AUA CS108, Statistics, Fall 2020

Lecture 33

Michael Poghosyan

11 nov 2020

Contents

- Method of Moments (MoM, MME)
- Maximum Likelihood Method (MLE)

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate.

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

The Method of Moments, m-D case

Now assume that our Parameter θ is *m*-Dimensional, or, in other way, we have *m* Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the *m*-D case.

So assume $\theta = (\theta_1, \theta_2)$.

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

So assume $\theta = (\theta_1, \theta_2)$. Then we need 2 Equations to find θ_1 and θ_2.

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

So assume $\theta = (\theta_1, \theta_2)$. Then we need 2 Equations to find θ_1 and θ_2. The MoM says: first try to solve the following system:

$$\begin{cases} \textbf{1-st order Theoretical Moment} = \textbf{1-st order Empirical Moment} \\ \textbf{2-nd order Theoretical Moment} = \textbf{2-nd order Empirical Moment} \end{cases}$$

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

So assume $\theta = (\theta_1, \theta_2)$. Then we need 2 Equations to find θ_1 and θ_2. The MoM says: first try to solve the following system:

$$\begin{cases} \textbf{1-st order Theoretical Moment} = \textbf{1-st order Empirical Moment} \\ \textbf{2-nd order Theoretical Moment} = \textbf{2-nd order Empirical Moment} \end{cases}$$

The LHS of these equations, in general, will depend on θ_1 and θ_2.

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

So assume $\theta = (\theta_1, \theta_2)$. Then we need 2 Equations to find θ_1 and θ_2. The MoM says: first try to solve the following system:

$$\begin{cases} \textbf{1-st order Theoretical Moment} = \textbf{1-st order Empirical Moment} \\ \textbf{2-nd order Theoretical Moment} = \textbf{2-nd order Empirical Moment} \end{cases}$$

The LHS of these equations, in general, will depend on θ_1 and θ_2. If we can find from here θ_1 and θ_2, we take them as the MoM Estimators.

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

So assume $\theta = (\theta_1, \theta_2)$. Then we need 2 Equations to find θ_1 and θ_2. The MoM says: first try to solve the following system:

$$\begin{cases} \textbf{1-st order Theoretical Moment} = \textbf{1-st order Empirical Moment} \\ \textbf{2-nd order Theoretical Moment} = \textbf{2-nd order Empirical Moment} \end{cases}$$

The LHS of these equations, in general, will depend on θ_1 and θ_2. If we can find from here θ_1 and θ_2, we take them as the MoM Estimators. Otherwise, if we cannot express from this system θ_1 or θ_2, then we try to solve another pair of this kind of equations, using the possible smallest order Moments.

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

So assume $\theta = (\theta_1, \theta_2)$. Then we need 2 Equations to find θ_1 and θ_2. The MoM says: first try to solve the following system:

$$\begin{cases} \textbf{1-st order Theoretical Moment} = \textbf{1-st order Empirical Moment} \\ \textbf{2-nd order Theoretical Moment} = \textbf{2-nd order Empirical Moment} \end{cases}$$

The LHS of these equations, in general, will depend on θ_1 and θ_2. If we can find from here θ_1 and θ_2, we take them as the MoM Estimators. Otherwise, if we cannot express from this system θ_1 or θ_2, then we try to solve another pair of this kind of equations, using the possible smallest order Moments. Say, if the 1st Moment equation is not giving a result, solve the system with 2nd and 3rd Moments.

The Method of Moments, m-D case

Now assume that our Parameter θ is m-Dimensional, or, in other way, we have m Parameters to Estimate. Let me give the 2D case, and you will be able to extend to the m-D case.

So assume $\theta = (\theta_1, \theta_2)$. Then we need 2 Equations to find θ_1 and θ_2. The MoM says: first try to solve the following system:

$$\begin{cases} \textbf{1-st order Theoretical Moment} = \textbf{1-st order Empirical Moment} \\ \textbf{2-nd order Theoretical Moment} = \textbf{2-nd order Empirical Moment} \end{cases}$$

The LHS of these equations, in general, will depend on θ_1 and θ_2. If we can find from here θ_1 and θ_2, we take them as the MoM Estimators. Otherwise, if we cannot express from this system θ_1 or θ_2, then we try to solve another pair of this kind of equations, using the possible smallest order Moments. Say, if the 1st Moment equation is not giving a result, solve the system with 2nd and 3rd Moments. Or, if in the initial system 2nd order Moment equation is not saying anything, use 1st and 3rd Moments, etc.

Example

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

Solution: OTB

Notes

Note: If you do not want to obtain the *MoM Estimator*, but just the *MoM Estimate*, just plug the Observations into the Empirical Moments, and then solve the Moments Equality Equations.

Notes

Note: If you do not want to obtain the *MoM Estimator*, but just the *MoM Estimate*, just plug the Observations into the Empirical Moments, and then solve the Moments Equality Equations.

Note: Sometimes we are interested not in the Estimate of the Parameter itself, but in the Estimate for some function of that Parameter.

Notes

Note: If you do not want to obtain the *MoM Estimator*, but just the *MoM Estimate*, just plug the Observations into the Empirical Moments, and then solve the Moments Equality Equations.

Note: Sometimes we are interested not in the Estimate of the Parameter itself, but in the Estimate for some function of that Parameter. Say, we want to Estimate $h(\theta)$, where θ is our unknown Parameter.

Notes

Note: If you do not want to obtain the *MoM Estimator*, but just the *MoM Estimate*, just plug the Observations into the Empirical Moments, and then solve the Moments Equality Equations.

Note: Sometimes we are interested not in the Estimate of the Parameter itself, but in the Estimate for some function of that Parameter. Say, we want to Estimate $h(\theta)$, where θ is our unknown Parameter. Then one of the approaches is the Plug-in Principle: find an Estimator $\hat{\theta}$ for θ, say, using the MoM, and then plug that in h, to obtain $h(\hat{\theta})$ as an Estimator for $h(\theta)$.

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Assume that for the Model $\mathcal{F}_\theta$, the Expectation $\mathbb{E}_\theta(X)$ $(X \sim \mathcal{F}_\theta)$, as a function of θ, has a Continuous Inverse function.

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Assume that for the Model $\mathcal{F}_\theta$, the Expectation $\mathbb{E}_\theta(X)$ ($X \sim \mathcal{F}_\theta$), as a function of θ, has a Continuous Inverse function. Say, if we denote $e(\theta) = \mathbb{E}_\theta(X)$, then we assume that the function $e^{-1}(\cdot)$ is continuous.

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Assume that for the Model $\mathcal{F}_\theta$, the Expectation $\mathbb{E}_\theta(X)$ ($X \sim \mathcal{F}_\theta$), as a function of θ, has a Continuous Inverse function. Say, if we denote $e(\theta) = \mathbb{E}_\theta(X)$, then we assume that the function $e^{-1}(\cdot)$ is continuous. Then the MoM Estimator for θ will be Consistent.

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Assume that for the Model $\mathcal{F}_\theta$, the Expectation $\mathbb{E}_\theta(X)$ ($X \sim \mathcal{F}_\theta$), as a function of θ, has a Continuous Inverse function. Say, if we denote $e(\theta) = \mathbb{E}_\theta(X)$, then we assume that the function $e^{-1}(\cdot)$ is continuous. Then the MoM Estimator for θ will be Consistent.

Indeed, to find the MoM Estimator for θ, we need to solve

$$e(\theta) = \overline{X}_n,$$

the solution of which we denote by $\hat{\theta}_n$.

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Assume that for the Model $\mathcal{F}_\theta$, the Expectation $\mathbb{E}_\theta(X)$ ($X \sim \mathcal{F}_\theta$), as a function of θ, has a Continuous Inverse function. Say, if we denote $e(\theta) = \mathbb{E}_\theta(X)$, then we assume that the function $e^{-1}(\cdot)$ is continuous. Then the MoM Estimator for θ will be Consistent.

Indeed, to find the MoM Estimator for θ, we need to solve

$$e(\theta) = \overline{X}_n,$$

the solution of which we denote by $\hat{\theta}_n$. This gives

$$\hat{\theta}_n = e^{-1}(\overline{X}_n).$$

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Assume that for the Model $\mathcal{F}_\theta$, the Expectation $\mathbb{E}_\theta(X)$ ($X \sim \mathcal{F}_\theta$), as a function of θ, has a Continuous Inverse function. Say, if we denote $e(\theta) = \mathbb{E}_\theta(X)$, then we assume that the function $e^{-1}(\cdot)$ is continuous. Then the MoM Estimator for θ will be Consistent.

Indeed, to find the MoM Estimator for θ, we need to solve

$$e(\theta) = \overline{X}_n,$$

the solution of which we denote by $\hat{\theta}_n$. This gives

$$\hat{\theta}_n = e^{-1}(\overline{X}_n).$$

Now, by the WLLN, $\overline{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}(X_1) = e(\theta)$,

Consistency of the MoM Estimator

Let's consider the case when we use the 1st order Moments.

Assume that for the Model $\mathcal{F}_\theta$, the Expectation $\mathbb{E}_\theta(X)$ ($X \sim \mathcal{F}_\theta$), as a function of θ, has a Continuous Inverse function. Say, if we denote $e(\theta) = \mathbb{E}_\theta(X)$, then we assume that the function $e^{-1}(\cdot)$ is continuous. Then the MoM Estimator for θ will be Consistent.

Indeed, to find the MoM Estimator for θ, we need to solve

$$e(\theta) = \overline{X}_n,$$

the solution of which we denote by $\hat{\theta}_n$. This gives

$$\hat{\theta}_n = e^{-1}(\overline{X}_n).$$

Now, by the WLLN, $\overline{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}(X_1) = e(\theta)$, so

$$\hat{\theta}_n = e^{-1}(\overline{X}_n) \xrightarrow{\mathbb{P}} e^{-1}(e(\theta)) = \theta.$$

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

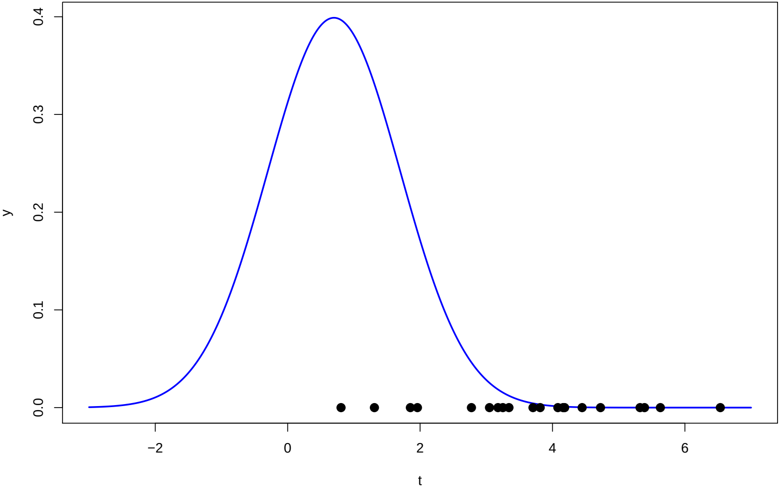
$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

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**.

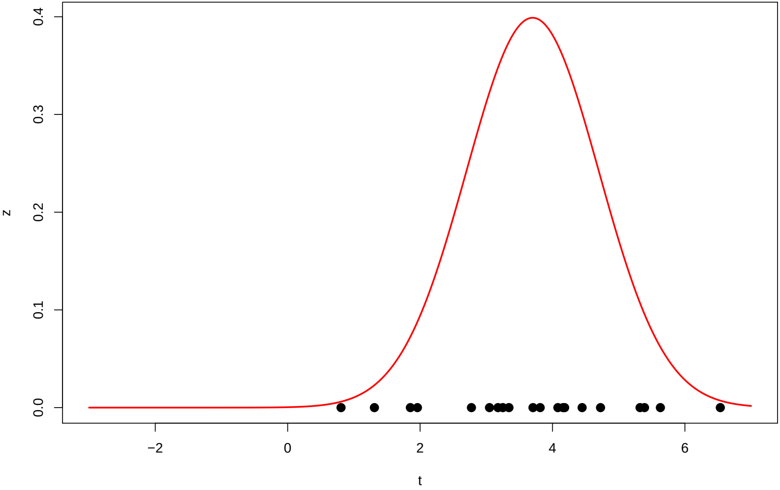
Idea of the MLE

Guess which Distribution makes our Data most Probable:



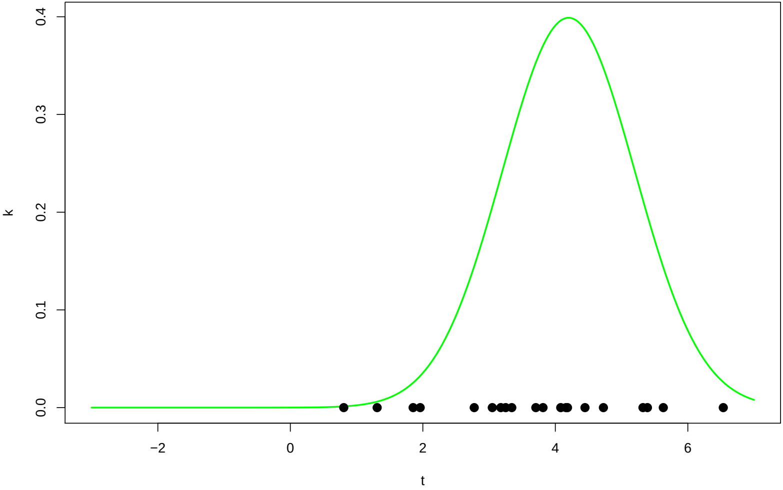
Idea of the MLE

Guess which Distribution makes our Data most Probable:



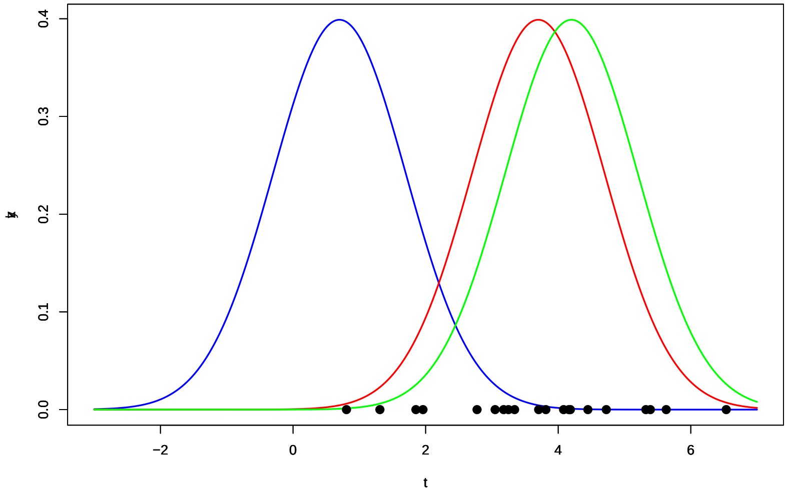
Idea of the MLE

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 important for you to refresh you knowledge from Calc1 + Calc3 about how to find the maximum points of a function $\ell(\theta)$ for $\theta \in \Theta$, considering:

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