

# Lecture 21

## Method of Moments & Maximum Likelihood Estimation

---

STAT 330 - Iowa State University

1 / 20

### Estimating Parameters

#### 2 General Methods for estimating parameters:

1. Method of moments estimation (MoM)
2. Maximum likelihood estimation (MLE)

2 / 20

## Method of Moments (MoM)

### Method of Moments (MoM)

Definition: *also called population moment*

- The  $k^{th}$  *moment* of a R.V X is defined as  $\mu_k = E(X^k)$
- The  $k^{th}$  *sample moment* is defined as  $m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$

The **method of moments (MoM)** estimators for parameters are found by equating (known) sample moments to (unknown) population moments, and then solving for the parameters in terms of the data.

- If our model has more than one unknown parameter, we need to make equations with more than one moment.
- In general, need  $k$  equations to derive MoM estimators for  $k$  parameters.

## MOM Cont.

To obtain MoM estimators for  $k$  parameters: Set the sample moments ( $m_k$ ) equal to population moments ( $\mu_k$ ), and solve.

- $m_1 = \mu_1 \rightarrow \frac{1}{n} \sum x_i = E(X)$
- $m_2 = \mu_2 \rightarrow \frac{1}{n} \sum x_i^2 = E(X^2)$
- ⋮
- $m_k = \mu_k \rightarrow \frac{1}{n} \sum x_i^k = E(X^k)$

' **Note:**

- MoM estimators may be biased
- Sometimes you can get estimates outside of parameter space

## MoM Examples

## MoM Example

Independent &  
↓ identical

Example 1: Let  $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Geo}(p)$

Estimate one parameter  $p \rightarrow$  need the first moment.

- 1<sup>st</sup> (population) moment:  $\mu_1 = E(X) = \frac{1}{p}$ .
- 1<sup>st</sup> sample moment is  $m_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$

Set 1<sup>st</sup> moment equal 1<sup>st</sup> sample moment, and solve for  $p$ .

$$\frac{1}{p} = \bar{X} \rightarrow \hat{p}_{MoM} = \underbrace{\frac{1}{\bar{X}}}_{\substack{\text{MoM estimator} \\ \text{for } p}}$$

5 / 20

## MoM Examples Cont.

Example 2: Let  $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$

Estimate two parameters  $\rightarrow$  need first two moments

Set the first two moments equal to the first two sample moments.

1.  $\frac{1}{n} \sum_{i=1}^n X_i = E(X)$
2.  $\underbrace{\frac{1}{n} \sum_{i=1}^n X_i^2}_{\text{sample moments}} = \underbrace{E(X^2)}_{\text{moment}}$

For our random variables,  $E(X) = \mu$  and  $\text{Var}(X) = \sigma^2$

From Eq 1, we have  $\frac{1}{n} \sum X_i = E(X) = \mu$

$$\rightarrow \hat{\mu}_{MoM} = \frac{1}{n} \sum X_i$$

$$\rightarrow \hat{\mu}_{MoM} = \underbrace{\bar{X}}_{\substack{\text{MoM estimator} \\ \text{for } \mu}}$$

6 / 20

## MoM Examples Cont.

$$\text{Var}(X) = E(X^2) - E(X)^2 \rightarrow E(X^2) = \text{Var}(X) + E(X)^2 = \sigma^2 + \mu^2$$

From Eq 2. we have: ~~2<sup>nd</sup> sample moment~~ ~~2<sup>nd</sup> moment~~

$$\overbrace{\frac{1}{n} \sum_{i=1}^n X_i^2}^{\text{2<sup>nd</sup> sample moment}} = \overbrace{E(X^2)}^{\text{2<sup>nd</sup> moment}} = \sigma^2 + \mu^2$$

$$\rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \mu^2$$

$$\rightarrow \hat{\sigma}_{\text{MoM}}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}_{\text{MoM}}^2$$

$$= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$$

$$= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

*is actually the  
biased version  
of sample variance.*

7 / 20

## Maximum Likelihood Estimation (MLE)

## Likelihood Function

We have  $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x)$ , where  $f_X(x)$  has (unknown) parameter  $\theta$ .

The model for our data is the *joint distribution* of  $X_1, \dots, X_n$

$$f_X(x_1, \dots, x_n) = \prod_{i=1}^n f_X(x_i)$$

since  $X$ 's are indep  
joint = product  
of  
marginals

Sum  
function

When the joint distribution is viewed as a function of the unknown parameter, it is referred to as the *likelihood function*

$$\mathcal{L}(\theta) = \prod_{i=1}^n f_X(x_i)$$

8 / 20

## Likelihood Example

Example 3:  $X_1, \dots, X_n \stackrel{iid}{\sim} Pois(\lambda)$

The marginal distribution of each  $X_i$  is

$$f_X(x) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

pmf for a  
single  $\times$   
(marginal dist)

The joint distribution/likelihood function is

$$\begin{aligned}\mathcal{L}(\lambda) &= f(x_1, \dots, x_n) = \prod_{i=1}^n f_X(x_i) \\ &= \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\ &= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\end{aligned}$$

$$\begin{aligned}\prod_{i=1}^n e^{-\lambda} &= e^{-\sum_{i=1}^n \lambda} = e^{-n\lambda} \\ \prod_{i=1}^n x_i &= \lambda^{\sum_{i=1}^n x_i}\end{aligned}$$

9 / 20

## Maximum Likelihood Estimation (MLE)

### Definition

A *maximum likelihood estimator*  $\hat{\theta}_{MLE}$  of  $\theta$  is the function that "maximizes the likelihood (probability) of the data"

Thus, the MLE maximizes the joint distribution model or likelihood function:

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \mathcal{L}(\theta) = \underset{\theta \in \Theta}{\operatorname{argmax}} \prod_{i=1}^n f(x_i)$$

i.e.) find the  $\theta$  that makes  $L(\theta)$  the largest  
→ find  $\theta$  that is "most likely"

10 / 20

### MLE Examples

Example 4: Flip a coin 10 times. Let  $X$  be the # of heads obtained. A reasonable model for  $X$  is  $Bin(n = 10, p)$  where  $p$  is our unknown parameter that we would like to estimate.

Suppose we observe the value  $x = 3$ . (only 1 data value).

Since there's only 1 data value, the likelihood/joint distribution is just the marginal distribution  $f(x)$ :

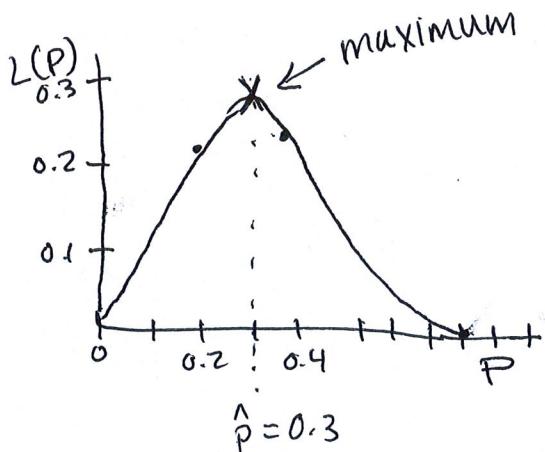
$$\begin{aligned}\mathcal{L}(p) = f(x) &= \binom{10}{x} p^x (1-p)^{10-x} \\ &= \binom{10}{3} p^3 (1-p)^{10-3} \\ &= 120p^3(1-p)^7\end{aligned}$$

*this is my likelihood function*  
 $l(p)$

11 / 20

## MLE Examples Cont.

What value of  $p$  maximizes the likelihood?



$$p = 0.2 \rightarrow L(0.2) = 0.201$$

$$p = 0.3 \rightarrow L(0.3) = 0.267$$

$$p = 0.4 \rightarrow L(0.4) = 0.215$$

$$p = 0.9 \rightarrow L(0.9) = 0.000008$$

Based on my data ( $x=3$ )

MLE for  $p$  is  $\hat{p}_{\text{MLE}} = 0.3$

"the most likely value for my unknown parameter  $p^{12/20}$ "

## General Calculation of MLE

- Maximizing the likelihood from  $L(\theta)$  when there are multiple observed values becomes difficult.  $L(\theta) = \prod_{i=1}^n f(x_i)$
- The common trick is to use the *log-likelihood function* instead:

why  
 - using logs  
 turns mult.  
 problem into  
 addition  
 . much easier  
 to maximize  
 $L(\theta) = a \cdot b \cdot c$

$$\ell(\theta) = \log L(\theta)$$

where  $\ell(\cdot)$  is the natural-log

→ Since  $\ell(\cdot)$  is increasing, the same  $\theta$  that maximizes log-likelihood  $\ell(\cdot)$  also maximizes the likelihood  $L(\theta)$

- Use calculus to find the maximum of  $\ell(\theta)$

$$\begin{aligned} \log L(\theta) &= \log(a \cdot b \cdot c) \\ &= \log(a) + \log(b) + \log(c) \end{aligned}$$

## General Calculation of MLE cont.

### Finding MLE:

1. Find the likelihood function:  $\mathcal{L}(\theta) = \prod_{i=1}^n f(x_i)$
2. Find the log-likelihood function:  $\ell(\theta) = \log \mathcal{L}(\theta)$
3. Take the first derivative:  $\ell'(\theta) = \frac{d}{d\theta} \ell(\theta)$
4. Set  $\ell'(\theta) = 0$  and solve for  $\theta$   
→ this is your  $\hat{\theta}_{MLE}$
5. Check if second derivative  $\ell''(\theta) < 0$  to make sure  $\hat{\theta}_{MLE}$  is maximum

14 / 20

## MLE Examples

## MLE Examples

Example 5: Roll a (6-sided) die until you get a 6, and record the number of rolls. Repeat for 100 trials. For  $i = 1, \dots, 100$ ,

$X_i = \# \text{ of rolls until you obtain a 6 in the } i^{\text{th}} \text{ trial}$

$X_i \stackrel{iid}{\sim} \text{Geo}(p)$  and  $f(x_i) = p(1-p)^{x_i-1}$

Data: Goal: estimate parameter  $P$ .

x	1	2	3	4	5	6	7	8	9
#	18	20	8	9	9	5	8	3	5
x	11	14	15	16	17	20	21	27	29
#	3	3	3	1	1	1	1	1	1

15 / 20

## MLE Examples Cont.

- Find the likelihood function  $\mathcal{L}(p)$ : (joint distribution)

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

- Find the log-likelihood function  $\ell(p) = \log \mathcal{L}(p)$ :

$$\ell(p) = \log \mathcal{L}(p)$$

$$\ell(p) = \log \mathcal{L}(p) = \log (p^n (1-p)^{\sum x_i - n})$$

$$= \log(p^n) + \log((1-p)^{\sum x_i - n})$$

$$= n \log(p) + (\sum x_i - n) \log(1-p)$$

log-likelihood function

recall  
 $\log(a^n) = n \log(a)$

$\log(a \cdot b) = \log a + \log b$

+ log b

16 / 20

## MLE Examples Cont.

3. Take the 1<sup>st</sup> derivative w.r.t  $p$ :  $\ell'(p)$ :

$$\begin{aligned}\ell'(p) &= \frac{d}{dp} \ell(p) = \frac{d}{dp} n \log(p) + \left( \sum_{i=1}^n x_i - n \right) \log(1-p) \\ &= \frac{n}{p} + \frac{(\sum x_i - n)}{1-p} (-1) \\ &= \frac{n}{p} - \frac{\sum x_i - n}{1-p}\end{aligned}$$

17 / 20

## MLE Example

4. Set  $\ell'(p) = 0$  and solve for  $p$ :

$$\begin{aligned}\frac{d}{dp} \ell(p) &\stackrel{\text{set}}{=} 0 \\ \Rightarrow \frac{n}{p} - \frac{\sum x_i - n}{1-p} &\stackrel{\text{set}}{=} 0 \\ \Rightarrow \frac{n}{p} &= \frac{\sum x_i - n}{1-p} \\ \Rightarrow \frac{1-p}{p} &= \frac{\sum x_i - n}{n} \\ \Rightarrow \frac{1}{p} - 1 &= \bar{x} - 1 \\ \Rightarrow \frac{1}{p} &= \bar{x} \\ \Rightarrow p &= \frac{1}{\bar{x}} \quad \Rightarrow \hat{p}_{\text{MLE}} = \frac{1}{\bar{x}}\end{aligned}$$

18 / 20

## MLE Examples Cont.

5. 2<sup>nd</sup> derivative test to confirm we have maximum:

$$\begin{aligned}
 \frac{d^2}{dp^2} \ell(p) &= \frac{d}{dp} \left[ \frac{d}{dp} \ell(p) \right] \Big|_{p=\hat{p}} \\
 \frac{d^2}{dp^2} \ell(p) &= \frac{d}{dp} \left[ \frac{n}{p} - \frac{\sum x_i - n}{1-p} \right] \Big|_{p=\hat{p}_{\text{mle}}} \\
 &= \frac{-n}{p^2} - \frac{(\sum x_i - n)(-1)}{(1-p)^2} \Big|_{p=\hat{p}_{\text{mle}}} \\
 &= \frac{-n}{p^2} - \frac{\sum x_i - n}{(1-p)^2} \Big|_{p=\hat{p}_{\text{mle}}} \\
 &= \frac{-n}{\hat{p}^2} - \frac{\sum x_i - n}{(1-\hat{p})^2} \stackrel{20}{<} 0
 \end{aligned}$$

$x \sim \text{Geo}(p)$

$\text{Im}(x) = \{1, 2, 3, \dots\}$

so  $\sum x_i$  is at least  $n \cdot 1 = n$

so  $\sum x_i - n \stackrel{19/20}{\geq} 0$

## MLE Example Cont.

So we have a maximum at  $\hat{p}_{\text{mle}}$

Hence  $\hat{p}_{\text{mle}} = \frac{1}{\bar{x}}$

Plug in the data into our MLE:

$$\begin{aligned}
 \hat{p}_{\text{mle}} &= \frac{1}{\bar{x}} = \frac{1}{5.68} \\
 &= 0.176
 \end{aligned}$$

Data

$$\sum x_i = 568$$

$$n = 100$$

$$\bar{x} = 5.68$$