Introduction to Statistics and Data Science using *eStat*

Chapter 5 Probability Distribution

5.4 Continuous Random Variable

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

5.4 Continuous Random Variable

- Consider a statistical experiment that measures how long it takes for an office worker to get to work from home.
  => Past experience shows that the commuting time usually takes about 30 minutes to get to the work in usual traffic

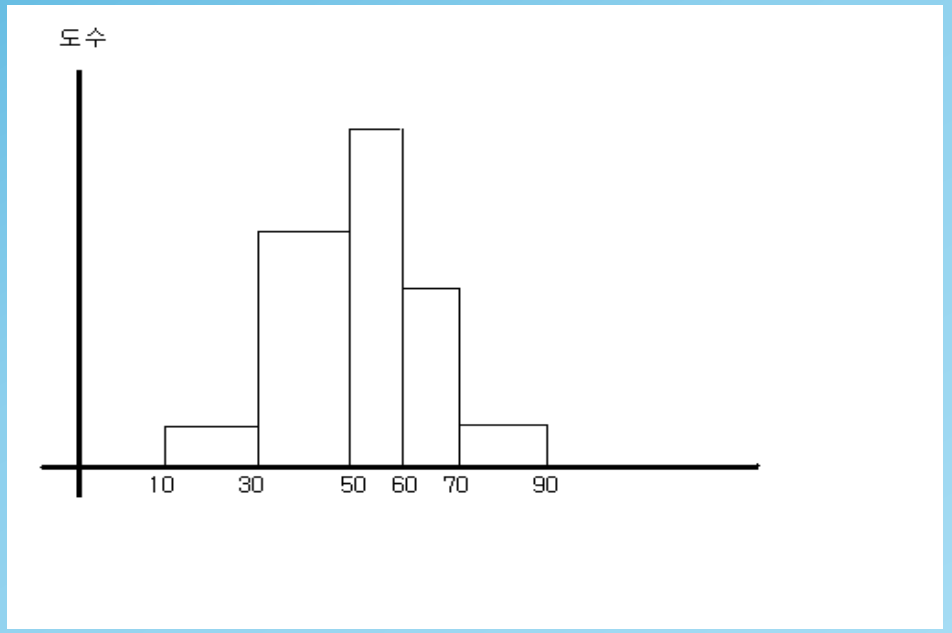
- Define a random variable X as the 'time to work place'.
  => infinite number of possible values for a random variable
  => this is called a continuous random variable.

- Probability calculation at each point is meaningless which is zero.

- Probability of an interval is of interest,
  => 'What is probability of a commuting time between 25 and 35 minutes?'

5.4 Continuous Random Variable

Interval ($a \le X < b$)	Frequency	Probability
$10 \le X < 30$	5	5/100
$30 \le X < 50$	30	30/100
$50 \le X < 60$	40	40/100
$60 \le X < 70$	20	20/100
$70 \le X < 90$	5	5/100
Total	100	1

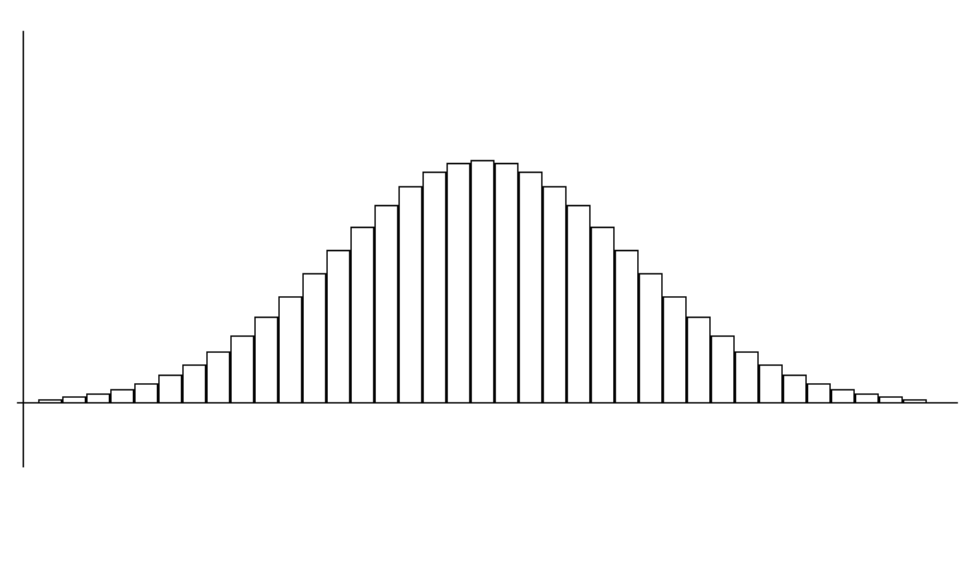


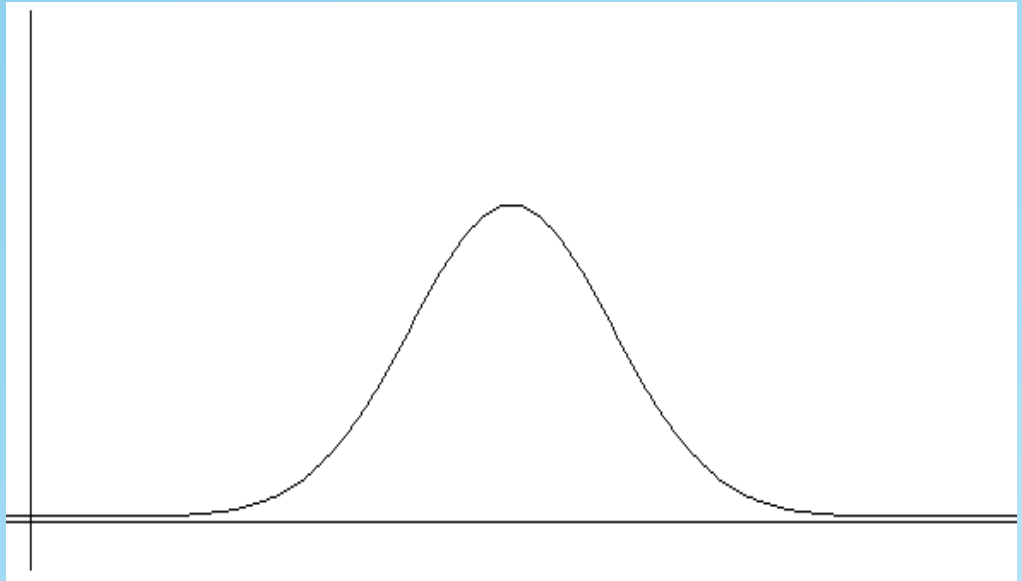
- Using this frequency table, 'probability of commuting time between 30 and 60 minutes' is as follows.

$$P(30 \le X < 60) = 30/100 + 40/100 = 70/100$$

5.4 Continuous Random Variable

- If you use this table, you cannot calculate probability of the commuting time between 25 and 35 minutes.
  ⇨ this probability will require a detail frequency table and histogram which is narrower in the interval by obtaining more data.
- If you increase the number of data and close to zero width of the interval, this histogram will be approximated to a continuous function.
  ⇨ probability distribution function of a continuous random variable.





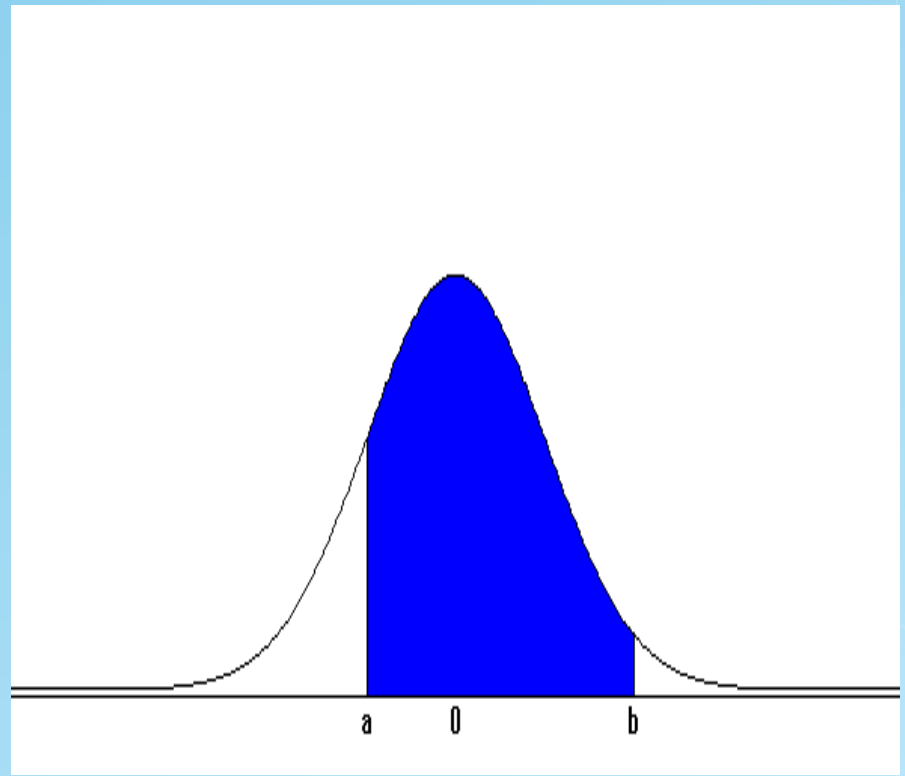
5.4 Continuous Random Variable

- If the probability distribution function of continuous random variable can be expressed as a function f(x), the desired probability can be obtained without finding the frequency table and histogram.
- This area under this function f(x) should be 1 because the addition of all probabilities is 1.

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

- The probability of the random variable X at interval (a, b), P(a < X < b), can be obtained as the area between (a, b) of f(x) which is the integral.

$$P(a < X < b) = \int_{a}^{b} f(x)dx$$



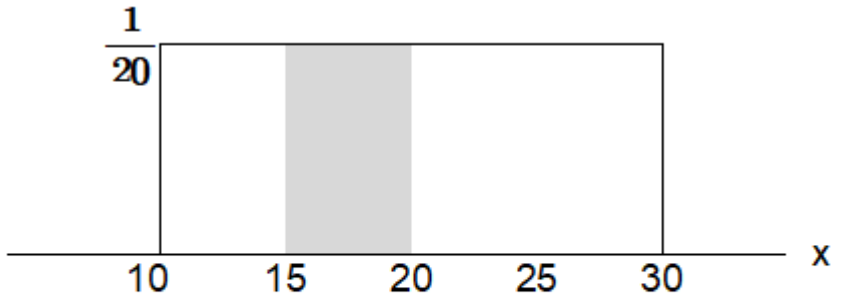
5.4 Continuous Random Variable

[Ex 5.4.1] The time takes to order a pizza and get home has the same possibility as any time from 10 to 30 minutes. Let X be the time it takes to deliver a pizza. Find the probability distribution function of X and draw a picture. Find the probability of delivery between 15 and 20 minutes.

<Answer>

- **Since the random variable X has the same possibility as any number between 10 and 30, the pdf is called a uniform distribution between 10 and 30 denoted as Uniform(10,30).**
- **The probability of delivery in 15 to 20 is the area of the shaded rectangle.**
  P(15 < X < 20) = (20 - 15) × (1 / 20) = 0.25

$$f(x)=\begin{cases}1/(30-10), & 10<x<30\\ 0, & \text{기타}\end{cases}$$



<Figure 5.4.5> Uniform distribution on (10,30) and the probability of P(15 < X < 20)

Thank you