Defining the likelihood of data

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- X_i was drawn from a distribution with density function $f(X_i|\theta)$.
  (or mass)
- Sample: $(X_1, X_2, \ldots, X_n)$

Likelihood question:

How likely is the sample $(X_1, X_2, \ldots, X_n)$ given the parameter θ?

this follows from generic definition when X_i are iid.

Likelihood function, $L(\theta)$:

this is the definition of $L(\theta)$ in all scenarios

$$L(\theta) = f(X_1, X_2, \ldots, X_n|\theta) = \boxed{\prod_{i=1}^{n} f(X_i|\theta)}$$

This is just a product, since X_i are iid.

Maximum Likelihood Estimator

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$, drawn from a distribution $f(X_i|\theta)$.

def The **Maximum Likelihood Estimator (MLE)** of θ is the value of θ that maximizes $L(\theta)$. → i.e. maximizes the likelihood of the observed data.

$$\boxed{\theta_{MLE} = \arg\max_{\theta} L(\theta)}$$

Maximum Likelihood Estimator

Consider a sample of n iid random variables $X_1, X_2, \dots, X_n$, drawn from a distribution $f(X_i|\theta)$.

<u>def</u> The Maximum Likelihood Estimator (MLE) of θ is the value of θ that maximizes $L(\theta)$.

$$\theta_{MLE} = \arg\max_{\theta} L(\theta)$$

Likelihood of your sample

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

For continuous X_i, $f(X_i|\theta)$ is PDF, and for discrete X_i, $f(X_i|\theta)$ is PMF

Maximum Likelihood Estimator

Consider a sample of n iid random variables $X_1, X_2, \dots, X_n$, drawn from a distribution $f(X_i|\theta)$.

<u>def</u> The Maximum Likelihood Estimator (MLE) of θ is the value of θ that maximizes $L(\theta)$.

$$\theta_{MLE} = \arg\max_{\theta} L(\theta)$$

The argument θ
that maximizes $L(\theta)$

argmax and log likelihood

MLE: Bernoulli

Computing the MLE

$\theta_{MLE} = \arg\max_{\theta} LL(\theta)$

General approach for finding θ_{MLE} , the MLE of θ:

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

2. Differentiate $LL(\theta)$ w.r.t. (each) θ

$$\frac{\partial LL(\theta)}{\partial \theta}$$

To maximize: $\frac{\partial LL(\theta)}{\partial \theta} = 0$

3. Solve resulting equations

(algebra or computer)

4. Make sure derived $\hat{\theta}_{MLE}$ is a maximum
   - Check $LL(\theta_{MLE} \pm \epsilon) < LL(\theta_{MLE})$
   - Often ignored in expository derivations
   - We'll ignore it here too (and won't require it in class)

$LL(\theta)$ is often easier to differentiate than $L(\theta)$.

Maximum Likelihood with Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \ldots, X_n$.
What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|p)$$

?

$$f(X_i|p) = \begin{cases} p & \text{if } X_i = 1 \\ 1-p & \text{if } X_i = 0 \end{cases}$$

function as expressed is not differentiable! not what we want! :(

2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0

3. Solve resulting equations

Maximum Likelihood with Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \dots, X_n$.
What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.
- $f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$

1. Determine formula for $LL(\theta)$
2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0
3. Solve resulting equations

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|p)$$

$$f(X_i|p) = \begin{cases} p & \text{if } X_i = 1 \\ 1-p & \text{if } X_i = 0 \end{cases}$$

$f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$ where $X_i \in \{0,1\}$

expanded $\begin{cases} X_i = 1? \quad f(X_i = 1|p) = p^1(1-p)^{1-1} = p^1(1-p)^0 = p \\ X_i = 0? \quad f(X_i = 0|p) = p^0(1-p)^{1-0} = p^0(1-p)^1 = 1-p \end{cases}$

✅
- Is differentiable with respect to p
- Valid PMF over discrete domain

Maximum Likelihood with Bernoulli

$\left.\begin{array}{l} \log ab = \log a + \log b \\ \log c^d = d \log c \end{array}\right]$ properties of log

Consider a sample of n iid RVs $X_1, X_2, \dots, X_n$.
What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.
- $f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$

1. Determine formula for $LL(\theta)$
2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0
3. Solve resulting equations

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|p) = \sum_{i=1}^{n} \underbrace{\log\left(p^{X_i}(1-p)^{1-X_i}\right)}_{\log p^{X_i} + \log (1-p)^{1-X_i}}$$

$$= \sum_{i=1}^{n} \left[X_i \log p + (1 - X_i)\log(1-p)\right]$$

$$\log p \sum_{i=1}^{n} X_i + \log(1-p) \sum_{i=1}^{n} 1 - \log(1-p) \sum_{i=1}^{n} X_i$$

$$= Y(\log p) + (n - Y)\log(1-p), \text{ where } Y = \sum_{i=1}^{n} X_i$$

Maximum Likelihood with Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \dots, X_n$.

What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.
- $f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} [X_i \log p + (1 - X_i)\log(1-p)]$$

$$= Y(\log p) + (n - Y)\log(1-p), \text{ where } Y = \sum_{i=1}^{n} X_i$$

2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0

$$\frac{\partial LL(\theta)}{\partial p} = Y\frac{1}{p} + (n - Y)\frac{-1}{1-p} = 0$$

3. Solve resulting equations

Maximum Likelihood with Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \dots, X_n$.

What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.
- $f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} [X_i \log p + (1 - X_i)\log(1-p)]$$

$$= Y(\log p) + (n - Y)\log(1-p), \text{ where } Y = \sum_{i=1}^{n} X_i$$

2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0

$$\frac{\partial LL(\theta)}{\partial p} = Y\frac{1}{p} + (n - Y)\frac{-1}{1-p} = 0$$

$\frac{Y}{p} = \frac{n-Y}{1-p}$

$Y - \cancel{Yp} = np - \cancel{Yp} \rightarrow p = \frac{Y}{n}$

3. Solve resulting equations

$$p_{MLE} = \frac{1}{n}Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

MLE of the Bernoulli parameter, p_{MLE}, is the unbiased estimate of the mean, $\bar{X}$ (sample mean)

Quick check

- You draw n iid random variables $X_1, X_2, \ldots, X_n$ from the distribution F, yielding the following sample:

$$[0, 0, 1, 1, 1, 1, 1, 1, 1, 1] \qquad (n = 10)$$

- Suppose distribution $F = \text{Ber}(p)$ with unknown parameter p.

1. What is p_{MLE}, the MLE of the parameter p?

A. 1.0
B. 0.5
C. 0.8
D. 0.2
E. None/other

$$p_{MLE} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Quick check

- You draw n iid random variables $X_1, X_2, \ldots, X_n$ from the distribution F, yielding the following sample:

$$[0, 0, 1, 1, 1, 1, 1, 1, 1, 1] \qquad (n = 10)$$

- Suppose distribution $F = \text{Ber}(p)$ with unknown parameter p.

1. What is p_{MLE}, the MLE of the parameter p? C. 0.8
2. What is the likelihood $L(\theta)$ of this specific sample?

$f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$ where $X_i \in \{0,1\}$

$$L(\theta) = \prod_{i=1}^{n} f(X_i|p) \quad \text{where } \theta = p$$

$$= p^8(1-p)^2 = 0{,}8^{8}\ 0{,}2^{2} = 0{,}0067$$

MLE: Poisson and Uniform

Maximum Likelihood with Poisson

Consider a sample of n iid RVs $X_1, X_2, \dots, X_n$.
What is $\theta_{MLE} = \lambda_{MLE}$?

recall that $\log ab = \log a + \log b$
$\log \frac{a}{b} = \log a - \log b$

- Let $X_i \sim \text{Poi}(\lambda)$.
- PMF: $f(X_i|\lambda) = \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} \log\left(\frac{e^{-\lambda}\lambda^{X_i}}{X_i!}\right) = \sum_{i=1}^{n} (-\lambda \log e + X_i \log \lambda - \log X_i!)$$

$$= -n\lambda + \log(\lambda) \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \log(X_i!)$$

(using natural log, $\ln e = 1$)

Maximum Likelihood with Normal

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$. $\quad f(X_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(X_i-\mu)^2/(2\sigma^2)}$

What is $\theta_{MLE} = (\mu_{MLE}, \sigma^2_{MLE})$?

1. Determine formula for $LL(\theta)$
2. Differentiate $LL(\theta)$ w.r.t. (each) θ, set to 0
3. Solve resulting equations

$$LL(\theta) = -\sum_{i=1}^{n} \log(\sqrt{2\pi}\sigma) - \sum_{i=1}^{n} [(X_i-\mu)^2/(2\sigma^2)]$$

with respect to μ

$$\frac{\partial LL(\theta)}{\partial \mu} = \sum_{i=1}^{n} [2(X_i-\mu)/(2\sigma^2)] = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0$$

with respect to σ

$$\frac{\partial LL(\theta)}{\partial \sigma} = -\sum_{i=1}^{n}\frac{1}{\sigma} + \sum_{i=1}^{n} 2(X_i-\mu)^2/(2\sigma^3) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$$

Maximum Likelihood with Normal

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$. $\quad f(X_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(X_i-\mu)^2/(2\sigma^2)}$

What is $\theta_{MLE} = (\mu_{MLE}, \sigma^2_{MLE})$?

3. Solve resulting equations

Two equations, two unknowns:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0 \qquad -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$$

First, solve for μ_{MLE}:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i - \frac{1}{\sigma^2}\sum_{i=1}^{n}\mu = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i = n\mu \quad \Rightarrow \quad \mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

unbiased

another sample mean!

Maximum Likelihood with Normal

Consider a sample of n iid random variables $X_1, X_2, \dots, X_n$.

- Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$. $\quad f(X_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(X_i-\mu)^2/(2\sigma^2)}$

What is $\theta_{MLE} = (\mu_{MLE}, \sigma^2_{MLE})$?

3. Solve resulting equations

Two equations, two unknowns: $\quad \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0 \qquad -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$

First, solve for μ_{MLE}: $\quad \frac{1}{\sigma^2}\sum_{i=1}^{n}X_i - \frac{1}{\sigma^2}\sum_{i=1}^{n}\mu = 0 \quad \Rightarrow \quad \sum_{i=1}^{n}X_i = n\mu \quad \Rightarrow \quad \mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n}X_i$

unbiased

Next, solve for σ_{MLE}: $\quad \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{n}{\sigma} \quad \Rightarrow \sum_{i=1}^{n}(X_i-\mu)^2 = \sigma^2 n \quad \Rightarrow \sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_{MLE})^2$

biased!!

brand new content

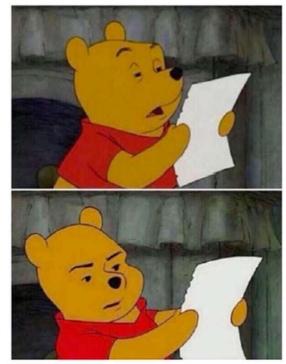
MLE: Multinomial

Okay, just one more MLE with the Multinomial

Consider a sample of n iid random variables where:

- Each element is drawn from one of m outcomes. $P(\text{outcome } i) = p_i$, where $\sum_{i=1}^{m} p_i = 1$ } this is the classic description of multinomial
- X_i = # of trials with outcome i, where $\sum_{i=1}^{m} X_i = n$

Staring at my math homework like



Let's give an example!