ASDS Statistics, YSU, Fall 2020

Lecture 25

Michael Poghosyan

28 Nov 2020

Contents

- The Method of Maximum Likelihood Estimation

The Maximum Likelihood Method

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$.

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times.

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times. Let the outcome be

$$HHHHHHH.$$

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times. Let the outcome be

$$HHHHHHH.$$

What is your best guess for p?

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times. Let the outcome be

$$HHHHHHH.$$

What is your best guess for p?

Well, of course, you are correct, best guess is $p = 1$.

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times. Let the outcome be

$$HHHHHHH.$$

What is your best guess for p?

Well, of course, you are correct, best guess is $p = 1$. But it is possible also that this outcome is obtained from a coin with $p = 0.9$.

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times. Let the outcome be

$$HHHHHHH.$$

What is your best guess for p?

Well, of course, you are correct, best guess is $p = 1$. But it is possible also that this outcome is obtained from a coin with $p = 0.9$. Or with $p = 0.8$.

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times. Let the outcome be

$$HHHHHHH.$$

What is your best guess for p?

Well, of course, you are correct, best guess is $p = 1$. But it is possible also that this outcome is obtained from a coin with $p = 0.9$. Or with $p = 0.8$. Even with $p = 0.2$.

Example

Example: Assume we have a coin. The Probability of *Heads* is $p \in [0, 1]$. We toss that 7 times. Let the outcome be

$$HHHHHHH.$$

What is your best guess for p?

Well, of course, you are correct, best guess is $p = 1$. But it is possible also that this outcome is obtained from a coin with $p = 0.9$. Or with $p = 0.8$. Even with $p = 0.2$.

Ok, let's do some calculations.

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) =$$

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.9)^7 \approx 0.48$$

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.9)^7 \approx 0.48$$

And what if $p = 0.8$?

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.9)^7 \approx 0.48$$

And what if $p = 0.8$?

$$\text{If } p = 0.8, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.8)^7 \approx 0.21$$

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.9)^7 \approx 0.48$$

And what if $p = 0.8$?

$$\text{If } p = 0.8, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.8)^7 \approx 0.21$$

Of course, we could have also the above outcome if $p = 0.2$?

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.9)^7 \approx 0.48$$

And what if $p = 0.8$?

$$\text{If } p = 0.8, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.8)^7 \approx 0.21$$

Of course, we could have also the above outcome if $p = 0.2$? But the chances are

$$\text{If } p = 0.2, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.2)^7 \approx 1.28e{-}05 = 1.28 \cdot 10^{-5}$$

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.9)^7 \approx 0.48$$

And what if $p = 0.8$?

$$\text{If } p = 0.8, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.8)^7 \approx 0.21$$

Of course, we could have also the above outcome if $p = 0.2$? But the chances are

$$\text{If } p = 0.2, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.2)^7 \approx 1.28e{-}05 = 1.28 \cdot 10^{-5}$$

And, of course, if $p = 1$, then

$$\text{If } p = 1, \text{ then} \quad \mathbb{P}(HHHHHHH) = 1^7 = 1.$$

Example, Cont'd

Assume $p = 0.9$. What is the Probabilty to obtain the above outcome?

$$\text{If } p = 0.9, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.9)^7 \approx 0.48$$

And what if $p = 0.8$?

$$\text{If } p = 0.8, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.8)^7 \approx 0.21$$

Of course, we could have also the above outcome if $p = 0.2$? But the chances are

$$\text{If } p = 0.2, \text{ then} \quad \mathbb{P}(HHHHHHH) = (0.2)^7 \approx 1.28e{-}05 = 1.28 \cdot 10^{-5}$$

And, of course, if $p = 1$, then

$$\text{If } p = 1, \text{ then} \quad \mathbb{P}(HHHHHHH) = 1^7 = 1.$$

So our guess was to **select the value of p giving the highest likelihood to our outcome**.

Idea of the Maximum Likelihood Method

Assume we have a Parametric Family of Distributions $\mathcal{F}_\theta$ with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$.

Idea of the Maximum Likelihood Method

Assume we have a Parametric Family of Distributions $\mathcal{F}_\theta$ with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$. We take a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

and want to use it to construct a good Estimator for θ.

Idea of the Maximum Likelihood Method

Assume we have a Parametric Family of Distributions $\mathcal{F}_\theta$ with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$. We take a Random Sample

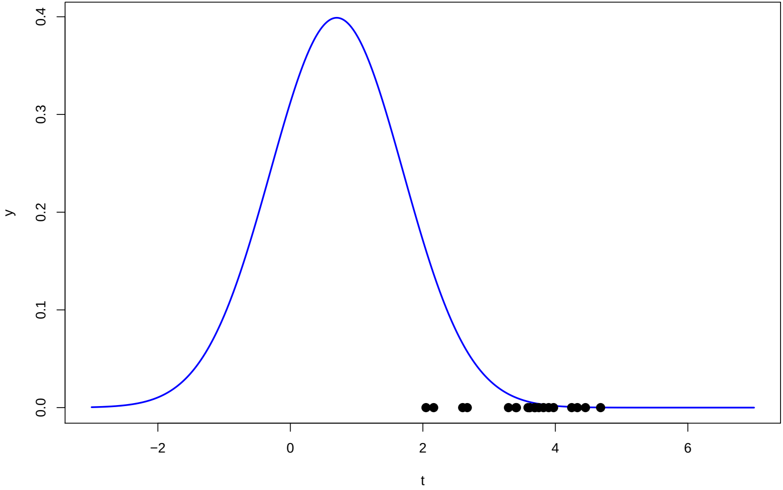
$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

and want to use it to construct a good Estimator for θ.

Idea of Maximum Likelihood Estimation: We choose that value of our parameter, under which **our Observation is the most Probable**.

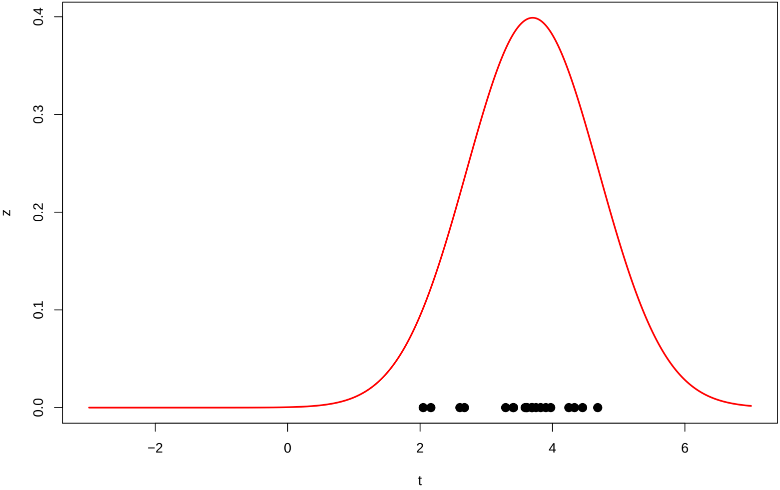
Idea of the MLE

Guess which Distribution makes our Data most Probable:



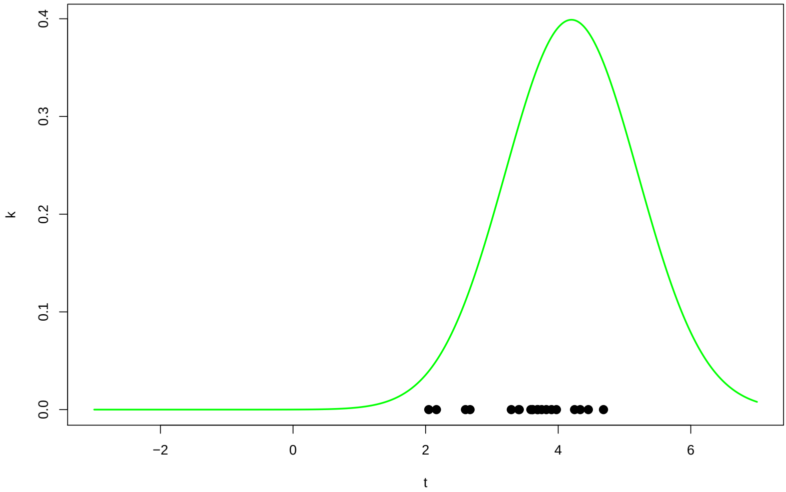
Idea of the MLE

Guess which Distribution makes our Data most Probable:



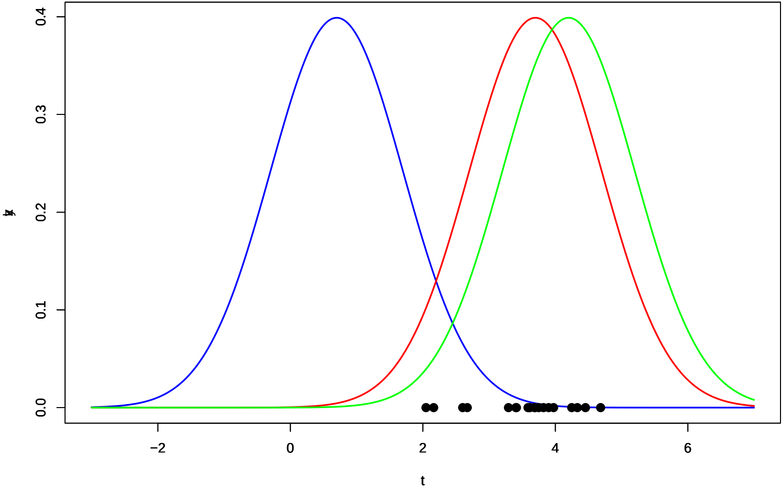
Idea of the MLE

Guess which Distribution makes our Data most Probable:



Idea of the MLE

Guess which Distribution makes our Data most Probable:



Problem Statement Again

Again, assume we have an Observation $x: \ x_1, ..., x_n$, from one of the Distributions of Parametric Family $\mathcal{F}_\theta$, with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$.

Problem Statement Again

Again, assume we have an Observation $x: \ x_1, ..., x_n$, from one of the Distributions of Parametric Family $\mathcal{F}_\theta$, with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$. Our aim is to Estimate θ.

Problem Statement Again

Again, assume we have an Observation $x: \ x_1, ..., x_n$, from one of the Distributions of Parametric Family $\mathcal{F}_\theta$, with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$. Our aim is to Estimate θ.

We, instead of our Observation, take a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to *generalize*, to have our method work also for *unseen Data*, to get a result for all possible Observations,

Problem Statement Again

Again, assume we have an Observation $x: \ x_1, ..., x_n$, from one of the Distributions of Parametric Family $\mathcal{F}_\theta$, with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$. Our aim is to Estimate θ.

We, instead of our Observation, take a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to *generalize*, to have our method work also for *unseen Data*, to get a result for all possible Observations, i.e., to construct an **Estimator** for θ.

Problem Statement Again

Again, assume we have an Observation $x:\ x_1, ..., x_n$, from one of the Distributions of Parametric Family $\mathcal{F}_\theta$, with the PD(M)F $f(x|\theta)$, $\theta \in \Theta$. Our aim is to Estimate θ.

We, instead of our Observation, take a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to *generalize*, to have our method work also for *unseen Data*, to get a result for all possible Observations, i.e., to construct an **Estimator** for θ.

And the Maximum Likelihood Method is saying: **choose that value of θ, under which it is most likely to get** $X_1, X_2, ..., X_n$.

Likelihood

Definition: The **Likelihood Function** for the above Model and Random Sample is the Joint PD(M)F of $X_1, \ldots, X_n$, **considered as a function of the parameter** θ, and **calculated at the Random Sample**, i.e., it is given by[1]

$$\mathcal{L}(\theta) = \mathcal{L}_n(X_1, \ldots, X_n|\theta) = f(X_1|\theta) \cdot f(X_2|\theta) \cdot \ldots \cdot f(X_n|\theta), \qquad \theta \in \Theta.$$

[1] Since X_k-s are independent

Likelihood

Definition: The **Likelihood Function** for the above Model and Random Sample is the Joint PD(M)F of $X_1, ..., X_n$, **considered as a function of the parameter** θ, and **calculated at the Random Sample**, i.e., it is given by[1]

$$\mathcal{L}(\theta) = \mathcal{L}_n(X_1, ..., X_n|\theta) = f(X_1|\theta) \cdot f(X_2|\theta) \cdot ... \cdot f(X_n|\theta), \qquad \theta \in \Theta.$$

The **Log-Likelihood Function** is the function

$$\ell(\theta) = \ell(X_1, ..., X_n|\theta) = \ln \mathcal{L}(\theta) = \sum_{k=1}^{n} \ln f(X_k|\theta), \qquad \theta \in \Theta.$$

[1] Since X_k-s are independent

Likelihood

Definition: The **Likelihood Function** for the above Model and Random Sample is the Joint PD(M)F of $X_1, ..., X_n$, **considered as a function of the parameter** θ, and **calculated at the Random Sample**, i.e., it is given by[1]

$$\mathcal{L}(\theta) = \mathcal{L}_n(X_1, ..., X_n|\theta) = f(X_1|\theta) \cdot f(X_2|\theta) \cdot ... \cdot f(X_n|\theta), \qquad \theta \in \Theta.$$

The **Log-Likelihood Function** is the function

$$\ell(\theta) = \ell(X_1, ..., X_n|\theta) = \ln \mathcal{L}(\theta) = \sum_{k=1}^{n} \ln f(X_k|\theta), \qquad \theta \in \Theta.$$

Also we define the **Negative Log-Likelihood Function** to be

$$-\ell(\theta) = -\ln \mathcal{L}(\theta).$$

[1]Since X_k-s are independent

Maximum Likelihood Method

Note: Likelihood is not a Probability - it can be larger than 1.

Maximum Likelihood Method

Note: Likelihood is not a Probability - it can be larger than 1. It is not a PDF either, it is a **function of the parameter** θ. Say, the integral of Likelihood over all possible θ-s can be different than 1.

Maximum Likelihood Method

Note: Likelihood is not a Probability - it can be larger than 1. It is not a PDF either, it is a **function of the parameter** θ. Say, the integral of Likelihood over all possible θ-s can be different than 1.

Now, the Maximum Likelihood Method suggests **to find a point that makes our Likelihood Maximal**:

Maximum Likelihood Method

Note: Likelihood is not a Probability - it can be larger than 1. It is not a PDF either, it is a **function of the parameter** θ. Say, the integral of Likelihood over all possible θ-s can be different than 1.

Now, the Maximum Likelihood Method suggests **to find a point that makes our Likelihood Maximal**:

Definition: The **Maximum Likelihood Estimator (MLE)** of the parameter θ is the value of θ that maximizes the Likelihood function for the given random sample $X_1, ..., X_n$, the global maximum point (in case it exists) of $\mathcal{L}(X_1, ..., X_n|\theta)$:

$$\hat{\theta}^{MLE} = \hat{\theta}_n^{MLE} = \underset{\theta \in \Theta}{argmax}\, \mathcal{L}(\theta).$$

Maximum Likelihood Method

Note: Likelihood is not a Probability - it can be larger than 1. It is not a PDF either, it is a **function of the parameter** θ. Say, the integral of Likelihood over all possible θ-s can be different than 1.

Now, the Maximum Likelihood Method suggests **to find a point that makes our Likelihood Maximal**:

Definition: The **Maximum Likelihood Estimator (MLE)** of the parameter θ is the value of θ that maximizes the Likelihood function for the given random sample $X_1, ..., X_n$, the global maximum point (in case it exists) of $\mathcal{L}(X_1, ..., X_n|\theta)$:

$$\hat{\theta}^{MLE} = \hat{\theta}_n^{MLE} = \underset{\theta \in \Theta}{argmax}\, \mathcal{L}(\theta).$$

And in the case if we have an Observation $x: \ x_1, x_2, ..., x_n$ from the above Model (from one of the Distributions of that Model), the **Maximum Likelihood Estimate** (again **MLE**) of the parameter θ is the value of $\hat{\theta}^{MLE}$ on our Observation.

MLE, Notes

Note: argmax means the **Argument of the Maximum**, the point(s) of the Maximum. In our case, **Global Max Point(s)**.

MLE, Notes

Note: argmax means the **Argument of the Maximum**, the point(s) of the Maximum. In our case, **Global Max Point(s)**.

Note: To find the **Maximum Likelihood Estimate** for θ, you can do the following steps:

- Either find the **Maximum Likelihood Estimator** for θ, and then plug the Observation values;

MLE, Notes

Note: argmax means the **Argument of the Maximum**, the point(s) of the Maximum. In our case, **Global Max Point(s)**.

Note: To find the **Maximum Likelihood Estimate** for θ, you can do the following steps:

- Either find the **Maximum Likelihood Estimator** for θ, and then plug the Observation values;
- Or first plug the Observation values into the Likelihood function, to get
$$\mathcal{L}(x_1, ..., x_n|\theta),$$
and then find the maximum point for this function, over $\theta \in \Theta$.

MLE, Notes

Note: argmax means the **Argument of the Maximum**, the point(s) of the Maximum. In our case, **Global Max Point(s)**.

Note: To find the **Maximum Likelihood Estimate** for θ, you can do the following steps:

- Either find the **Maximum Likelihood Estimator** for θ, and then plug the Observation values;
- Or first plug the Observation values into the Likelihood function, to get
$$\mathcal{L}(x_1, ..., x_n|\theta),$$
and then find the maximum point for this function, over $\theta \in \Theta$.

Note: Since the function $h(t) = \ln t$ is strictly increasing, we will have that

$$\underset{\theta \in \Theta}{argmax}\, \mathcal{L}(\theta) = \underset{\theta \in \Theta}{argmax} \ln \mathcal{L}(\theta) = \underset{\theta \in \Theta}{argmax}\, \ell(\theta),$$

i.e., the points of maximum of $\mathcal{L}(\theta)$ and $\ln \mathcal{L}(\theta)$ coincide.

MLE, Notes

Note: argmax means the **Argument of the Maximum**, the point(s) of the Maximum. In our case, **Global Max Point(s)**.

Note: To find the **Maximum Likelihood Estimate** for θ, you can do the following steps:

- Either find the **Maximum Likelihood Estimator** for θ, and then plug the Observation values;
- Or first plug the Observation values into the Likelihood function, to get
$$\mathcal{L}(x_1, ..., x_n|\theta),$$
and then find the maximum point for this function, over $\theta \in \Theta$.

Note: Since the function $h(t) = \ln t$ is strictly increasing, we will have that

$$\underset{\theta\in\Theta}{argmax}\, \mathcal{L}(\theta) = \underset{\theta\in\Theta}{argmax} \ln \mathcal{L}(\theta) = \underset{\theta\in\Theta}{argmax}\, \ell(\theta),$$

i.e., the points of maximum of $\mathcal{L}(\theta)$ and $\ln \mathcal{L}(\theta)$ coincide. And, in the rest, we will find the Max points of the **Log-Likelihd** function.

Examples

Example: Find the MLE for p in the *Bernoulli*(p) Model.

Solution: OTB

Examples

Example: Find the MLE for p in the *Bernoulli*(p) Model.

Solution: OTB

Example: Find the MLE Estimator for λ in the *Exp*(λ) Model.

Solution: OTB

Examples

Example: Find the MLE for p in the $Bernoulli(p)$ Model.

Solution: OTB

Example: Find the MLE Estimator for λ in the $Exp(\lambda)$ Model.

Solution: OTB

Examples

Example: Find the MLE Estimator for θ in the $\mathit{Unif}\,[0,\theta]$ Model.

Solution: OTB

Examples

Example: Find the MLE Estimator for θ in the *Unif* $[0, \theta]$ Model.

Solution: OTB

Example: Find the MLE Estimator for (μ, σ^2) in the $\mathcal{N}(\mu, \sigma^2)$ Model.

Solution: OTB

Examples

Example: Find the MLE Estimator for θ in the *Unif* $[0, \theta]$ Model.

Solution: OTB

Example: Find the MLE Estimator for (μ, σ^2) in the $\mathcal{N}(\mu, \sigma^2)$ Model.

Solution: OTB

Example: Assume we have an observation

$$0, 1, 1, 2, 1, 0, 0, 1, 1$$

from the following Model:

X	0	1	2
$\mathbb{P}(X = x)$	$\frac{\theta}{10}$	$\frac{\theta}{5}$	$1 - \frac{3\theta}{10}$,

where $\theta \in [0, \frac{10}{3}]$.

Examples

Example: Find the MLE Estimator for θ in the *Unif* $[0, \theta]$ Model.

Solution: OTB

Example: Find the MLE Estimator for (μ, σ^2) in the $\mathcal{N}(\mu, \sigma^2)$ Model.

Solution: OTB

Example: Assume we have an observation

$$0, 1, 1, 2, 1, 0, 0, 1, 1$$

from the following Model:

X	0	1	2
$\mathbb{P}(X = x)$	$\frac{\theta}{10}$	$\frac{\theta}{5}$	$1 - \frac{3\theta}{10}$,

where $\theta \in [0, \frac{10}{3}]$. Find the MLE Estimator and MLE Estimate for θ.

Solution: OTB