AUA CS 108, Statistics, Fall 2019

Lecture 28

Michael Poghosyan
YSU, AUA
michael@ysu.am, mpoghosyan@aua.am

28 Oct 2019

Contents

Last Lecture ReCap

- How to estimate 3D unknown Parameter using MoM?

Last Lecture ReCap

- How to estimate 3D unknown Parameter using MoM?
- Give some reasons to use MoM.

The Maximum Likelihood Method

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

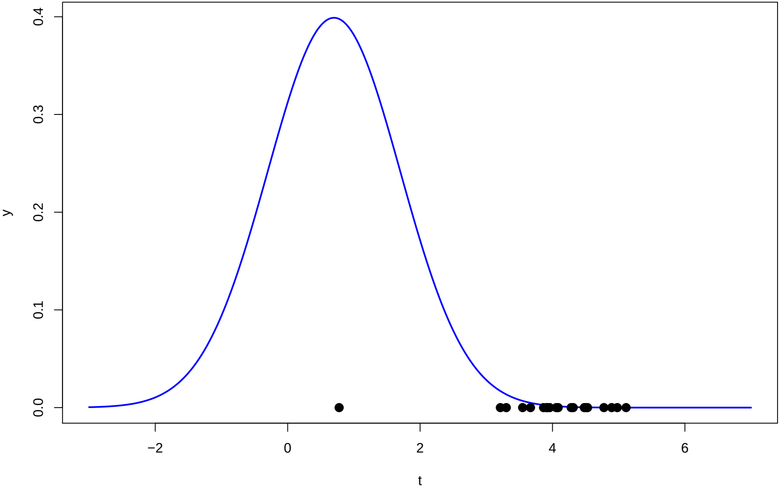
$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

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**.

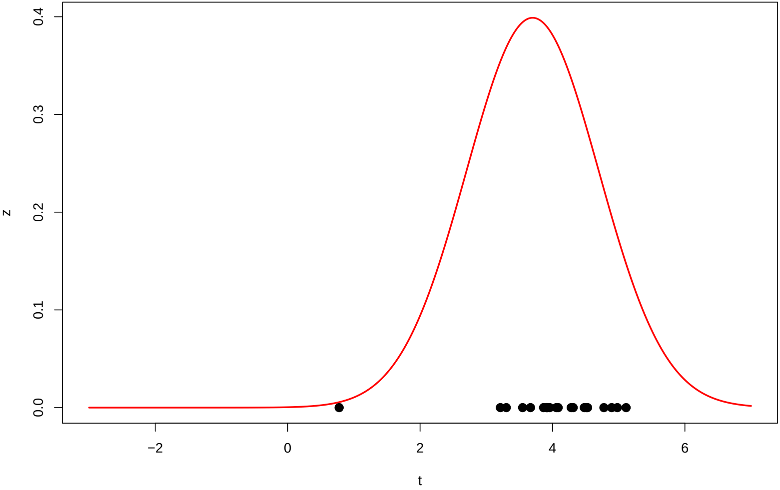
Idea of the MLE

Guess which Distribution makes our Data most Probable:



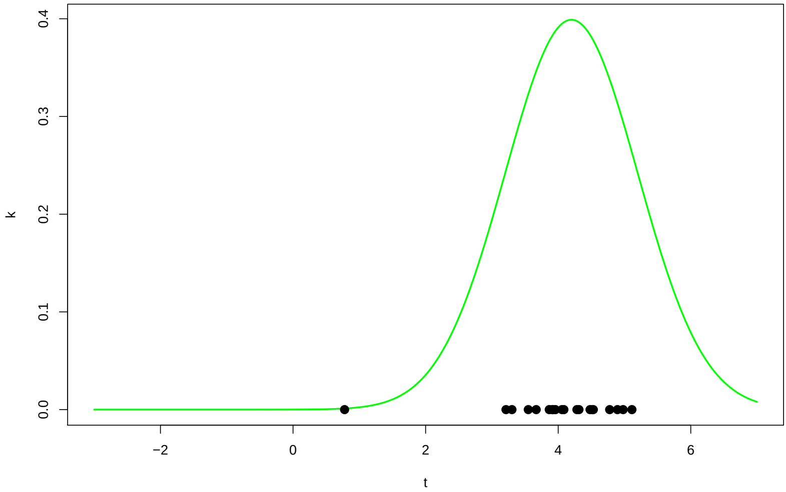
Idea of the MLE

Guess which Distribution makes our Data most Probable:



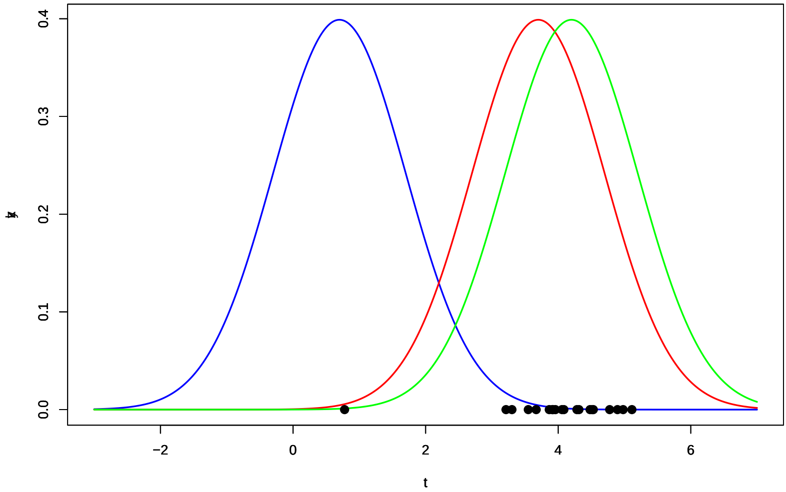
Idea of the MLE

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 $X_1, ..., X_n$, **considered as a function of the parameter** θ, and **calculated at the Random Sample**, i.e., it is given by[1]

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

[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.

Calc 1 + Calc 3 Refresher

Here it is desirable to have a slide about how to find the maximum points of a function $\ell(\theta)$ for $\theta \in \Theta$, considering:

- 1D Case
- *n*-D Case
- Sufficient Conditions.

Calc 1 + Calc 3 Refresher

Here it is desirable to have a slide about how to find the maximum points of a function $\ell(\theta)$ for $\theta \in \Theta$, considering:

- 1D Case
- *n*-D Case
- Sufficient Conditions.

I know that you can fill this slide, so I am keeping it to you*.

* In fact, I realized that one slide will not be enough, and was lazy to prepare them ⌣̈

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*(λ) Model.

Solution: OTB

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

Solution: OTB