2024-1

Multivariate Data Analysis

Youjin Shin
(yj.shinn@catholic.ac.kr)

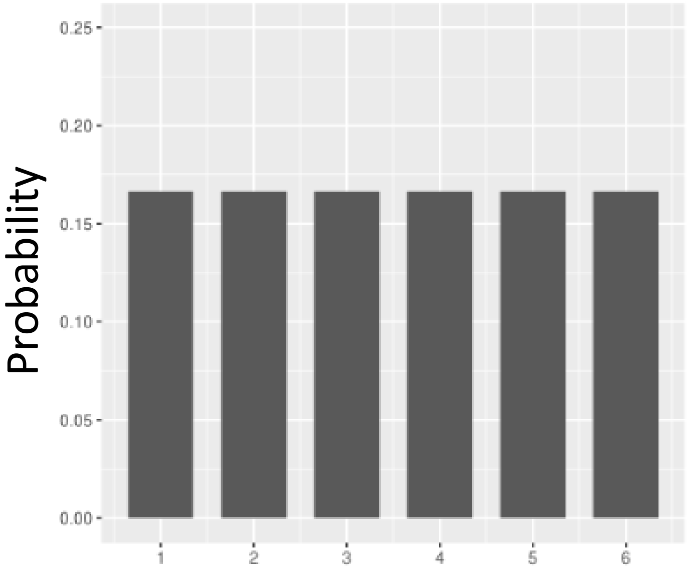
Probability vs. Likelihood

Probability

◆ Roll a dice vs. Toss a coin

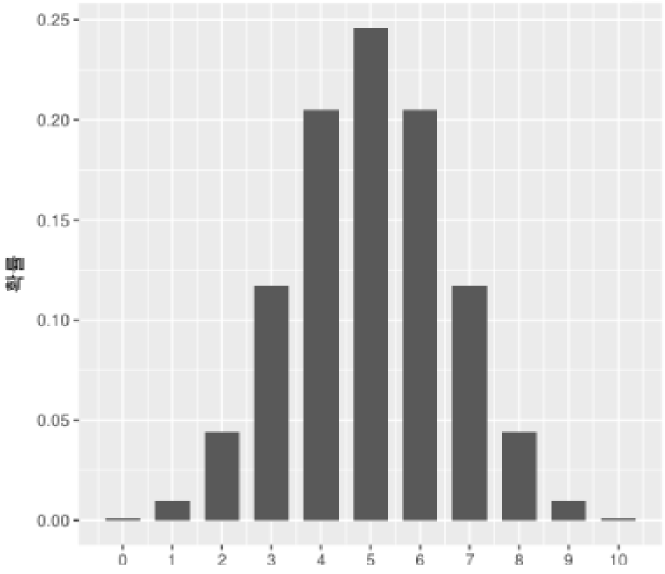
- **When you throw a dice**, the possible outcomes are 1, 2, 3, 4, 5, and 6, with each number having an equal probability of 1/6.
- **When you toss a coin 10 times**, the number of times heads can appear ranges from 0 to 10, with the respective probabilities being 0.001, 0.01, 0.044, 0.117, 0.205, 0.246, 0.205, 0.117, 0.044, 0.01, and 0.001.
- In both scenarios, there are 6 and 11 possible events, respectively, and the probability of each can be calculated, with the sum of the probabilities equaling 1.

Roll a dice



"pips" of the dice

Toss a coin 10 times



the number of getting heads

Probability

◆ Toss a coin : binomial distribution

- A discrete probability distribution that models the number of successes in a fixed number of independent trials of a binary experiment.
- A binary experiment is one that has exactly two outcomes, commonly referred to as "success" and "failure." In the context of the binomial distribution, the term "success" doesn't necessarily mean a good outcome; it just represents one of the two possible outcomes.
- $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

이항 분포는 고정된 수의 독립적인 베르누이 시행에서 성공 횟수를 모델링합니다. 성공 확률이 p이고 시행 횟수가 n일 때, k번 성공할 확률은 다음과 같이 주어집니다:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

where:

- $P(X=k)$ is the probability of getting exactly k successes.
- $\binom{n}{k}$ is the binomial coefficient, calculated as $\frac{n!}{k!(n-k)!}$, which gives the number of ways k successes can occur in n trials.
- p is the probability of success on a single trial.
- $1-p$ is the probability of failure on a single trial.
- n is the number of trials.
- k is the number of successes (ranging from 0 to n).

$$k=0:\ \frac{10!}{0!\cdot 10!}\left(\frac{1}{2}\right)^0\cdot\left(1-\frac{1}{2}\right)^{10}=0.0010$$

$$k=1:\ \frac{10!}{1!\cdot 9!}\left(\frac{1}{2}\right)^1\cdot\left(1-\frac{1}{2}\right)^{9}=0.0098$$

$$k=2:\ \frac{10!}{2!\cdot 8!}\left(\frac{1}{2}\right)^2\cdot\left(1-\frac{1}{2}\right)^{8}=0.0439$$

$$\vdots$$

$$k=10:\ \frac{10!}{10!\cdot 0!}\left(\frac{1}{2}\right)^{10}\cdot\left(1-\frac{1}{2}\right)^{0}=0.0010$$

Probability of Continuous Events

◆ Probability of Continuous Events

- **The Probabilities of Certain Events are All 0**

  : The probability of selecting specific number is 0 since the number of possible numbers that can be selected between consecutive integers is infinite (there are infinitely many numbers between 1 and 6).

- (Ex) Let's assume we randomly select any number between 1 and 6. There are countless real natural numbers between 1 and 6.
  - Q: What is the probability of picking exactly 5?
  - A: Since there are infinite numbers between 1 and 6, the probability of picking exactly 5 is $1/\infty=0$.

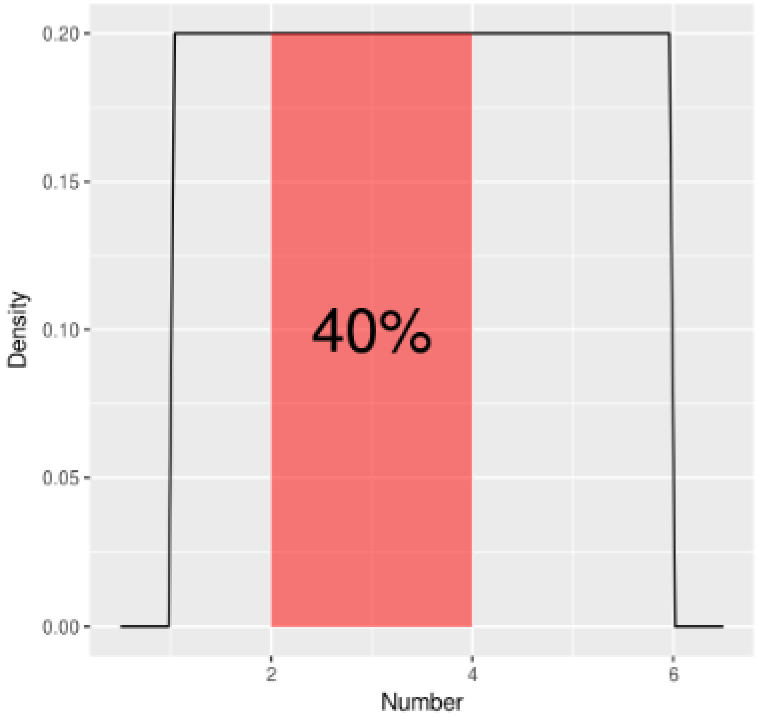
- Probability = 0 is meaningless to analyze.
- we need to consider another approach : **Probability of belonging to a specific interval**

Probability Density Function

- **Probability of belonging to a specific interval**
  - The probability of picking exactly 5 : 0%
  - The probability of picking a number between 4 and 5 : 20%
  - The probability of picking a number between 1 and 3 : 40%

- **Probability Density Function (PDF)**
  - We can utilize the probability of belonging to a certain interval, rather than the probability of exact point.
  - Probability Density Function (PDF) is a function used to calculate the probability of belonging to a certain interval

Probability Density Function
(Roll a dice)



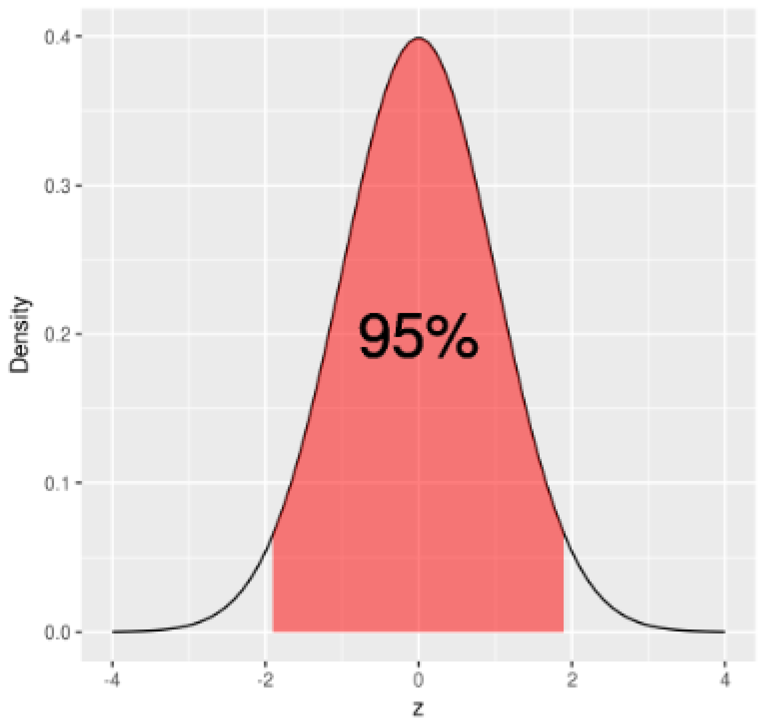
Probability Density Function

◆ **PDF of standard normal distribution**

➢ In the standard normal distribution with mean 0 and variance 1, PDF is expressed as $\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

➢ The probability that z is within -1.96 and 1.96 is 95%

Standard Normal Distribution



Likelihood

◆ **PDF of standard normal distribution**

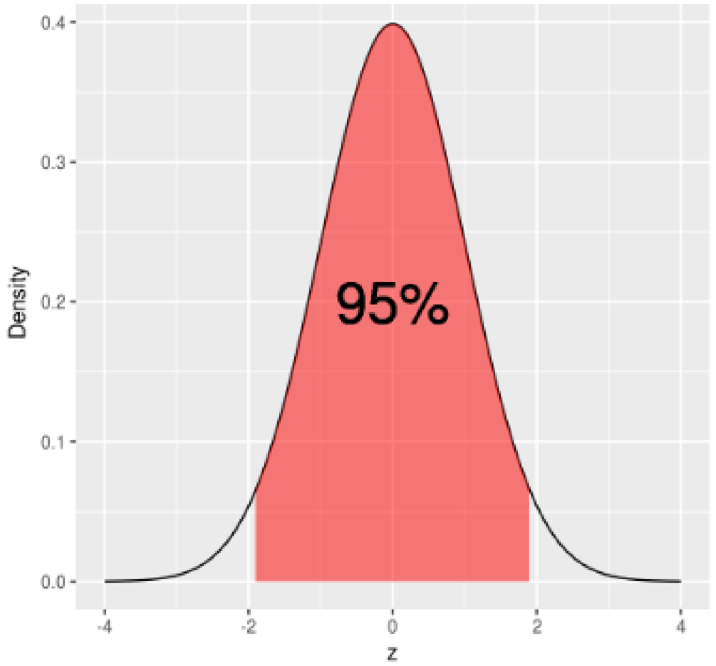
- Now we know that the probability of picking the exact number is 0%.
- Then, is it impossible to express the probability of a certain event?

◆ According to the PDF of the normal distribution,

- we can say that the Z value is most likely to be near 0, followed by near 1.
- It is highly unlikely that the Z value is 999.

=> likely… unlikely…. => Likelihood

Standard Normal Distribution



Likelihood

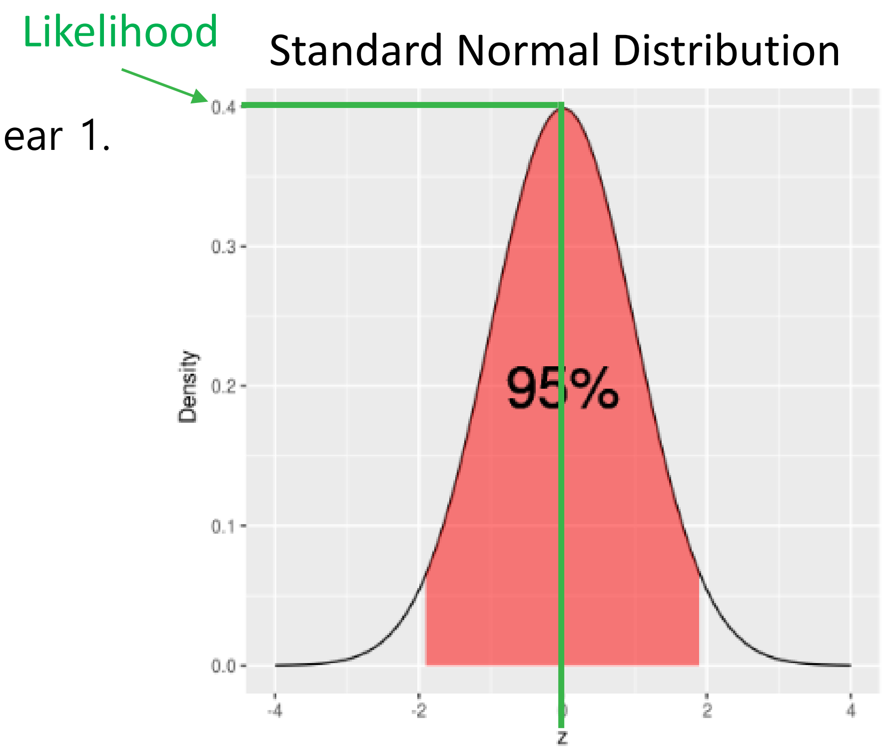
◆ **PDF of standard normal distribution**

- Now we know that the probability of picking the exact number is 0%.
- Then, is it impossible to express the probability of a certain event?

◆ According to the PDF of the normal distribution,

- we can say that the Z value is most likely to be near 0, followed by near 1.
- It is highly unlikely that the Z value is 999.

=> likely... unlikely.... => Likelihood



Likelihood

◆ Likelihood

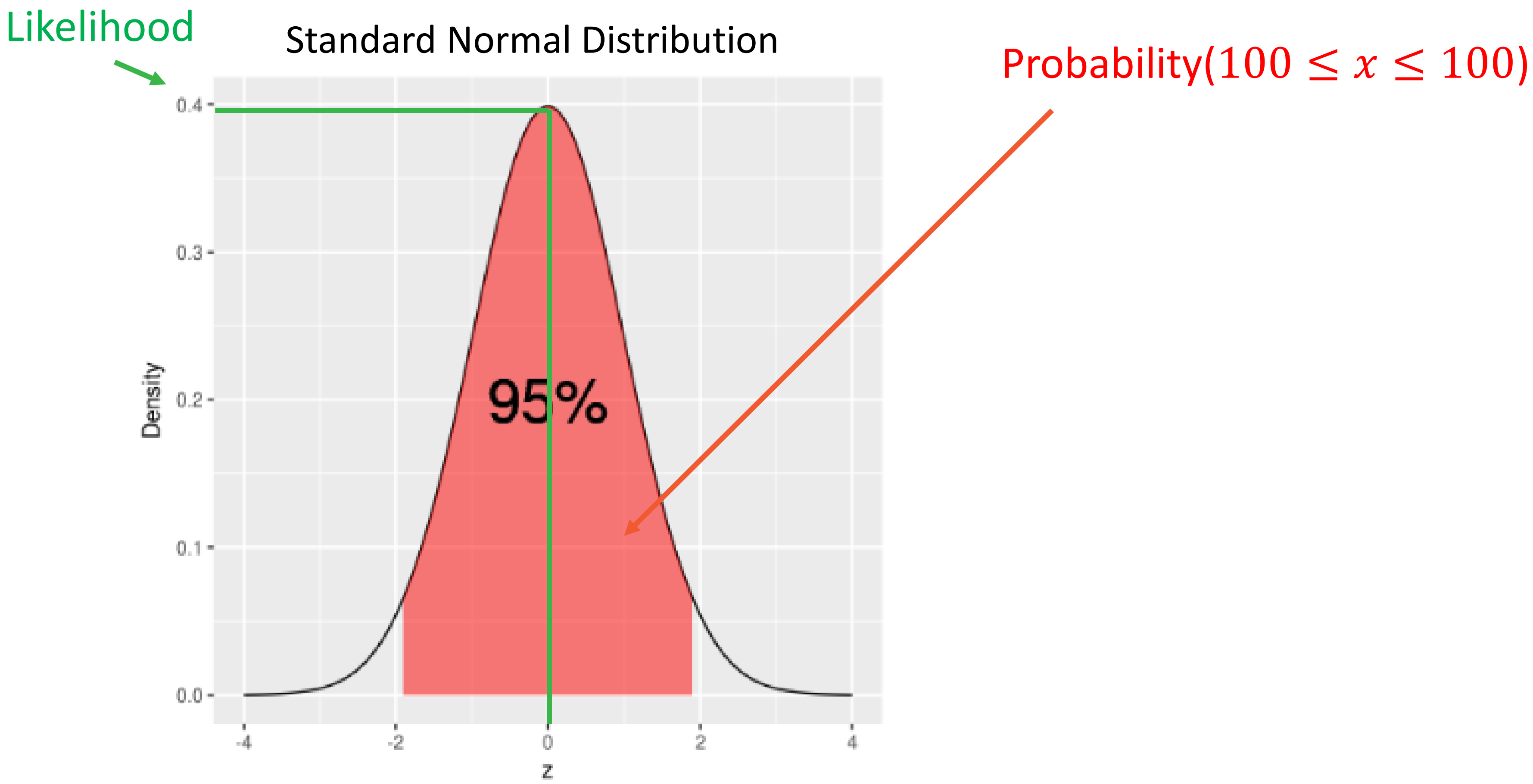
- It tells us how "likely" our observed data is given certain parameters of the model
- The higher the “likelihood” value, the more likely the event will occur
  - ✓ We cannot say that the probability will be high.
- A measure of how well a statistical model or set of parameters explains observed data
- Unlike probability, it does not sum or integrate to 1

◆ Probability = Likelihood?

- **For discrete events** (like rolling dice), the terms "probability" and "likelihood" can often be used interchangeably. Probability directly quantifies the chance of each specific outcome.

- **For continuous events** (like measurements modeled by normal distributions), “likelihood“ is not the probability of specific points. “probability” pertains to intervals, not specific points, because the probability of any exact value is 0. Instead, we use a probability density function (pdf) to describe the density of probability across an interval. "Likelihood," in this context, refers to the value of the pdf at a specific point, which is crucial for statistical inference but does not represent probability in the traditional sense.

Probability vs. Likelihood



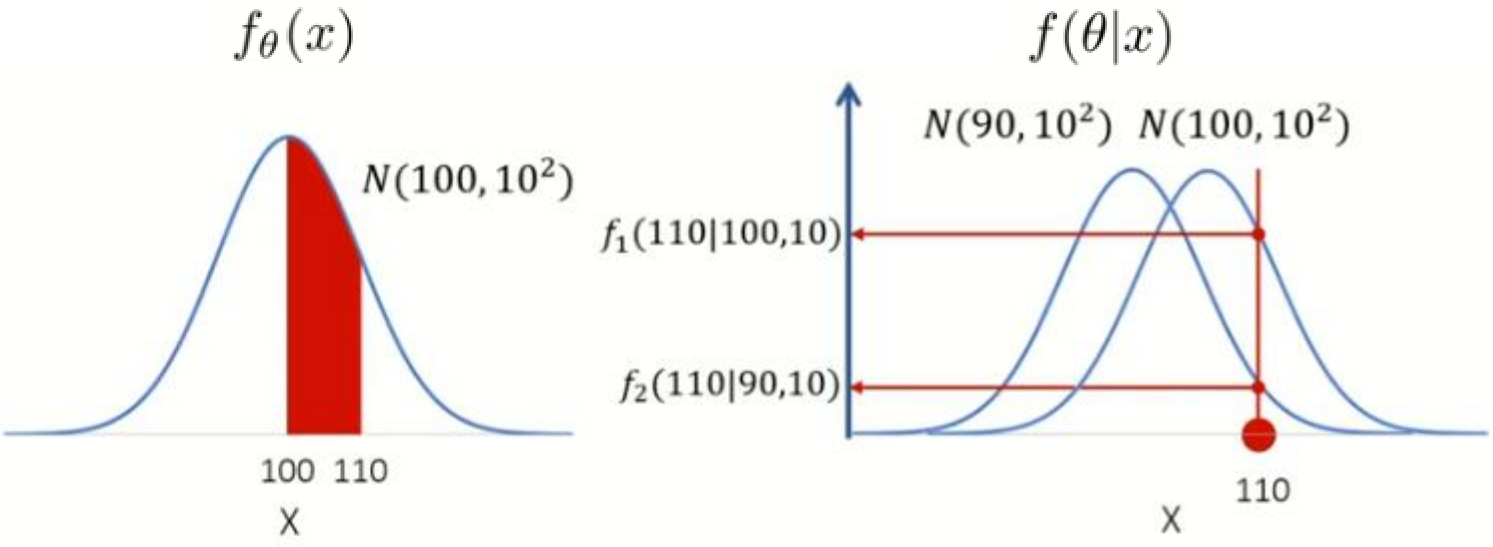
Probability vs. Likelihood

◆ Probability

- Given the parameters of the distribution, the probability that sample data appears

◆ Likelihood?

- Given specific sample data, (we don't know the parameter) finding a function by changing parameters.
- The goal is to estimate the underlying population parameters by optimizing the model parameters without prior knowledge of these parameters



Probability vs. Likelihood

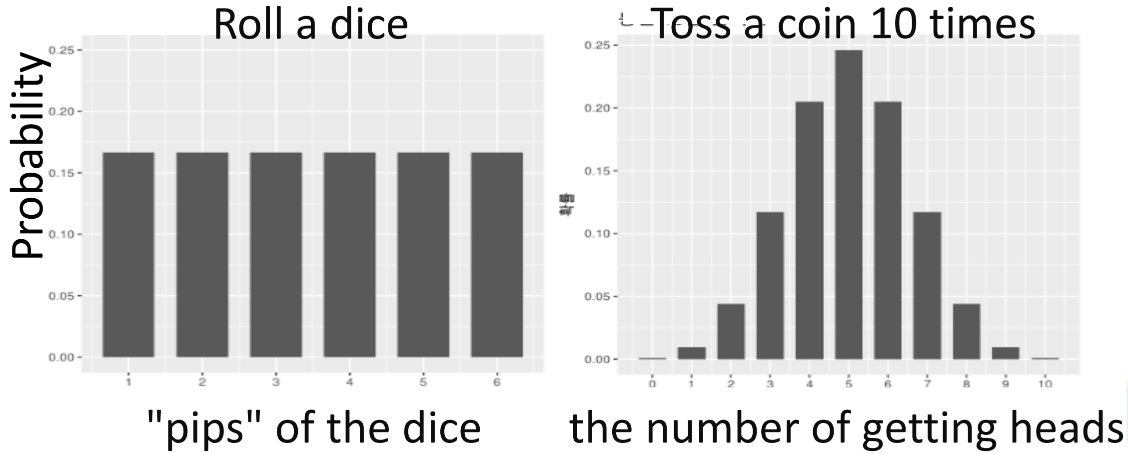
◆ Probability and Likelihood when events occur multiple times

1. What is the probability of throwing a dice 3 times and getting 1, 3, and 6 respectively?
   - $\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$

2. What is the probability of 'tossing a coin 10 times' 3 times, and getting heads 2, 5, and 7 times?

$$Pr(K = k) = [0.0010, 0.0098, \boxed{0.0439}, 0.1172, 0.2051, \boxed{0.2461}, 0.2051, \boxed{0.1172}, 0.0439, 0.0098, 0.0010]$$

   - 0.0439 × 0.2461 × 0.1172 = 0.0013



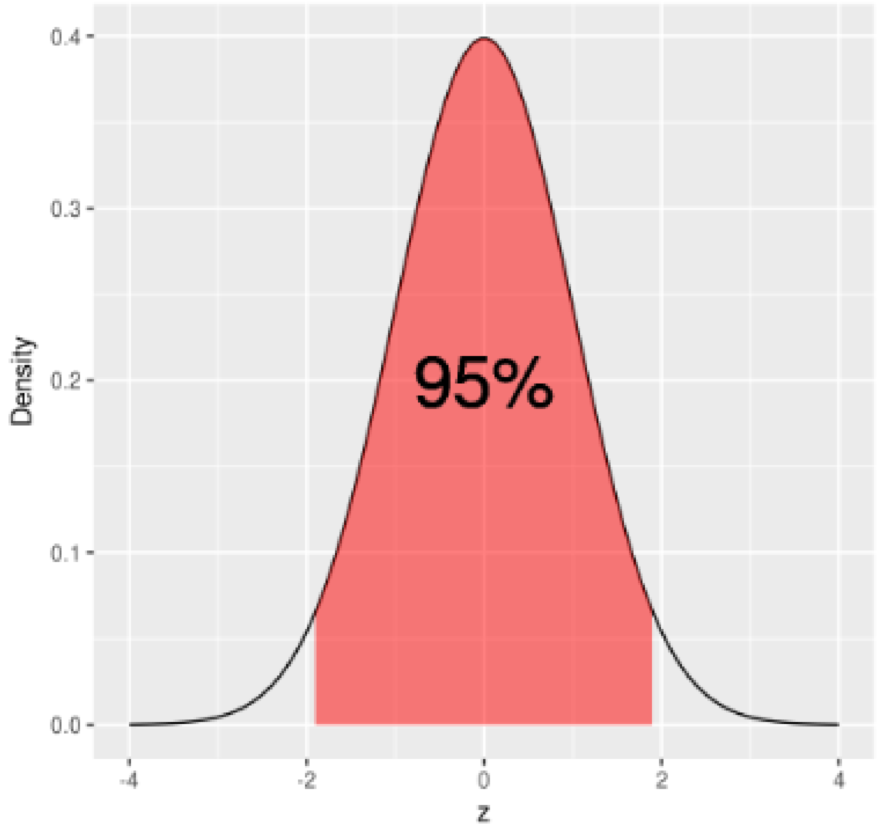
Probability vs. Likelihood

◆ Probability vs. Likelihood of continuous events occurring multiple times (Standard Normal Distribution)

1. What is the probability of getting -1, 0, 1 in Standard Normal Distribution?
   - ➢ 0%, 0%, 0% = 0%

2. What is the likelihood of getting -1, 0, 1 in Standard Normal Distribution?
   (Please refer to the normal distribution PDF)
   - ➢ 0.24 × 0.40 × 0.24 = 0.0230

PDF of the
Standard Normal Distribution



Maximum Likelihood Estimator, (MLE)

◆ Maximum Likelihood Estimator

- A method used for estimating the parameters of a probability distribution
- The goal of MLE is to find the best parameter values for your model that make the observed data most likely to occur

◆ How MLE Works

① **Model Assumption:** assuming a specific form for the probability distribution (e.g., Normal distribution, Binomial distribution) of the data, which is characterized by certain parameters (e.g., mean and variance for Normal distribution).

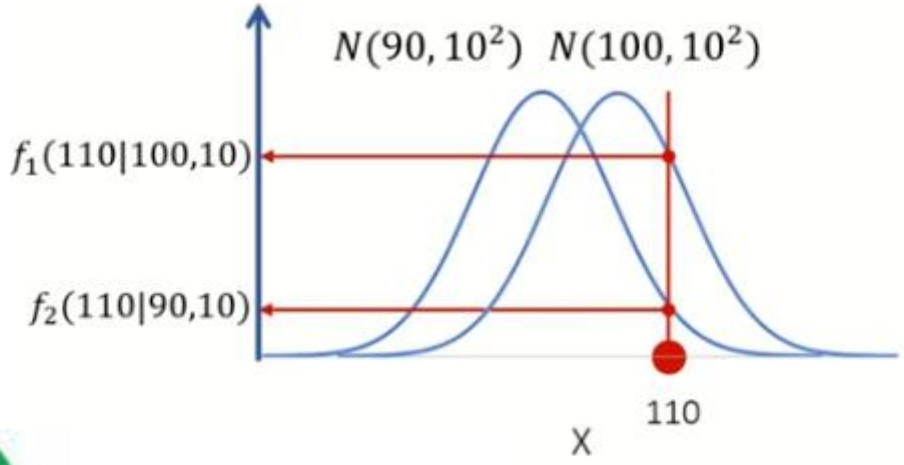
② **Likelihood Function:** constructing a likelihood function, $L(\theta|x)$, which is the probability of observing your data given the parameters θ. This function is essentially the product of the probability densities (or mass functions, for discrete variables) of all observed data points, considering the assumed distribution.

③ **Maximization:** finding the parameter values that maximize the likelihood function. This process often involves taking the logarithm of the likelihood function to convert the product into a sum (making it easier to work with), resulting in what's called the log-likelihood. Then, you differentiate this log-likelihood with respect to the parameters and solve for the values that set the derivatives to zero, adjusting for any constraints (like parameters that must be positive).

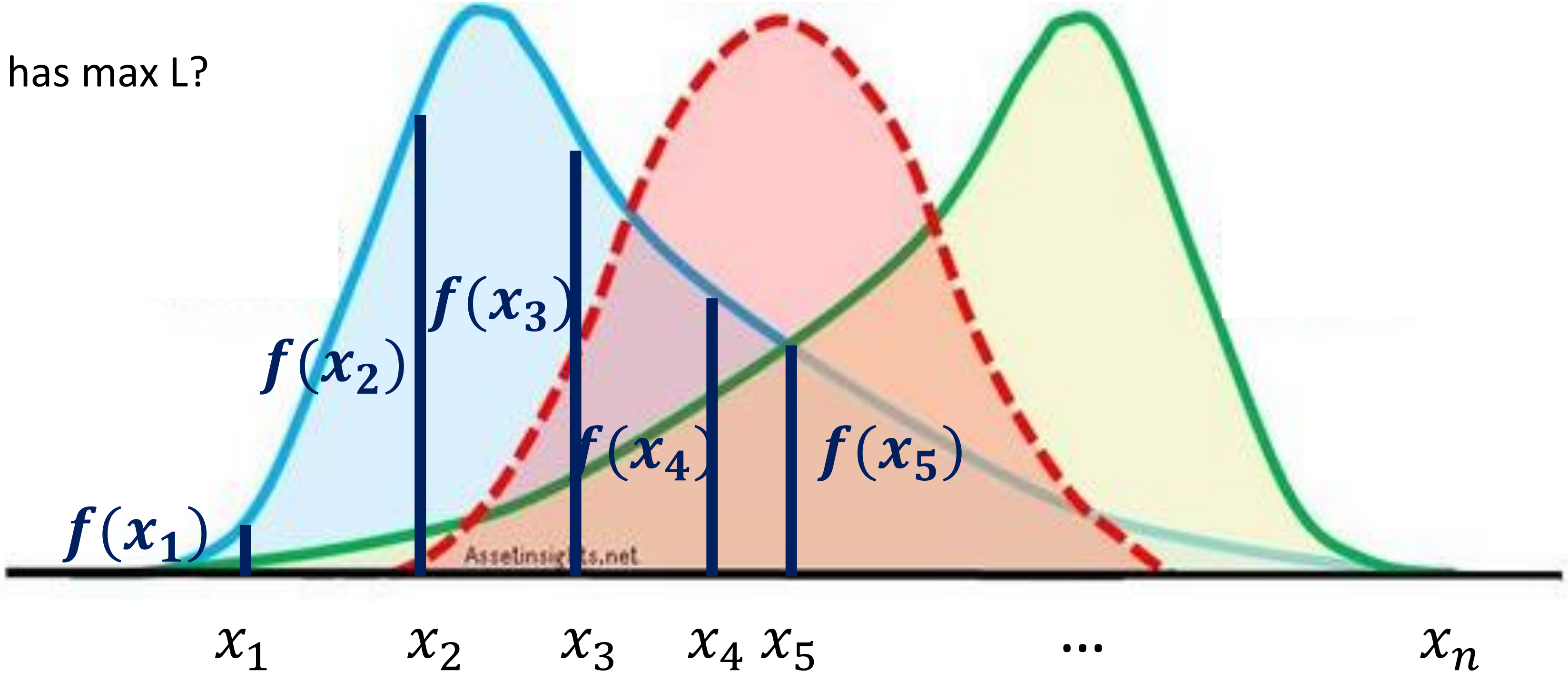
Maximum Likelihood Estimator, (MLE)

◆ Maximum Likelihood Estimator

$$L(\theta|\boldsymbol{x}) = f(x_1) \times f(x_2) \times \cdots \times f(x_n)$$
$$= \prod_{i=1}^{n} f(x_i;\theta)$$



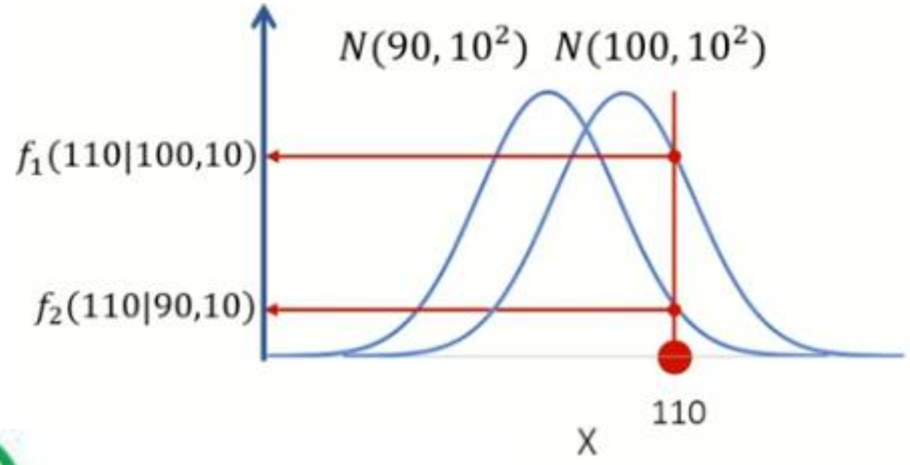
Which distribution has max L?



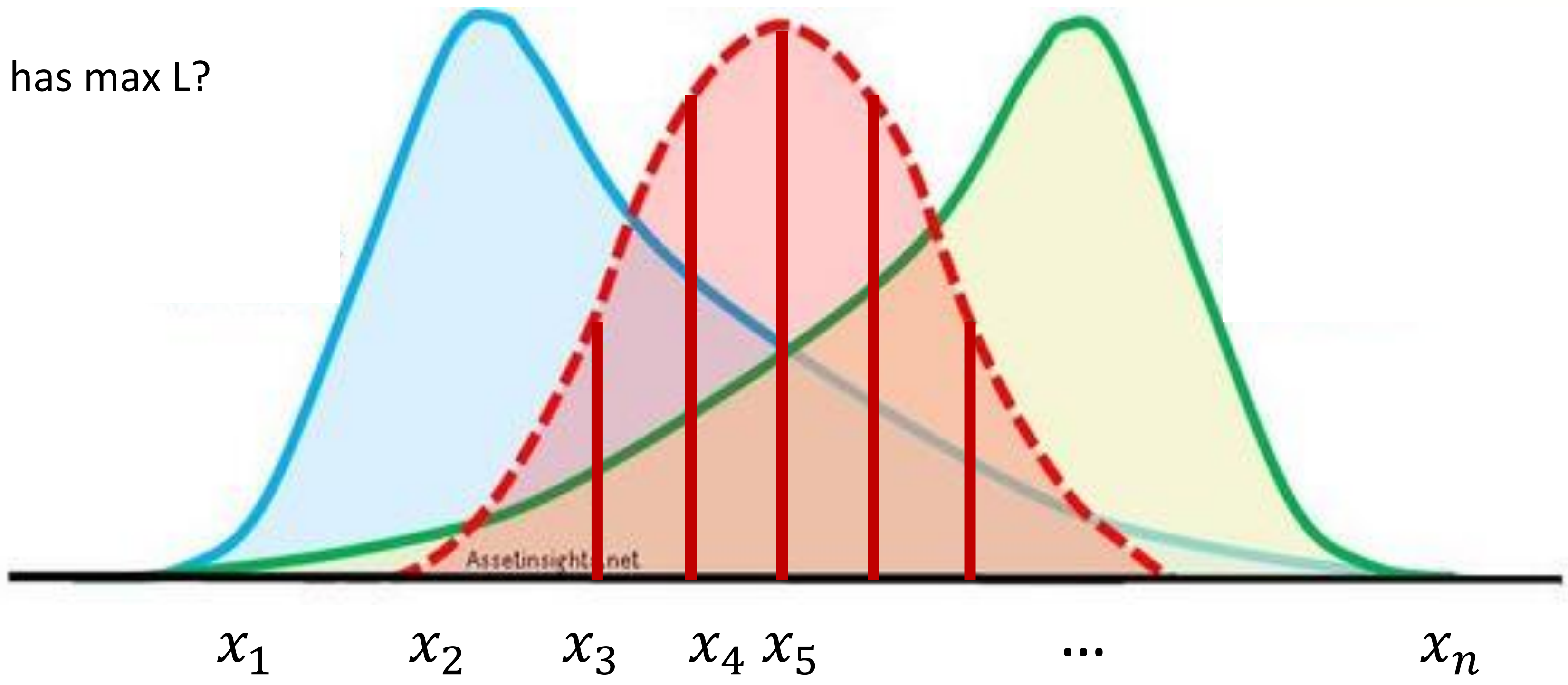
Maximum Likelihood Estimator, (MLE)

◆ **Maximum Likelihood Estimator**

$$L(\theta|\boldsymbol{x}) = f(x_1) \times f(x_2) \times \cdots \times f(x_n)$$
$$= \prod_{i=1}^{n} f(x_i;\theta)$$



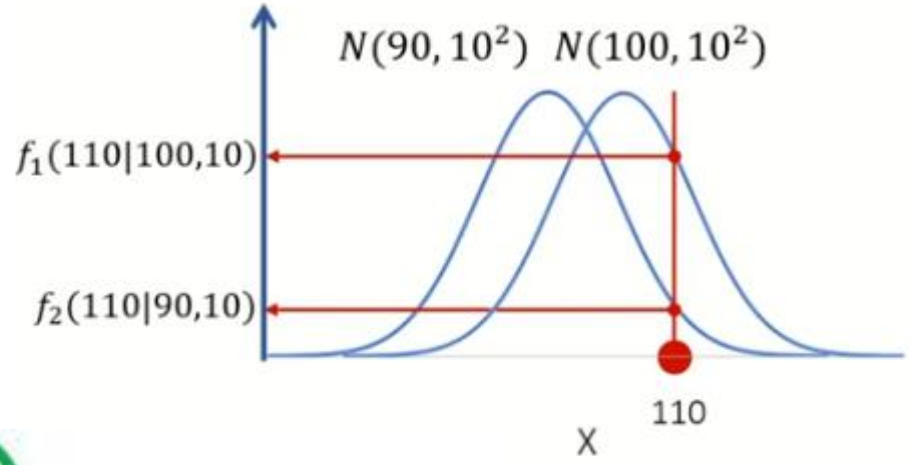
Which distribution has max L?



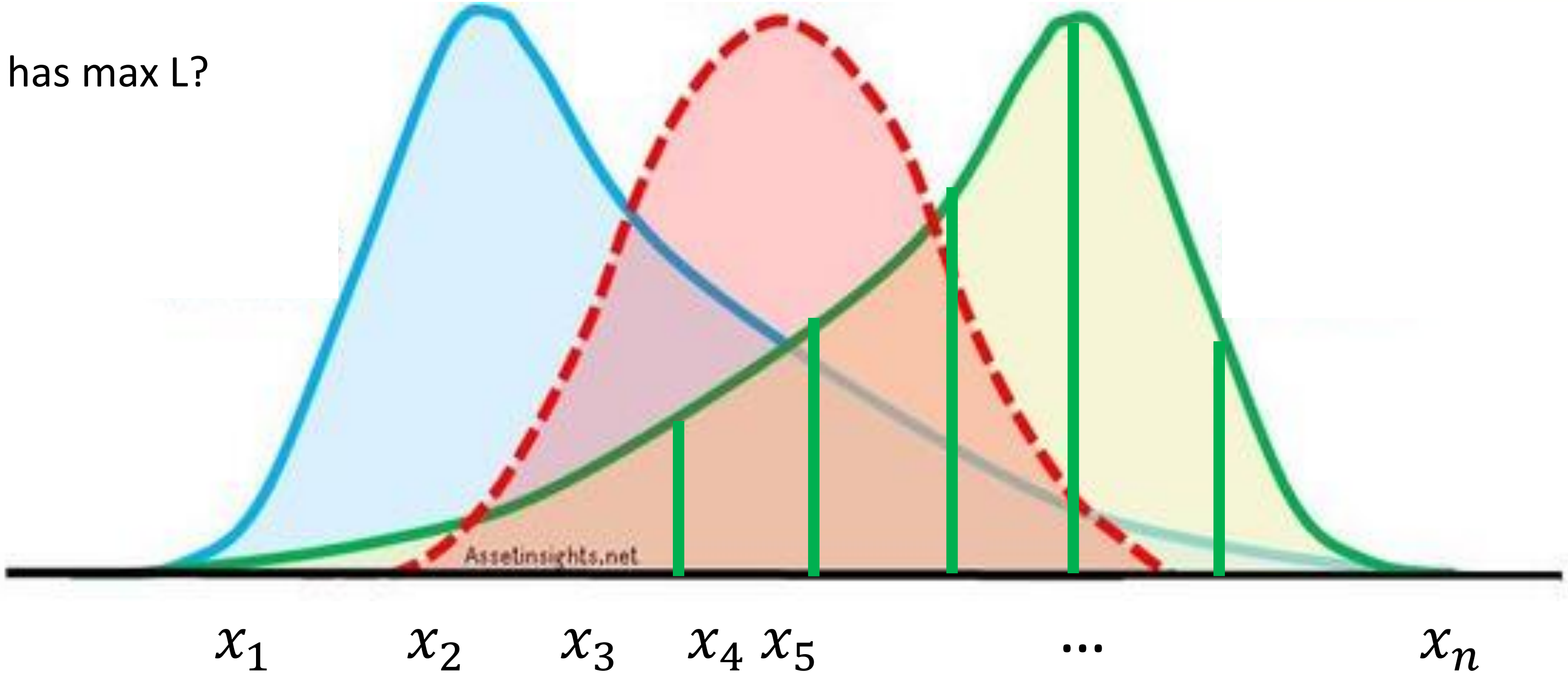
Maximum Likelihood Estimator, (MLE)

◆ **Maximum Likelihood Estimator**

$$L(\theta|\boldsymbol{x}) = f(x_1) \times f(x_2) \times \cdots \times f(x_n)$$
$$= \prod_{i=1}^{n} f(x_i;\theta)$$



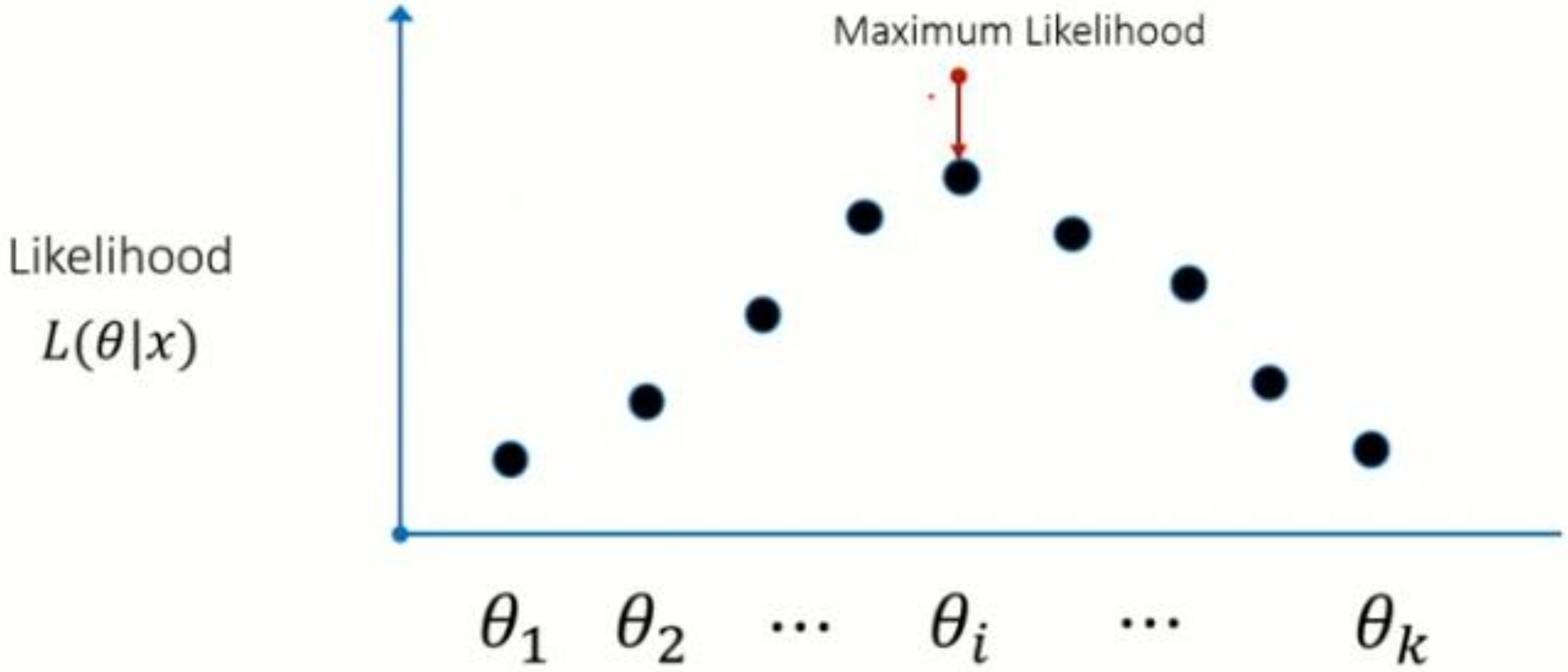
Which distribution has max L?



Maximum Likelihood Estimator, (MLE)

◆ **Maximum Likelihood Estimator**

➢ **Finding theta (=parameters of the distribution) that makes the Likelihood be max!! => MLE**



Bye Bye