calculate

$$E(Y_{\max}) = \int_0^b y f_{\max}(y) dy = \int_0^b y \left(ny^{n-1} \frac{1}{b^n} \right) dy = \frac{n}{b^n} \int_0^b y^n dy = \frac{n}{b^n} \left[\frac{y^{n+1}}{n+1} \right]_0^b$$
$$= \frac{n}{b^n} \left(\frac{b^{n+1}}{n+1} \right) = \left(\frac{n}{n+1} \right) b$$

Since this is not equal to b, the $\hat{b}_{\text{MLE}}$ is a biased estimator for b

(c) Use (b) to find an unbiased estimator of b

We can convert this to an unbiased estimator by multiplying it by $\frac{n+1}{n}$

$$\text{Hence } \hat{b} = \left(\frac{n+1}{n} \right) Y_{\max} = \frac{n}{n+1} \max(Y_1, Y_2, \ldots, Y_n)$$

Is an unbiased estimator for b

3. Geometric Example

Example 2:

Suppose we have a population which has a geometric distribution with parameter p

Take samples $Y_1, \dots, Y_n$ from this distribution

Find the MLE for p

Recall:

The pmf for $Y \sim$ Geom (p) is

$$p(k) = P(Y = k) = (1-p)^{k-1}p$$

STEP 1: Our likelihood function is:

$$L(Y_1, \dots, Y_n | p) = p(Y_1)p(Y_2)\cdots p(Y_n) = \prod_{i=1}^{n} (1-p)^{Y_i - 1} p$$

Note: $\prod_{i=1}^{n}$ means take the product of all the $(1-p)^{Y_i-1}p$. It's like $\sum_{i=1}^{n}$ but with products

STEP 2: To maximize this, we need to use calculus: Differentiate with respect to p and set the derivative equal to zero.

This becomes messy in general, but much easier with:

Trick: Take logarithms

$$\begin{aligned}\ln L(Y_1,\ldots,Y_n|p) &= \ln\prod_{i=1}^{n}(1-p)^{Y_i-1}p = \sum_{i=1}^{n}\ln(1-p)^{Y_i-1}+\sum_{i=1}^{n}\ln p\\ &= n\ln p+\sum_{i=1}^{n}(Y_i-1)\ln(1-p)\end{aligned}$$

Taking the derivative with respect to p:

$$\begin{aligned}\frac{d}{dp}\ln L(Y_1,...,Y_n|p) =& \frac{n}{p}+\sum_{i=1}^{n}(Y_i-1)\left(\frac{-1}{1-p}\right) = \frac{n}{p}-\frac{1}{1-p}\left(\sum_{i=1}^{n}Y_i-\sum_{i=1}^{n}1\right)\\ =& \frac{1}{1-p}\left(\sum_{i=1}^{n}Y_i-n\right)=0\end{aligned}$$

STEP 3: Solve for p, since that's what we're trying to estimate

$$\begin{aligned}\frac{1}{1-p}\left(\sum_{i=1}^{n}Y_i-n\right) &= \frac{n}{p}\\ p\left(\sum_{i=1}^{n}Y_i-n\right) &= n(1-p)\\ p\left(\sum_{i=1}^{n}Y_i\right)-\cancel{np} &= n-\cancel{np}\\ p &= \frac{n}{\sum_{i=1}^{n}Y_i}=\frac{1}{\bar{Y}}\end{aligned}$$

STEP 4: **Answer:**

Thus the MLE for the parameter p is $\hat{p}_{\text{MLE}}=\frac{1}{\bar{Y}}$

Note: This is the same estimator as the method of moments

Note: The trick works because if f is any positive function, then $\frac{d}{dx}\ln(f(x)) = \frac{f'(x)}{f(x)}$ so if the left-hand-side is zero, then $f'(x) = 0$ and vice-versa.

4. Poisson Example

Example 3:

Suppose we have a population which has a Poisson distribution with parameter λ.

Take samples $Y_1, \ldots, Y_n$ from this distribution.

Find the MLE for λ

STEP 1: Using the Poisson pmf, The likelihood function is:

$$
\begin{aligned}
L(Y_1, \ldots, Y_n|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{Y_i}}{Y_i!} &= e^{-\lambda}\frac{\lambda^{Y_1}}{Y_1!} \cdots e^{-\lambda}\frac{\lambda^{Y_n}}{Y_n!} \\
&= e^{-n\lambda}\lambda^{Y_1+\cdots+Y_n}\frac{1}{Y_1!\cdots Y_n!} = e^{-n\lambda}\lambda^{n\bar{Y}}\prod_{i=1}^{n}\frac{1}{Y_i!}
\end{aligned}
$$

STEP 2: To maximize this with respect to λ, we will once again maximize the log likelihood function.

$$
\begin{aligned}
\ln L(Y_1, \ldots, Y_n|\lambda) &= \ln(e^{-n\lambda}) + \ln(\lambda^{n\bar{Y}}) + \ln\left(\prod_{i=1}^{n}\frac{1}{Y_i!}\right) \\
&= -n\lambda + n\bar{Y}\ln(\lambda) + \ln\left(\prod_{i=1}^{n}\frac{1}{Y_i!}\right)
\end{aligned}
$$

Taking the derivative with respect to λ:

$$\frac{d}{d\lambda}\ln L(Y_1, \ldots, Y_n|\lambda) = -n + \frac{n\bar{Y}}{\lambda}$$

STEP 3: Setting this equal to 0, we get $\lambda = \bar{Y}$

Therefore our MLE estimator is $\hat{\lambda}_{\text{MLE}} = \bar{Y}$ which is the same estimator we got using the method of moments.