Lecture 23

Method of Moments & Maximum Likelihood Estimation

Manju M. Johny

STAT 330 - Iowa State University

Estimating Parameters

2 General Methods for estimating parameters:

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

Method of Moments (MoM)

Method of Moments (MoM)

Definition:

- 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 (μ_k), and solve.

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

  $\vdots$
- $m_k = \mu_k \rightarrow \frac{1}{n}\sum x_i^k = E(X^k)$

' **Note:**

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

MoM Examples

MoM Example

Example 1: Let $X_1, \ldots, X_n \overset{iid}{\sim} Geo(p)$

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

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

Set 1^{st} moment equal 1^{st} sample moment, and solve for p.

$$\frac{1}{p} = \bar{X} \rightarrow \hat{p}_{MoM} = \frac{1}{\bar{X}}$$

MoM Examples Cont.

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

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

For our random variables, $E(X) = \mu$ and $Var(X) = \sigma^2$
From Eq 1, we have $\frac{1}{n}\sum X_i = E(X) = \mu$
$\rightarrow \hat{\mu}_{MoM} = \frac{1}{n}\sum X_i$
$\rightarrow \hat{\mu}_{MoM} = \bar{X}$

MoM Examples Cont.

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

From Eq 2. we have:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2) = \sigma^2 + \mu^2$$

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

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

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

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

Maximum Likelihood Estimation (MLE)

Likelihood Function

We have $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$, where $f_X(x)$ has (unknown) parameter θ.

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

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

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

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

Likelihood Example

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

The marginal distribution of each X_i is

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

The joint distribution/likelihood function is

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

Maximum Likelihood Estimation (MLE)

Definition

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

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

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

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 $f(x)$:

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

MLE Examples Cont.

What value of p maximizes the likelihood?

General Calculation of MLE

- Maximizing the likelihood from $L(\theta)$ when there are multiple observed values becomes difficult.
- The common trick is to use the *log-likelihood function* instead:

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

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

  $\rightarrow$ Since $\ell(\cdot)$ is increasing, the same θ that maximizes log-likelihood $\ell(\cdot)$ also maximizes the likelihood $\mathcal{L}(\theta)$
- Use calculus to find the maximum of $\ell(\theta)$

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 θ

   $\rightarrow$ this is your $\hat{\theta}_{MLE}$
5. Check if second derivative $\ell^{''}(\theta) < 0$ to make sure $\hat{\theta}_{MLE}$ is maximum

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 $i = 1, \dots 100$,

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

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

Data:

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

MLE Examples Cont.

1. Find the likelihood function $\mathcal{L}(p)$:

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

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

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

MLE Examples Cont.

3. Take the 1^{st} derivative w.r.t p: $\ell'(p)$:

$$\ell^i(p) = \frac{d}{dp}\ell(p) = \frac{d}{dp} nlog(p) + (\sum_{i=1}^{n} x_i - n)log(1-p)$$

MLE Example

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

$$\frac{d}{dp}\ell(p) \overset{set}{=} 0$$

MLE Examples Cont.

5. 2^{nd} derivative test to confirm we have maximum:
   $\frac{d^2}{dp^2}\ell(p)$

Plug in the data into our MLE: