Chapter 5

Probability Distribution

Professor Jung Jin Lee

Table of Contents

Chapter 5 Probability Distribution

5.1 Definition of Probability

- Similar events occur frequently or carried out around our lives.

  - A machine is producing products repeatedly at a production plant.
    => product is either a normal or a defective product,
    but not known what will be.

  - We order pizza at home every Sunday.
    => It usually takes about 30 minutes for a pizza to be delivered to the house,
    but the exact time is not known.

- What these examples have in common is as follows.
  ① The repetition of similar events.
  ② Various possible outcomes are known.
  ③ There is no telling what exactly will happen.

5.1 Definition of Probability

- Events with these three characteristics are subject to the study and application of statistics.
- An experiment in which a particular case occurs among several possible cases is called a statistical experiment.

- Examples of statistical experiment?
- Examples of non-statistical experiment?

5.1 Definition of Probability

- All possible sets of outcomes from a statistical experiment is a sample space.
  - Sample space is usually marked by S
  - S = {normal, defective}
- A subset of this sample space is referred to as event.
  - Events are denoted in English capitals A, B, and C ...
  - A = {defective}
- When the number of elements in a sample space is finite or can be counted indefinitely, this is called a discrete sample space.
- When the number of elements in a sample space is infinite and uncountably innumerable, this is called a continuous sample space.

5.1 Definition of Probability

- The concept of probability is used to indicate the possibility of an event occurring in a statistical experiment.
- Probability is the 'representation of the likelihood of an event occurring' between 0 and 1.
- If an event is likely to occur, the probability is expressed as a number close to 1. if it is unlikely to occur the probability is expressed as a number close to zero.
- Two definitions of probability,
  - classical definition
  - relative frequency definition

5.1 Definition of Probability

❖ Classical definition of probability

- Assume that <u>all elements in the sample space are likely to occur equally</u>.

- The probability of an A will occur in case of discrete sample space is

$$P(A) = \frac{\text{Number of elements belonging to event A}}{\text{Total number of elements in sample space}}$$

- The probability of an A will occur in case of continuous sample space is

$$P(A) = \frac{\text{Measurement of elements belonging to event A}}{\text{Measurement of elements in sample space}}$$

* Measurement can be length, area, volume etc.

5.1 Definition of Probability

[Ex 5.1.1] An office worker went on a business trip to a city and there are two restaurants (e.g. restaurant A, restaurant B) near his lodging.
- He was hesitating about which restaurant to go and threw a dice to count the number of points that appear on the top.
- If he had odd numbers, he would go to the restaurant A, and if he had even numbers, he would go to the restaurant B. What is the probability that the restaurant A would be picked?

<Answer>

- **The sample space in this statistical experiment, which counts the number of points on the top by throwing a dice, is {1, 2, 3, 4, 5, 6}.**
- **The number of odd events is {1, 3, 5}, so there are three elements.**
- **Therefore, the probability that restaurant A will be selected is 3/6 = 1/2.**

5.1 Definition of Probability

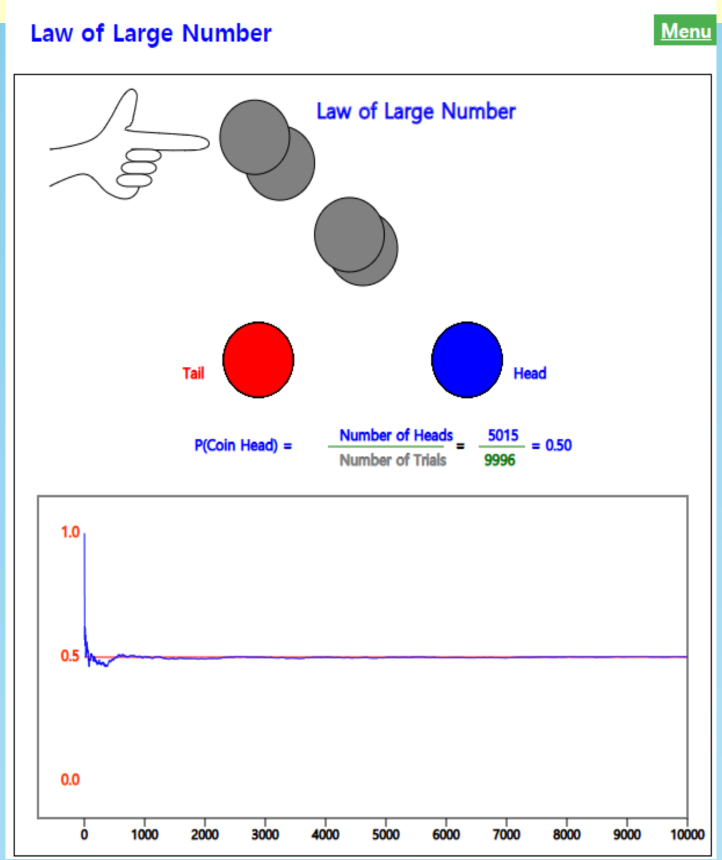
[Ex 5.1.2] Order pizza from home every Sunday. The time it takes for a pizza to be delivered to the house has the same possibility for any time from 10 to 30 minutes (you may have a decimal number).
- What is the probability that a pizza delivery will be delivered between 20 and 25 minutes?

<Answer>
- The sample space in this example is all values from 10 to 30 minutes {(10,30)}.
- The events where pizza is delivered between 20 and 25 minutes is { (20,25) }.
- Therefore, the probability of this event is (25-20) / (30-10) = 0.25 by measuring the distance of the interval.

5.1 Definition of Probability

❖ **Relative frequency definition of probability**

- The probability that event A will occur is the rate at which event A occurs when the statistical experiments are conducted under the same conditions repeatedly.



❖ Law of Large Number

- If a coin is thrown many times, the probability of {Head} event converges to half

5.2 Calculation of Probability

❖ Permutation

- The number of ways to select r objects out of n objects considering the order is called permutation and is calculated as follows.

$$ {}_{n}P_{r} = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} $$

- Therefore, the number of ways to list all n objects is as follows.

  0 ! = 1

$$ {}_{n}P_{n} = n(n-1)(n-2)\cdots 2\cdot 1 = n! $$

5.2 Calculation of Probability

❖ Combination

- The number of ways to select r objects out of n objects without considering the order is called combination and is calculated as follows.

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!} $$

5.2 Calculation of Probability

[Example 5.2.1] Four people A, B, C and D are intended to be placed on four side-by-side chairs. Obtain the total number of cases in which four people are placed, and the number in which A is placed on the leftmost. What is the probability that A is placed on the leftmost side?

<Answer>

- The number of elements in the sample space in this issue is as follows:
  (Number of people that can be placed on the leftmost)
  × (number of people except left who can be placed in the second position)
  × (number of people who can be placed in third place except for two left people)
  × (number of people, excluding the three on the left)
  $= 4 \times 3 \times 2 \times 1 = 4! = 24$
- The event in which A is placed on the left is the number of people placed in the second, third, and right positions except A, so $3 \times 2 \times 1 = 3!$.
- Therefore, the probability that A will be placed to the left is $3! / 4! = 6 / 24 = 0.25$.

5.2 Calculation of Probability

[Ex 5.2.2] A company has four security guards (A, B, C, D). Each morning, two of these guards are randomly selected, one at the front gate and the other at the rear guard. Obtain the total number of cases in which four people are placed at the front and rear gates and the number in which A is placed at the front gate. What is the probability that A will be placed at the front gate?

<Answer>

- The number of elements in the sample space in this issue is as follows:
  (number of people who can be placed at the front gate)
  × (number of people who can be placed in the rear except those in the front)
  = 4 × 3 = ${}_4P_2 = 12$.
- The number of elements where A will be placed at the front gate is ${}_3P_1 = 3$
- Since A can be placed at the front gate and one of the other three can be placed at the rear gate. That is, the probability that A will be placed at the front gate one day is

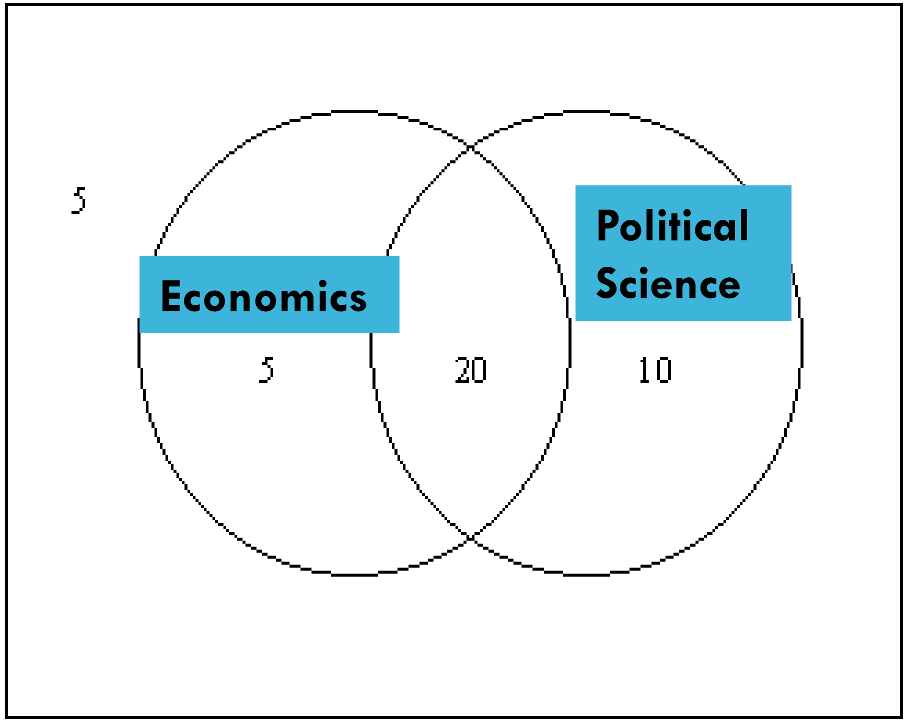
$$\frac{{}_3P_1}{{}_4P_2} = \frac{3\times1}{4\times3} = \frac{1}{4}$$

5.2 Calculation of Probability

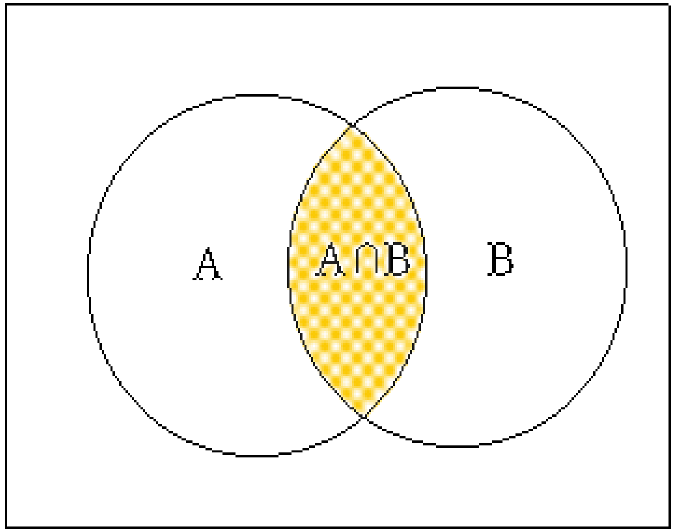
[Example 5.2.3] Out of 40 sophomores in statistics department this semester, 25 students are taking economics(A), 30 students are taking political science(B) and 20 students are taking both subjects. When I meet one of the sophomores, what is the probability of this student taking either economics or political science?

<Answer>

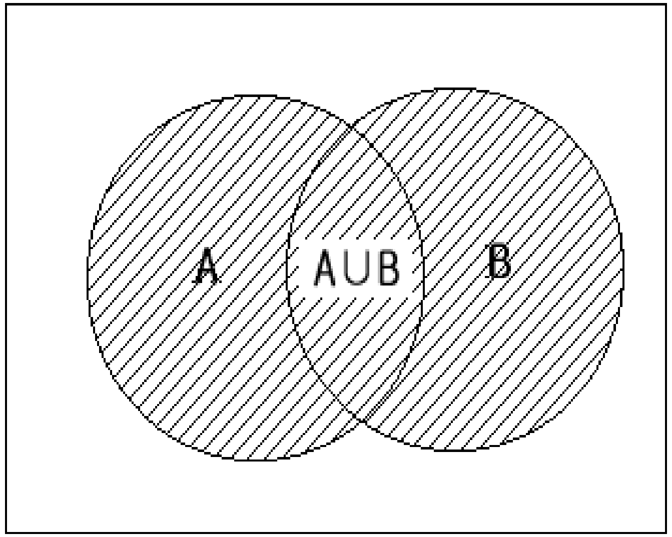
- Since there are 25 students who take economics and 20 students taking both courses, 25 - 20 = 5 students take only economics.
- Since there are 30 students who take political science, 30 - 20 = 10 students take only political science.
- The number of students taking economics or politics is 5 + 10 + 20 = 35.
- The probability of students taking economics or politics is 35 / 40.



5.2 Calculation of Probability



- Let's call the case of students taking economics A and the case of students taking political science B. Events that take both courses are marked as $A \cap B$ and are called an intersection event of A and B.
- The event in which a student takes a course in either economics or political science (one or both) is marked as $A \cup B$ and is called an union event of A and B



- The probability of $P(A \cup B)$ can also be calculated as

$$P(A \cup B) = P(A) + P(B) - P(A \cup B)$$
$$= 25/40 + 30/40 - 20/40 = 35/40$$

- The probability of taking either economics or political science, $P(A \cup B)$, can be calculated by adding the probability of taking each course and then by subtracting the probability of taking both courses.

5.2 Calculation of Probability

❖ Addition Rule of Probability

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

If $A \cap B = \emptyset$, then the rule becomes as follows.

$$P(A \cup B) = P(A) + P(B)$$

In this case, the events A and B are called **mutually exclusive events.**

5.2 Calculation of Probability

[Example 5.2.5] Of the 30 sophomores in the Department of Statistics, there are 10 males and 20 females, one of males is from the province and five of females are from the province.

1) When selecting a student, what is the probability that he is from a province?
2) When I selected a student, she was a female. What is the probability that this student is from a province?
3) When I selected a student, he was from a province. What's the probability of this student being a male?
4) When selecting a student, what is the probability that he is male and from Baku?

<Answer>

- To solve this problem, it is convenient to organize the information given into a partition table.

	Baku	Province	Total
Male	____	1	10
Female	____	5	20
Total	____	____	30

5.2 Calculation of Probability

- Calculate and insert the blanks along with the following.

	Baku(S)	Province(C)	Total
Male(M)	9	1	10
Female(F)	15	5	20
Total	24	6	30

1) P(C) = 6/30.
2) The probability that this student is from a province among girls is 5/20.
   This probability is expressed as P (C|F) and is called conditional probability.
3) The probability of a male from a province origin is P(M|C) = 1/6.

5.2 Calculation of Probability

4) The probability is $P(M \cap S)$ and the partition table shows that the answer is 9/30. Alternatively, the probability of being a male among all students can be first obtained (10/30) and then multiplied by the conditional probability of being from Baku ($P(S|M)) = 9/10$) among male. Namely

$$P(M \cap S) = P(M)\, P(S|M) = (10/30) \times (9/10) = 9/30$$

This expression shows that the conditional probability $P(S|M)$ can be calculated by dividing $P(M \cap S)$ by $P(M)$.

$$P(S|M) = \frac{P(M \cap S)}{P(M)} = \frac{9/30}{10/30} = \frac{9}{10}$$

$$P(M \cap S) = P(S)\, P(M|S) = (24/30) \times (9/24)$$

5.2 Calculation of Probability

❖ Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{if } P(B) \neq 0$$

❖ Multiplication Rule of Probability

$$P(A \cap B) = P(A)\,P(B|A)$$

If $P(B|A) = P(B)$, then $P(A \cap B) = P(A)\,P(B)$

** A and B are independent events

5.2 Calculation of Probability

[Ex 5.2.6] The Tiger of professional baseball team has the probability to beat the Lion team of 0.7 recently. What is the probabilities that the Tiger is winning both game on this evening double match? Assume that winning one game does not affect winning the next.

<Answer>

- Let us call the event that the Tiger wins the first game is A and the event that the Tiger wins the second game is B.
- Since A and B are independent of each other. the probability that the Tiger is winning both games is as follows,

$$P(A \cap B) = P(A)\,P(B) = 0.7 \times 0.7 = 0.49$$

5.2 Calculation of Probability

[Ex 5.2.7] The following is a table of 30 second-year students by gender and region of origin: Are the events of male and Baku origin independent of each other?

	Baku(S)	Province(C)	Total
Male(M)	5	5	10
Female(F)	10	10	20
Total	15	15	30

<Answer>

- Let us call the event of male as M, female as F, from Baku as S, and from other province as C.

  $P(M \cap S) = 5/30$, $P(M) = 10/30$, $P(S) = 15/30$ 이므로

  $P(M \cap S) = P(M)\, P(S)$
- Hence, gender and region are independent.
- Note that $P(M|S) = 5/15 = 1/3$, $P(M)=10/30$, therefore $P(M|S) = P(M)$.
- In this case, all items M and C, F and S, F and C are independent of each other and we call the two attributes, gender and region are independent.

5.2 Calculation of Probability

[Example 5.2.8] There is a box of six products, two of which are defective. What is the probability that at least one defective product will be found when three have been extracted for product testing? Assume that the product extracted once for inspection is a non-recovery extraction that is without replacement.

<Answer>

- Probability of finding one defect in three products is $({}_4C_2 \times {}_2C_1) / {}_6C_3 = 3/5$
- Probability of finding two other defective products is $({}_4C_1 \times {}_2C_2) / {}_6C_3 = 1/5$
- Thus, the probability that at least one defect will be found is 3/5 + 1/5 = 4/5.
- Another way to obtain this probability is to obtain the probability of an event in which there will be no defect (this is called a complementary event) and then subtract it from 1.

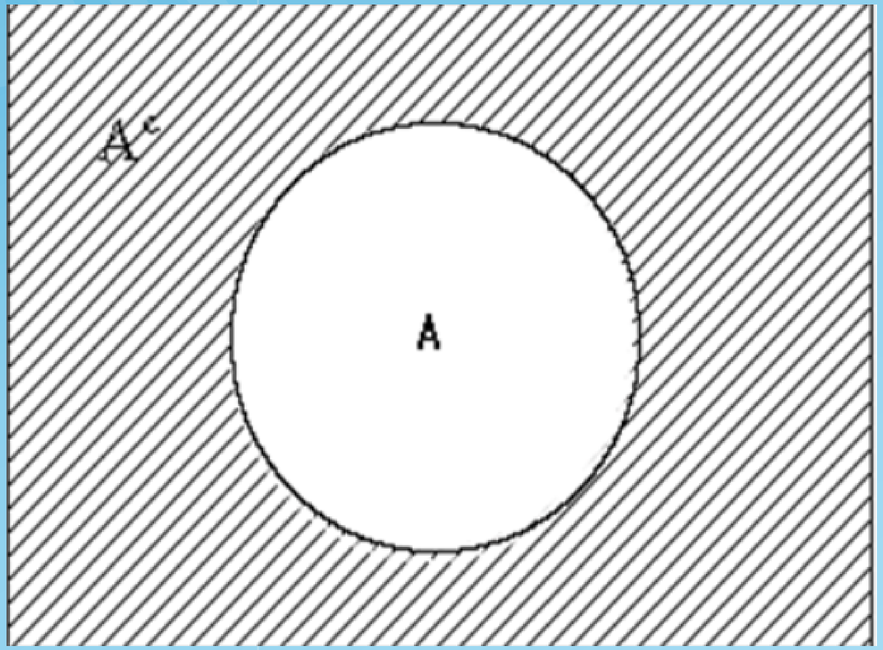
$$1 - ({}_4C_3 / {}_6C_3) = 1 - (4/20) = 4/5$$

5.2 Calculation of Probability

❖ **Probability of a complementary event**

If A^C denotes a complementary event of A, then P(A^C) can be calculated as follows.

$$P(A^C) = 1 - P(A)$$



5.3 Discrete Random Variable

- Consider a statistical experiment in which two coins are thrown repeatedly.
  If the coins are ideal, the sample space for this experiment is {'Tail-Tail', 'Tail-Head', 'Head-Tail' and 'Head-Head'}.
  The probability of each element of the sample space is 1/4.
- We are interested in counting the number of heads or tails. If X is defined as 'number of heads', the possible value of X can be 0, 1, or 2.
- A function that corresponds to one real number between [0,1] for each element of the sample space is called a random variable.

Sample Space	X=Number of {Head}
'Tail-Tail'	0
'Head-Tail'	1
'Tail-Head'	1
'Head-Head'	2

- If the possible values of random variable are finite or countably infinite, its is called a discrete random variable.
- If the possible values of random variable are uncountably infinite, it is called a continuous random variable.

5.3 Discrete Random Variable

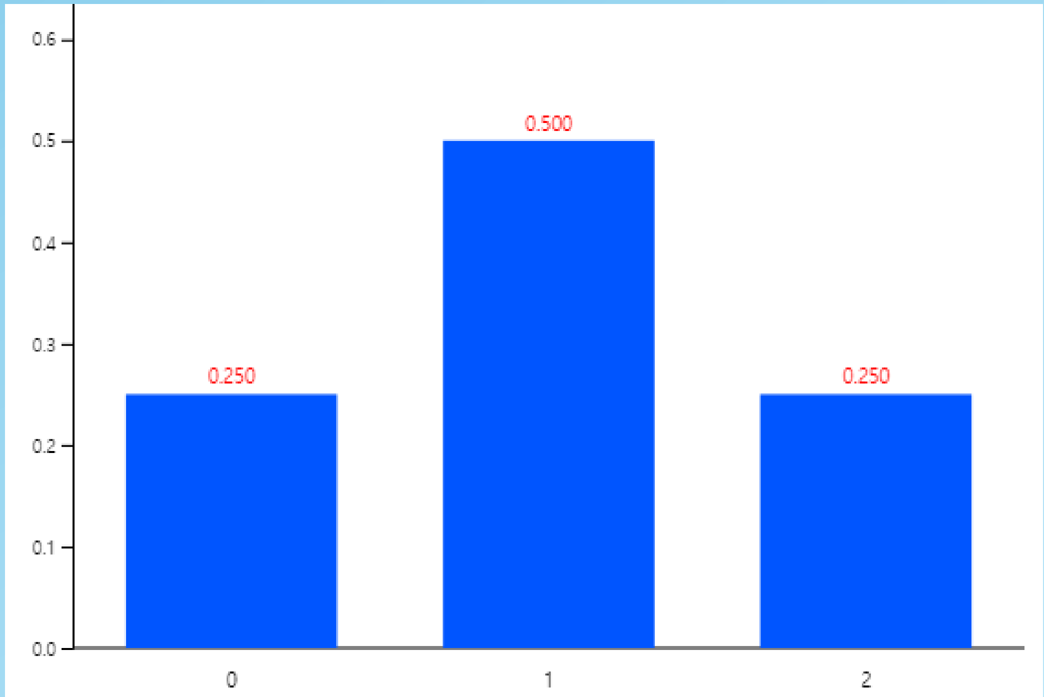
- The probability that the random variable X will be zero is 1/4 because it is the probability of an event {Tail-Tail}, the probability that X being 1 is 2/4 because P({Tail-Head, Head-Tail}) is 2/4, and the probability that X being 2 is 1/4 because P({Head-Head}) is 1/4.
- The summarized probability for the value of the random variable X is called the probability distribution function of X denoted as f(x).

1) Table style

X = x	P(X=x)
0	1/4
1	2/4
2	1/4
계	1

2) Function style

$$f(x) = \begin{cases} 1/4, & x=0 \\ 2/4, & x=1 \\ 1/4, & x=2 \end{cases}$$



5.3 Discrete Random Variable

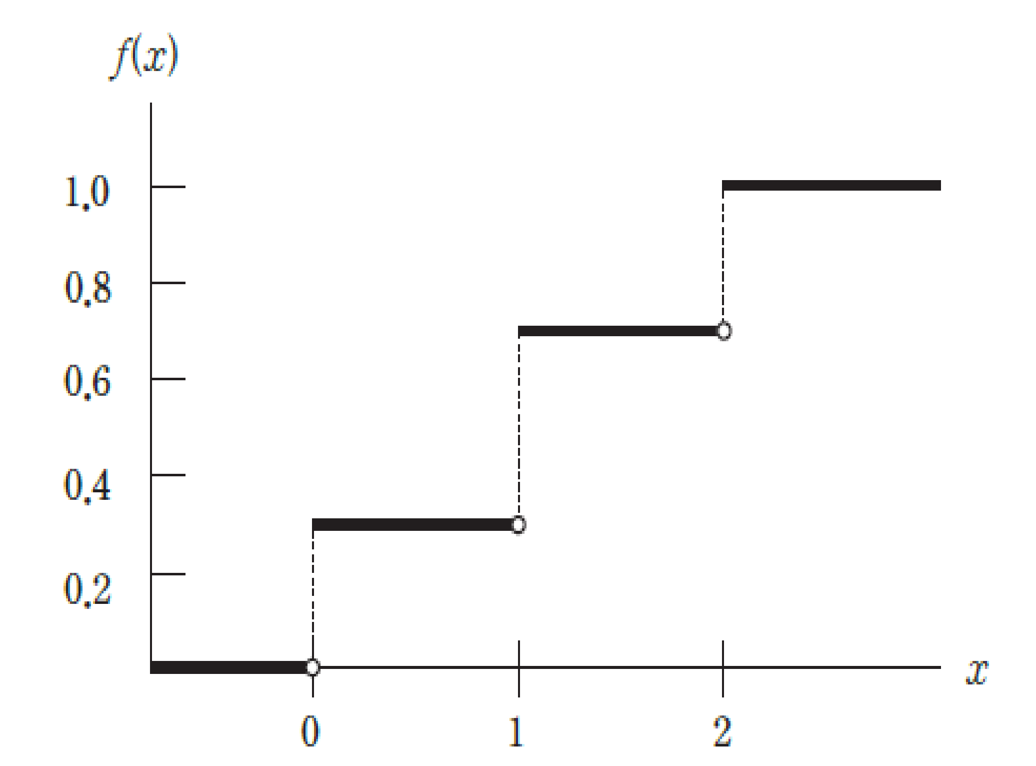
- The cumulative probability of $P(X \le x)$ as the value of random variable X increases is referred to as the **cumulative distribution function** $F(x)$.

1) Table style

$X = x$	$P(X \le x)$
0	1/4
1	3/4
2	4/4

2) Function style

$$F(x) = \begin{cases} 0, & x<0 \\ 1/4, & 0 \le x < 1 \\ 3/4, & 1 \le x < 2 \\ 1, & 2 \le x \end{cases}$$



5.3 Discrete Random Variable

[Example 5.3.1] There are 200 families living in a village. The number of visits to hospitals by each household over the past year is as follows. Obtain the probability distribution function and the cumulative distribution function of X = 'hospital visit'.

Hospital visit	0	1	2	3	4
Household	74	80	30	10	6

<Answer>

Probability distribution function

X = x	P(X=x)
0	0.37
1	0.40
2	0.15
3	0.05
4	0.03
Total	1.00

Cumulative distribution function

X = x	P(X≤x)
0	0.37
1	0.77
2	0.92
3	0.97
4	1.00

5.3 Discrete Random Variable

- If the possible values of the discrete random variable X are $x_1, x_2, \cdots, x_n$, mean and variance are also used for the measures of central tendency and dispersion. The mean of X, called usually expectation of X and denoted E(X) or μ, and the variance of X denoted as V(X) or σ^2 are defined as follows. The standard deviation of X, denoted σ, is the square root of the variance X.

$$E(X) = \mu = \sum_{i=1}^{n} x_i P(X=x_i)$$

$$V(X) = \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(X=x_i) = \sum_{i=1}^{n} x_i^2 P(X=x_i) - \mu^2$$

5.3 Discrete Random Variable

[Example 5.3.2] Find the expected value and variance of the random variable X = 'Number of Heads' when tossing a coin twice' which described in Table 5.3.2.

<Answer>

- Expectation and variance of X are as follows.

$$E(X) = \mu = \sum_{i=1}^{n} x_i P(X = x_i) = 0 \times \frac{1}{4} + 1 \times \frac{2}{4} + 2 \times \frac{1}{4} = 1$$

$$V(X) = \sum_{i=1}^{n} x_i^2 P(X = x_i) - \mu^2 = 0^2 \times \frac{1}{4} + 1^2 \times \frac{2}{4} + 2^2 \times \frac{1}{4} - 1^2 = \frac{1}{2}$$

5.3 Discrete Random Variable

- When knowing the expected value E(X) and variance V(X) of a random variable X, it is necessary to obtain the expected value and variance of aX+b of which a, b are constants.

Expectation and variance of aX + b

$$E(aX+b) = a\,E(X) + b$$

$$V(aX+b) = a^2 V(X)$$

5.3 Discrete Random Variable

[Example 5.3.3] The mean of a mid-term score on statistics was 60 points and the variance was 100. To adjust the score, the professor is thinking of the following alternative. Find the mean and variance of each alternative.

1) Add 20 points to each student's score.
2) Each student's score is multiplied by 1.4.
3) Multiply each student's score by 1.2 and add 10 points.

5.3 Discrete Random Variable

<Answer>

- X is the mid-term score and its mean and variance are E(X) = 60 and V(X) = 100.

1) Mean and variance of the new random variable X + 20 are as follows.
   $E(X + 20) = E(X) + 20 = 60 + 20$
   $V(X + 20) = V(X) = 100$
2) Mean and variance of the new random variable 1.4X are as follows.
   $E(1.4X) = 1.4\ E(X) = 1.4 \times 60 = 84$
   $V(1.4X) = 1.4^2\ V(X) = 1.96 \times 100 = 196$
3) Mean and variance of the new random variable 1.2X + 10 are as follows.e 1.4X.
   $E(1.2X + 10) = 1.2\ E(X) + 10 = 1.2 \times 60 + 10 = 82$
   $V(1.2X + 10) = 1.2^2\ V(X) = 1.44 \times 100 = 144$

5.3 Discrete Random Variable

☞ Standardized random variable

If the mean of a random variable X is μ, and the standard deviation is σ, then $Z = (X - \mu)/\sigma$ is a new random variable with a mean of 0 and a variance of 1. This new random variable is referred to as a standardised random variable.

5.3 Discrete Random Variable

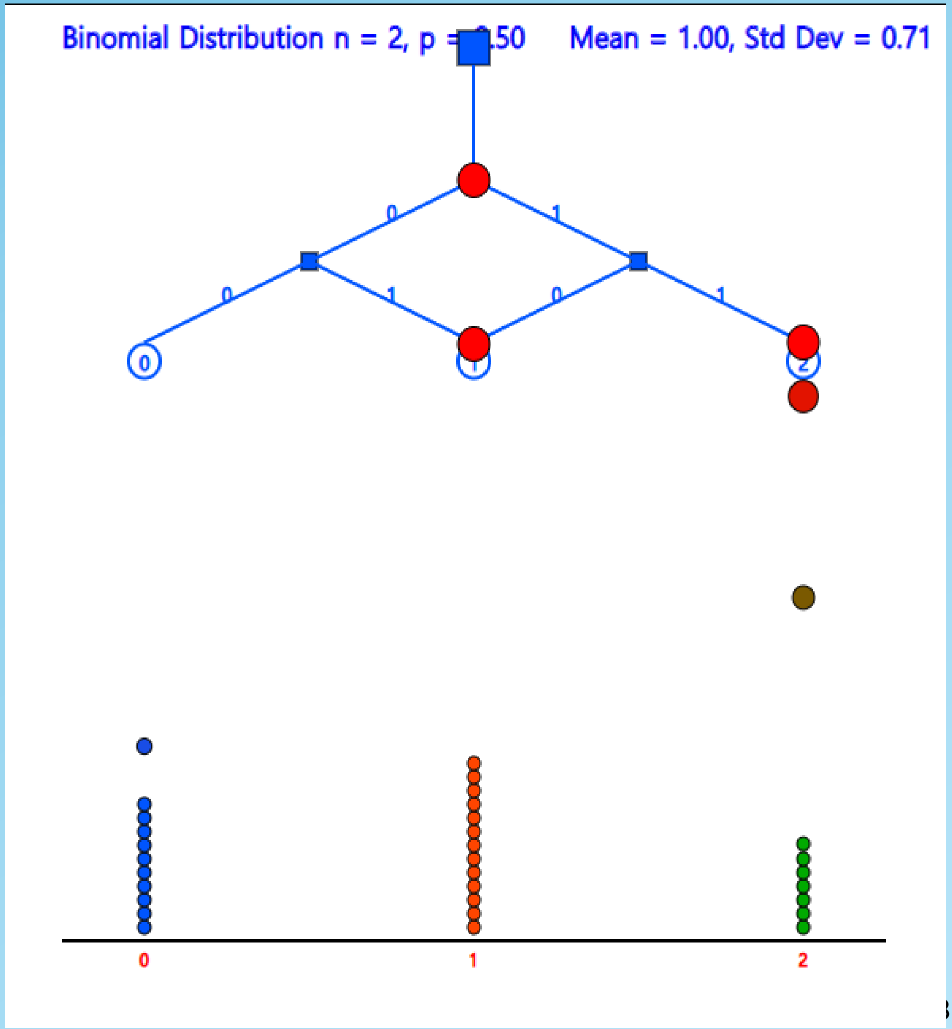
5.3.1 Binomial Distribution

- Examples similar to experiment that examine how many times the head comes out by tossing coins are observed very much around us.
  - Products are inspected and classified as defective or good.
  - Ask one voter about the pros and cons of a particular candidate.
- These experiments do not know what the results will be, but there are two possible outcomes such as {defective, normal}, {pros, cons}. However, the probability of outcomes in each experiment is different.
- These experiments are specifically called the Bernoulli trial, and one outcome of interest in the two is often referred to as 'success' and the other as 'failure'.
- Bernoulli trials are repeated and the number of 'success' is counted.
  - Throw a coin five times and examine the number of heads.
  - Inspect 100 products and count the number of defective products.
  - Count the number of voters in favor of a particular candidate among 50 voters.

5.3 Discrete Random Variable

5.4.1 Binomial Distribution - Simulation

- Drop a ball from the top and if it hit a bar, it has one-half chance to fall to the left (zero point) or right (one point).
- The dropped ball again falls to the left and right with a 1/2 chance.
- When you drop 100 balls, examine the sum of the total scores.



5.3 Discrete Random Variable

5.3.1 Binomial Distribution

- The 'counting of success' when performing this independently repeated Bernoulli trial with the same probability of success is called a binary random variable, and its distribution is called a binomial distribution.

[Ex 5.3.4] Four more baseball games will be played by the Tiger this season. If the Tiger has a 60% chance of winning every game, what is the probability of

1) losing all of them?
2) winning only once?
3) winning twice?
4) winning three times?
5) winning all four times?
6) Find the probability distribution function of the random variable X = 'the number of games the tiger wins'.

5.3 Discrete Random Variable

<Answer of Ex 5.3.4>

- This problem is the enforcement of Bernoulli trial in each game of 'win' and 'fail'. This Bernoulli trial is repeated four times. The sample space is all about winning or losing game and there are $2^4 = 16$all element. The sample space is shown as follows by marking the winning in O and the losing in X.

S = {'XXXX','OXXX','XOXX','XXOX','XXXO','OOXX','OXOX','OXXO',
'XOOX','XOXO','XXOO','OOOX','OOXO','OXOO','XOOO','OOOO'}

1) The event that the Tiger will loose all game is {'XXXX'} and the probability of this event iis (0.4)×(0.4)×(0.4)×(0.4) = $(0.4)^4$.
2) There are four events that the Tiger is winning once and losing three times such as {'OXXX', 'XOXX', 'XXOX', 'XXXO'}. These four cases are equal to the number of O's in a single seat when there are four seats which is ${}_4C_1$. Since the probability of each event is (0.6)×(0.4)×(0.4)×(0.4), the probability of the Tiger winning once is ${}_4C_1(0.6)(0.4)^3$.
3) There are six events that the Tiger is winning two times and losing two times such as {'OOXX', 'OXOX', 'OXXO', 'XOOX', 'XOXO', 'XXOO'}. These six cases are equal to the number of O's in two seats when there are four seats which is ${}_4C_2$. Since the probability of each event is (0.6)×(0.6)×(0.4)×(0.4), the probability of the Tiger winning twice is ${}_4C_2(0.6)^2(0.4)^2$.

5.3 Discrete Random Variable

<Answer of Ex 5.3.4>

4) There are four events that the Tiger is winning three times and losing one time such as {'OOOX', 'OOXO', 'OXOO', 'XOOO'}. These four cases are equal to the number of O's in three seats when there are four seats which is ${}_4C_3$. Since the probability of each event is (0.6)×(0.6)×(0.6)×(0.4), the probability of the Tiger winning three times is ${}_4C_3(0.6)^3(0.4)^1$.

5) There is one event that the Tiger is winning four times such as {'OOOO'}. This one case is equal to the number of O's in four seats when there are four seats which is ${}_4C_4$. Since the probability of each event is (0.6)×(0.6)×(0.6)×(0.6), the probability of the Tiger winning all four times is ${}_4C_4(0.6)^4$.

6) The probability distribution function of the random variable X = 'the number of games the Tiger wins' is a summary of the above probabilities.

x	P(X=x)	
0	${}_4C_0(0.4)^4$	= 0.0256
1	${}_4C_1(0.6)(0.4)^3$	= 0.1536
2	${}_4C_2(0.6)^2(0.4)^2$	= 0.3456
3	${}_4C_3(0.6)^3(0.4)$	= 0.3456
4	${}_4C_4(0.6)^4$	= 0.1296

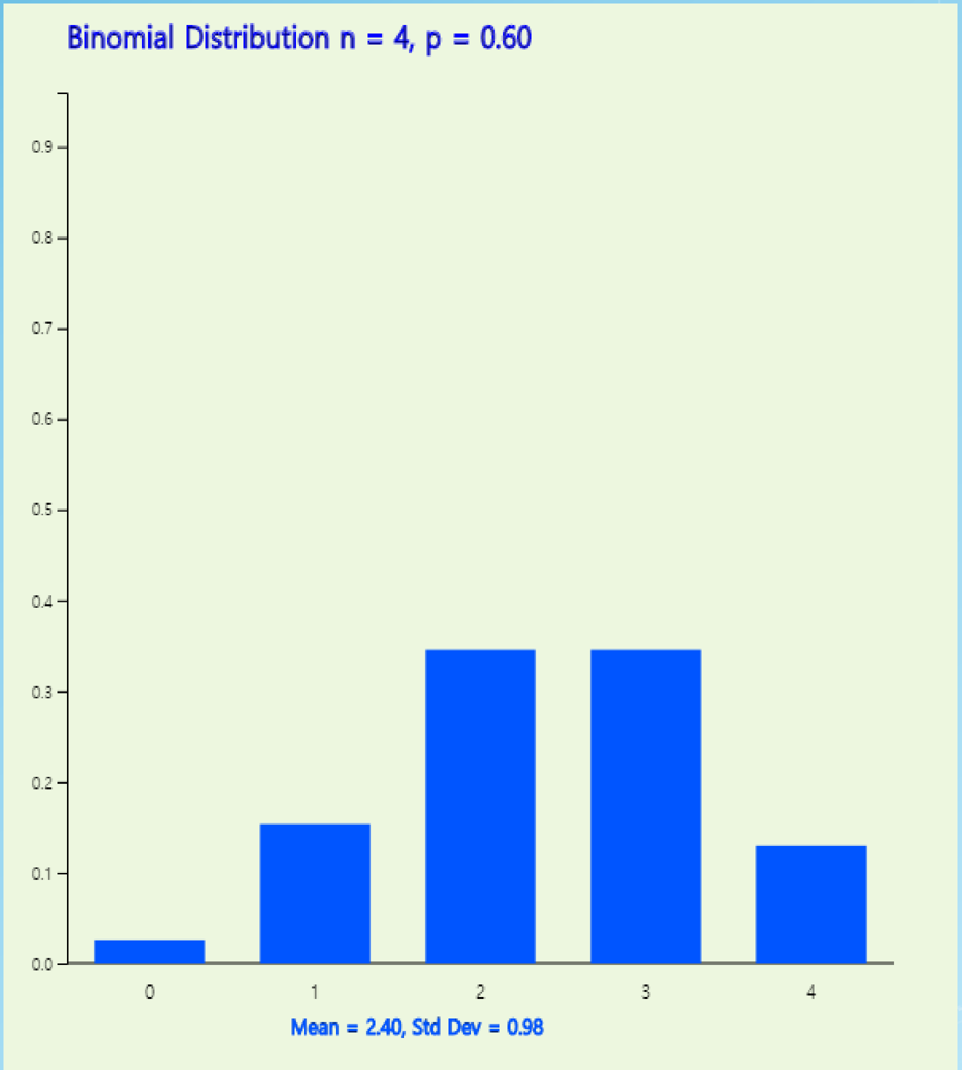
5.3 Discrete Random Variable

[Ex 5.3.5] By using『eStatU』, obtain the probability and probability distribution function of [Exe 5.3.4].

<Answer>

- Select a binomial distribution from 『eStatU』
- Enter n = 4, p = 0.6 and press the [Execute] button to display a binomial function graph.
- Click the 'Binary table' button.

n = 4	p = 0.600		
x	P(X = x)	P(X x)	P(X x)
0	0.0256	0.0256	1.0000
1	0.1536	0.1792	0.9744
2	0.3456	0.5248	0.8208
3	0.3456	0.8704	0.4752
4	0.1296	1.0000	0.1296



5.3 Discrete Random Variable

5.3.1 Binomial Distribution

☞ Binomial Distribution

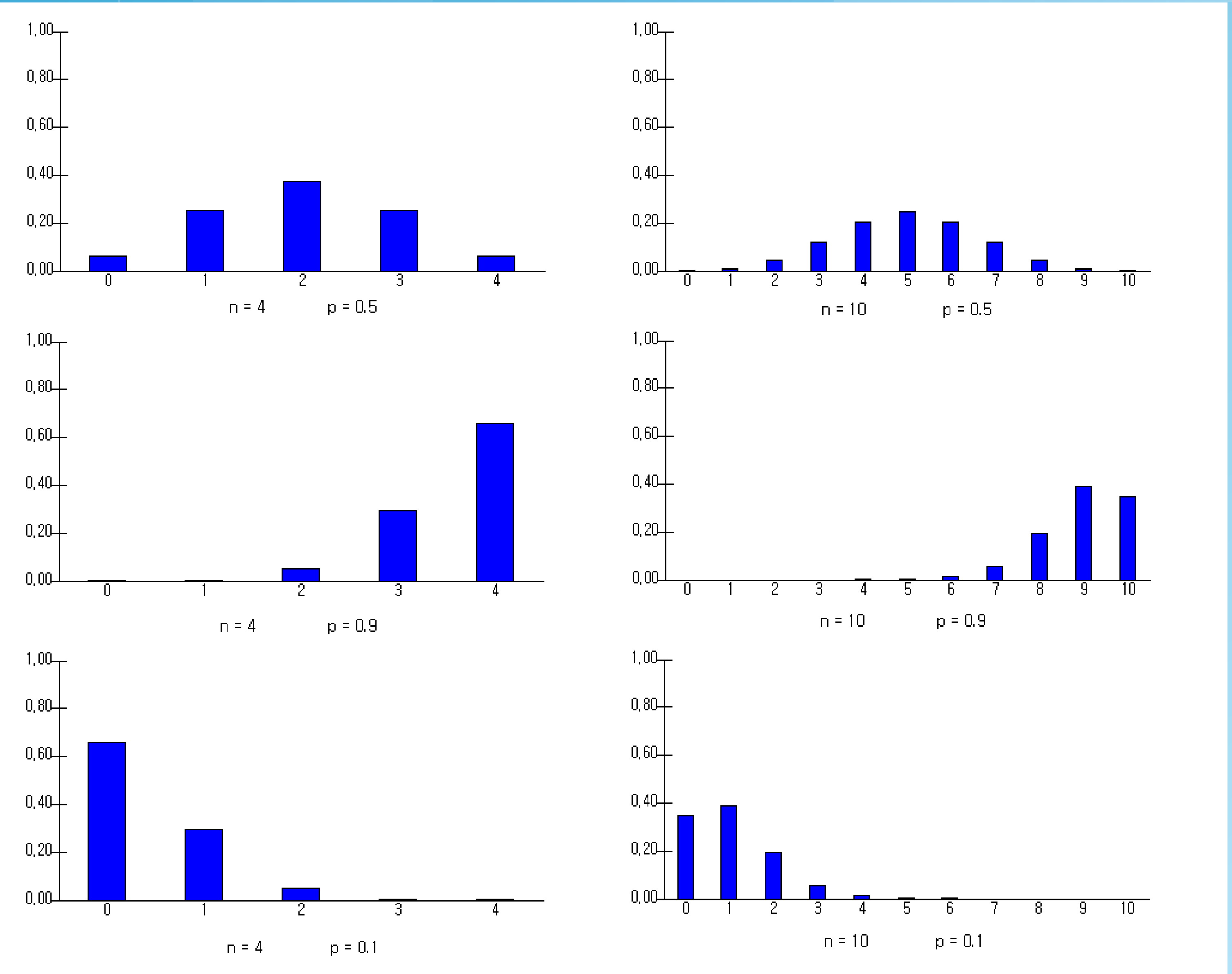
If the probability of success is p in a Bernoulli trial and the trial is repeated n times independently, the probability distribution function that the random variable X = 'the number of success' is x is as follows. It is called a binomial distribution and denoted as $B(n,p)$.

$$f(x) = {}_nC_x \; p^x \; (1-p)^{n-x} \quad , \qquad x = 0,1,2,\dots,n$$

The expectation and variance of the binomial distribution are E(X) = np, V(X) = $np(1-p)$.

5.3 Discrete Random Variable

5.3.1 Binomial Distribution



5.3.1 Binomial Distribution

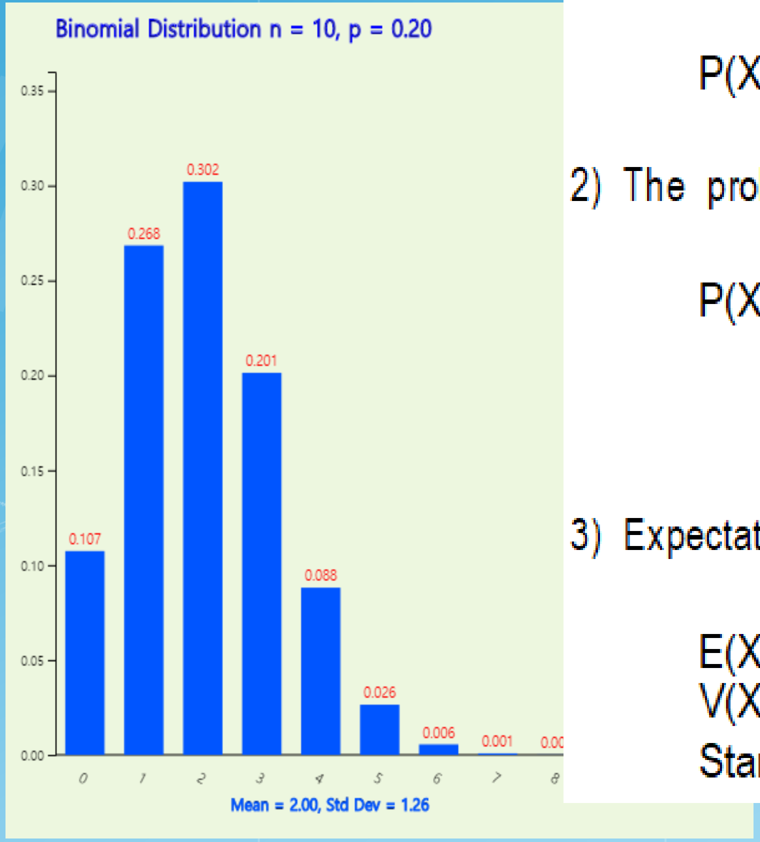
[Ex 5.3.6] Past experience shows that a salesperson from an insurance company has a 20% chance of meeting a customer and insuring that person. The salesperson is scheduled to meet 10 customers this morning. Calculate the following probabilities directly and check using『eStatU』

1) What are the probability that three customers will get insurance?
2) What is the probability that two or more customers will get insurance?
3) How many people on average would sign up? And its standard deviation?

5.3 Discrete Random Variable

5.3.1 Binomial Distribution

<Answer of 5.3.6>



- This is a Binomial distribution when n = 10, p = 0.2.

1) The probability that three customers will get insurance is as follows.

$$P(X=3) = {}_{10}C_3 (0.2)^3 (1-0.2)^{10-3} = 0.2013$$

2) The probability that two or more customers will get insurance may use the complement event as follows.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - {}_{10}C_0 (0.2)^0 (1-0.2)^{10} - {}_{10}C_1 (0.2)^1 (1-0.2)^{10-1} \\ &= 1 - 0.1074 - 0.2684 = 0.6242 \end{aligned}$$

3) Expectation and standard deviation are as follows..

$$E(X) = np = 10 \times 0.2 = 2$$
$$V(X) = np(1-p) = 10 \times 0.2 \times 0.8 = 1.6$$
$$\text{Standard deviation} = \sqrt{1.6} = 1.265$$

5.3 Discrete Random Variable

5.3.2 Poisson Distribution

- Consider the following examples that are frequently observed around us.
  - Number of calls made to an office between 9 and 10 a.m. daily.
  - Number of one-day traffic accidents occurring at a certain intersection.
  - Number of defective spots per unit area of the fabric.
- A random variable that represents this 'occurrence of events per unit time or unit area' is Poisson random variable and its distribution is Poisson distribution.
- Poisson distribution is used in many areas, and some examples are as follows.
  - Demand for a product sold every day at a certain store
  - The number of typos that occur on each page of a book
  - Number of accidents occurring during a week in a factory
  - Number of defectives per unit length of cloth
  - Number of radioactive particles released from radioactive materials

5.3 Discrete Random Variable

5.3.2 Poisson Distribution

Poisson Distribution

The distribution of a Poisson random variable X = 'Occurrence of success event per unit time or unit area' is as follows when the average number of success is λ.

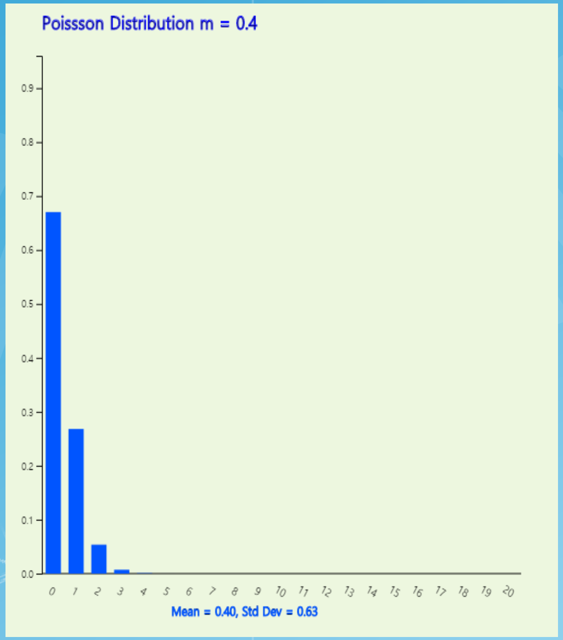
$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0,1,2, \ldots$$

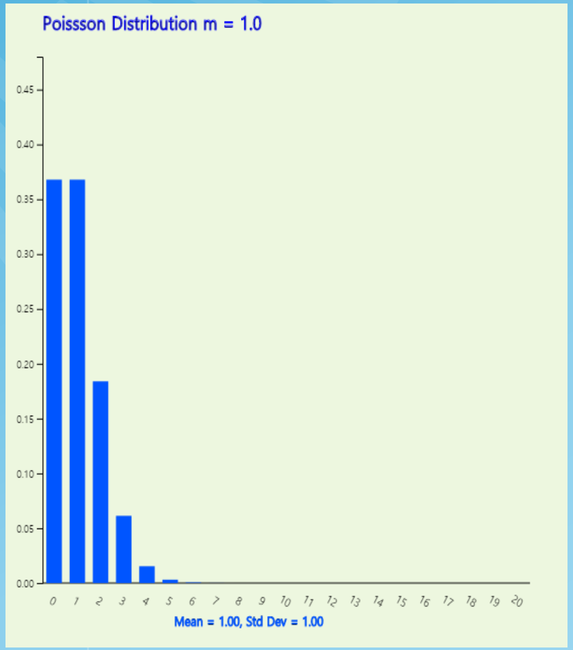
The expectation and variance of the Poisson random variable are as follows:

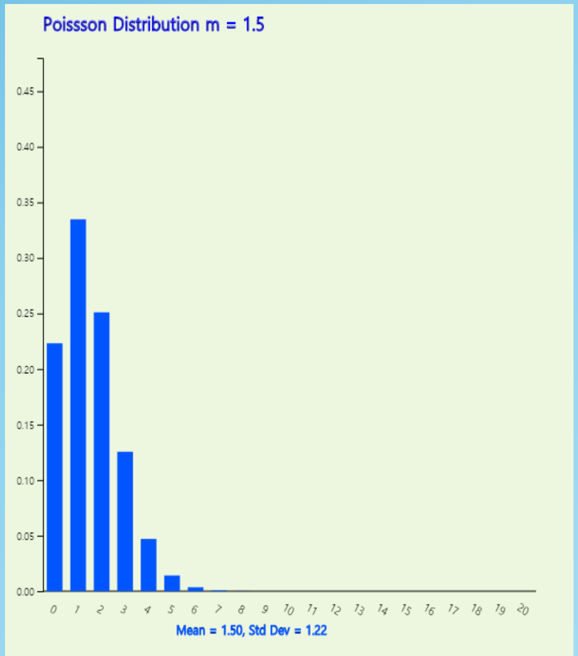
$$E(X) = \lambda$$
$$V(X) = \lambda$$

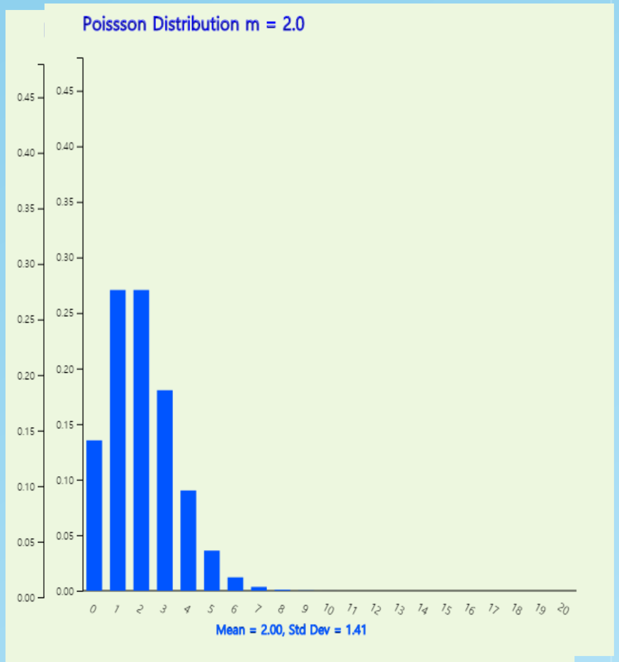
5.3 Discrete Random Variable

5.3.2 Poisson Distribution









5.3.2 Poisson Distribution

[Ex 5.3.8] Assume that cars per minute arriving at a highway toll gate during rush hour is the Poisson distribution of an average of five cars per one minute. One day, if you observe the toll gate for one minute during rush hour, calculate the following probabilities.
1) What is the probability that none of the cars will arrive?
2) What is the probability of five cars arriving?
3) What is the probability of more than two cars arriving?

Answer

- Let X be the Poisson random variable with $\lambda = 5$.

1) $P(X = 0) = f(0) = \dfrac{e^{-5}5^0}{0!} = 0.0067$

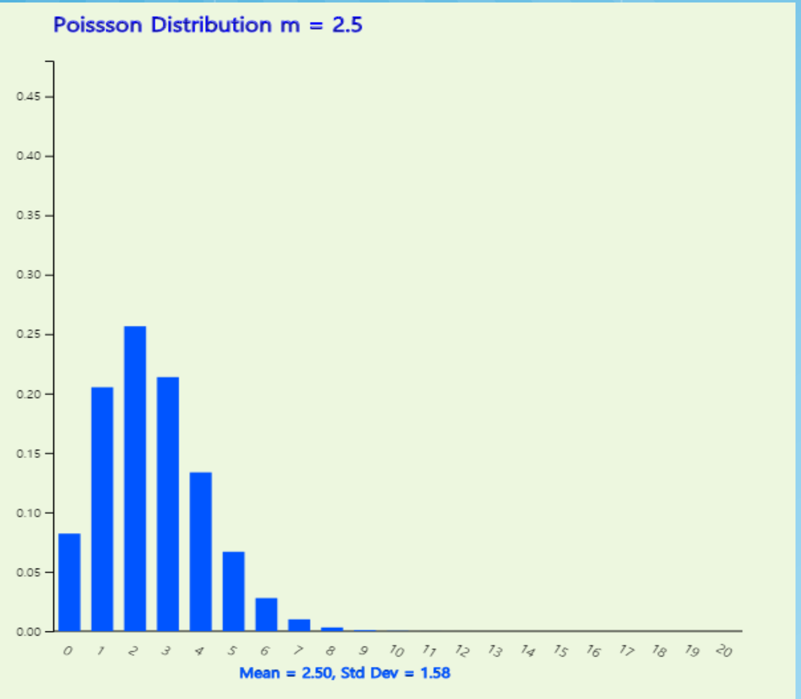
2) $P(X = 5) = f(5) = \dfrac{e^{-5}5^5}{5!} = 0.1755$

3) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X=0) - P(X=1)$
$= 1 - 0.0067 - 0.0337 = 0.9596$

5.3 Discrete Random Variable

[Ex 5.3.9] Assume that the average number of Typhoons passing through the southern part of the country is m = 2.5 times per year. Check the following probabilities using 『eStatU』.

1) What is the probability that a Typhoon will pass once this year?
2) What is the probability that this year's Typhoon will pass twice or four times?
3) What is the probability that this year's Typhoon will pass more than once?



m = 2.5			
x	P(X = x)	P(X x)	P(X x)
0	0.0821	0.0821	1.0000
1	0.2052	0.2873	0.9179
2	0.2565	0.5438	0.7127
3	0.2138	0.7576	0.4562
4	0.1336	0.8912	0.2424

5.3 Discrete Random Variable

<Answer of Ex 5.3.9>

- Let X be the Poisson random variable with m = 5.

1) $P(X = 0) = f(0) = \dfrac{e^{-5}5^0}{0!} = 0.0067$

2) $P(X = 5) = f(5) = \dfrac{e^{-5}5^5}{5!} = 0.1755$

3) $P(X \ge 2) = 1 - P(X \le 1) = 1 - P(X=0) - P(X=1)$

$= 1 - 0.0067 - 0.0337 = 0.9596$

5.3 Discrete Random Variable

5.3.3 Geometric Distribution

- The number of trials until the head of a coin appears may be of interest.
  - A candidate has a 40 percent approval rating in an election. When interviewing voters to hear opinions from those who oppose, what is the probability of finding someone who disagrees to meet at the fifth trials?
  - The defect rate in a factory-produced product is said to be about 5%. If you continue to inspect the product until you find a defect product to investigate the cause, what is the probability of finding it in 10^{th} trials?
- We don't know what the outcome of each experiment will be, but there are only two possible outcomes such as {probable, opposite}, {defective, normal}.
- If we call one outcome of interest is 'success' and the other is 'failure', the experiment is the repetition of Bernoulli trials until we have one 'success'. Number of Bernoulli trials until success is called the geometrical random variable and its distribution is the geometric distribution

5.3 Discrete Random Variable

5.3.3 Geometric Distribution

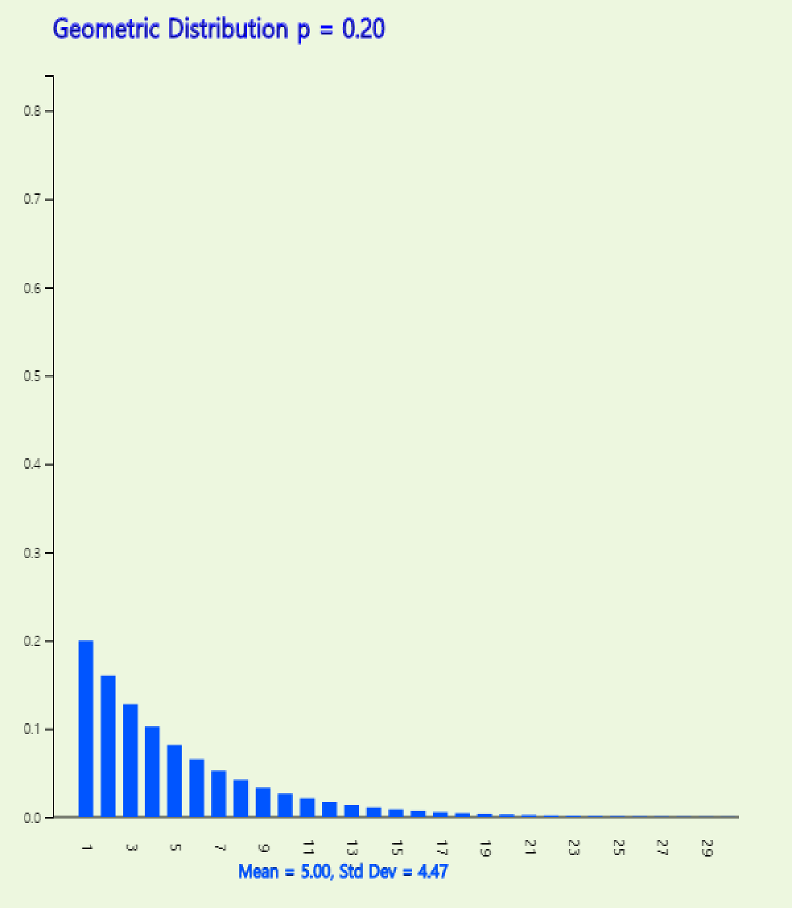
When the probability of 'success' in a Bernoulli trial is p and X is the number of Bernoulli trials until one success, the probability distribution of X is called a geometric distribution and its probability density function is as follows:

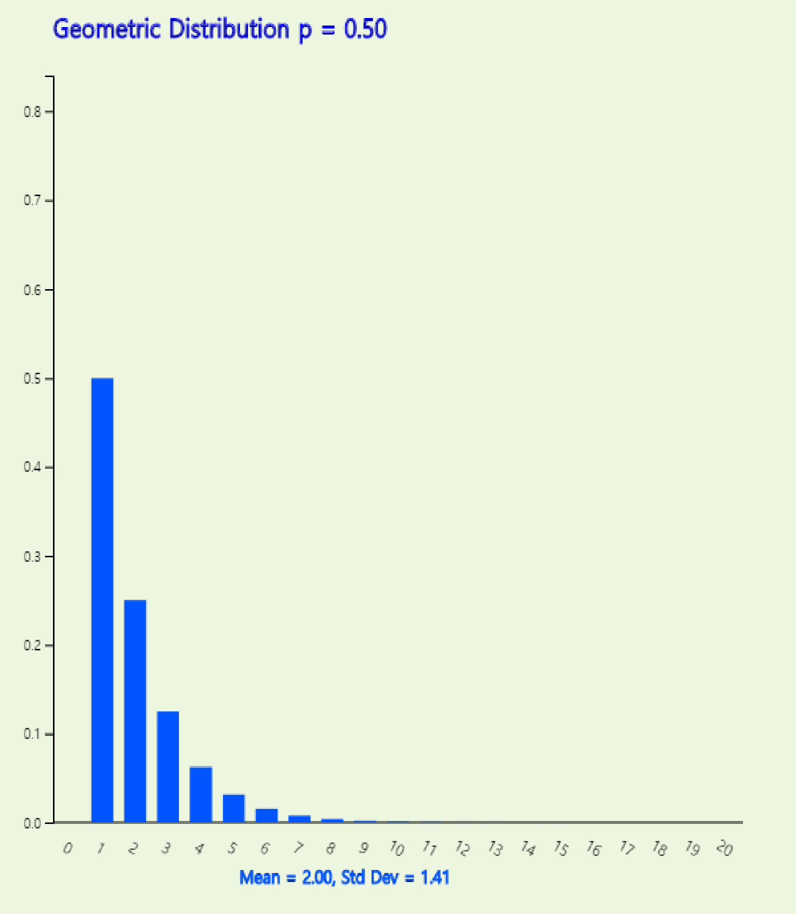
$$f(x) = (1-p)^{x-1}p, \qquad x = 1, 2, \ldots$$

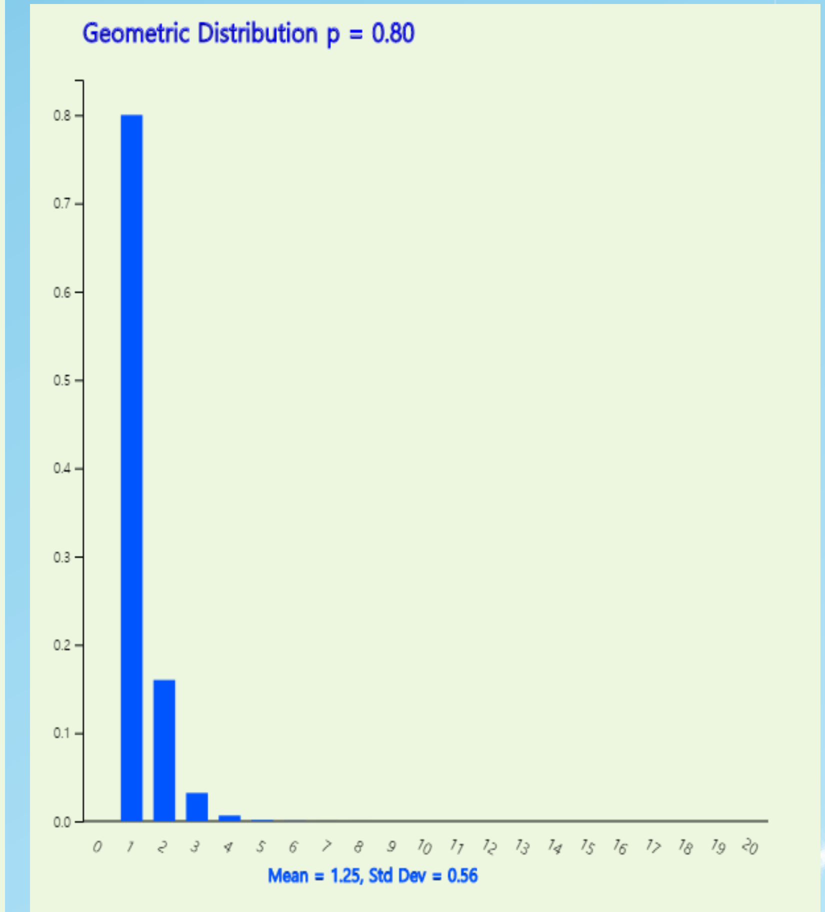
The expection and variance of X are $E(X) = \dfrac{1}{p}$, $V(X) = \dfrac{1-p}{p^2}$.

5.3 Discrete Random Variable

5.3.3 Geometric Distribution







5.3.3 Geometric Distribution

[Example 5.3.10] A candidate has a 40% approval rating in an election. When interviewing voters to hear the opinions of those who oppose the candidate, look for the next probabilities.

1) What is the probability of finding someone who is opposed in the first interview?
2) What is the probability of finding someone who is opposed in the fifth interview?

<Answer>

- Let X be the Geometric random variable with p = 0.4.

1) $P(X = 1) = f(1) = (1-0.4)^{1-1}0.4 = 0.4$

2) $P(X = 5) = f(5) = (1-0.4)^{5-1}0.4 = 0.0518$

5.3 Discrete Random Variable

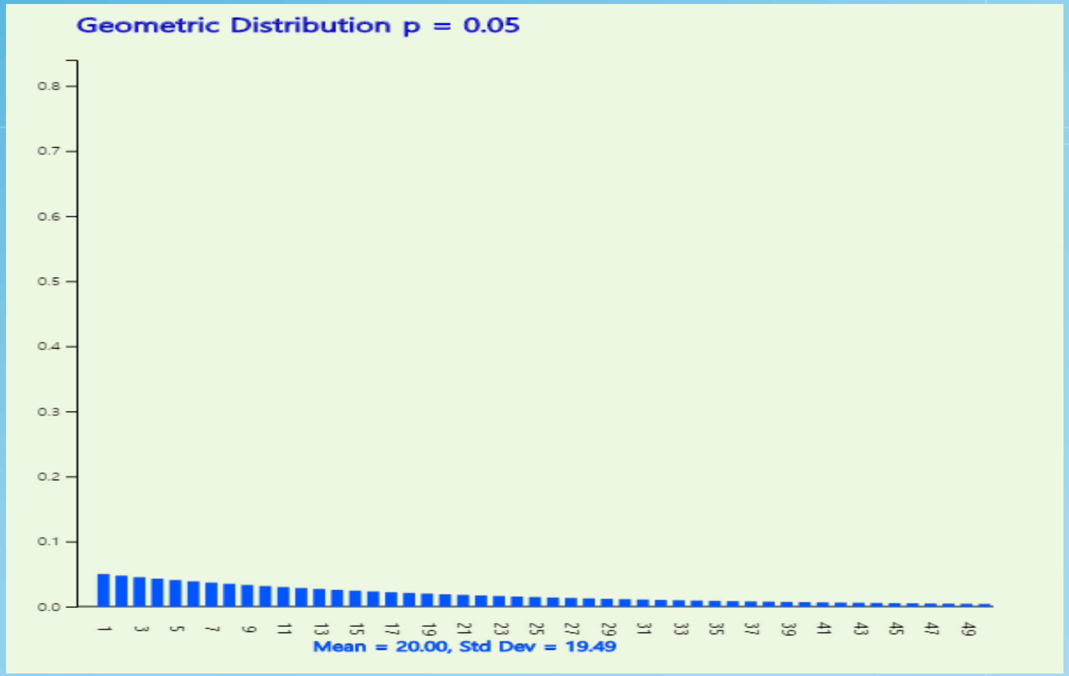
[Ex 5.3.11] The defect rate in a factory-produced product is said to be about 5 %. Use 『eStatU』 to obtain the following probabilities when continuing to inspect the product until it finds a defective product to investigate the cause of defective.

1) The probability of finding a defective product in three times,
2) What is the probability of finding defective products at least 3 times?

<Answer>

1) $P(X=3) = 0.0451$.
2) find $P(X \geq 3) = 0.9025$

$$P(X \geq 3) = 1 - P(X \leq 2)$$
$$= 1 - 0.0975 = 0.9025$$



p = 0.05			
x	P(X = x)	P(X x)	P(X x)
1	0.0500	0.0500	1.0000
2	0.0475	0.0975	0.9500
3	0.0451	0.1426	0.9025
4	0.0429	0.1855	0.8574
5	0.0407	0.2262	0.8145

5.3.4 HyperGeometric Distribution

- Consider a box consisting of 20 products and 15 of them are normal products and 5 are defective products. When three of the 20 products are sampled, the probability of having one normal products and two defective product is calculated using the combination as follows:

$$\frac{{}_{15}C_1 \times {}_{5}C_2}{{}_{20}C_3}$$

- The random variable that counts the number of ‘success’ in the finite population consisting of only ‘success’ and ‘failure’ is called **hypergeometric random variable** and its distribution is called **hypergeometric distribution**.

5.3 Discrete Random Variable

5.3.4 HyperGeometric Distribution

Hypergeometric Distribution

Consider a population of size N which consists of D 'success' and $N-D$ 'failure'. If we collect a sample of size n without replacement and X is the number of 'success' in the sample, then the distribution of X is called hypergeometric distribution and its probability distribution function is as follows:

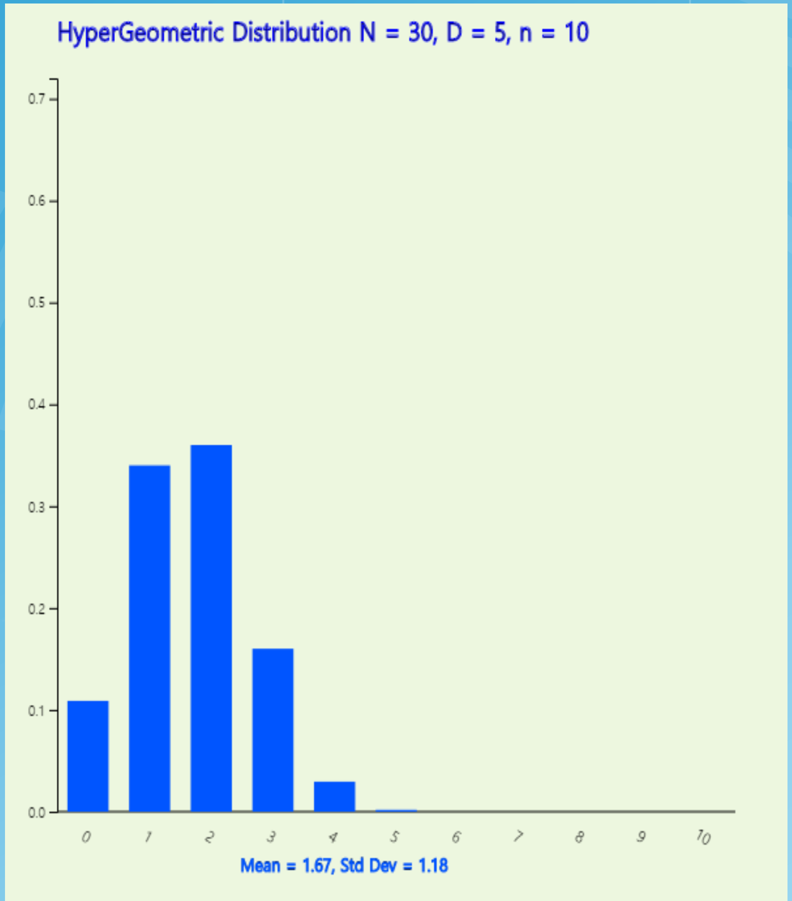
$$f(x) = \frac{{}_{D}C_{x} \; {}_{N-D}C_{n-x}}{{}_{N}C_{n}}$$

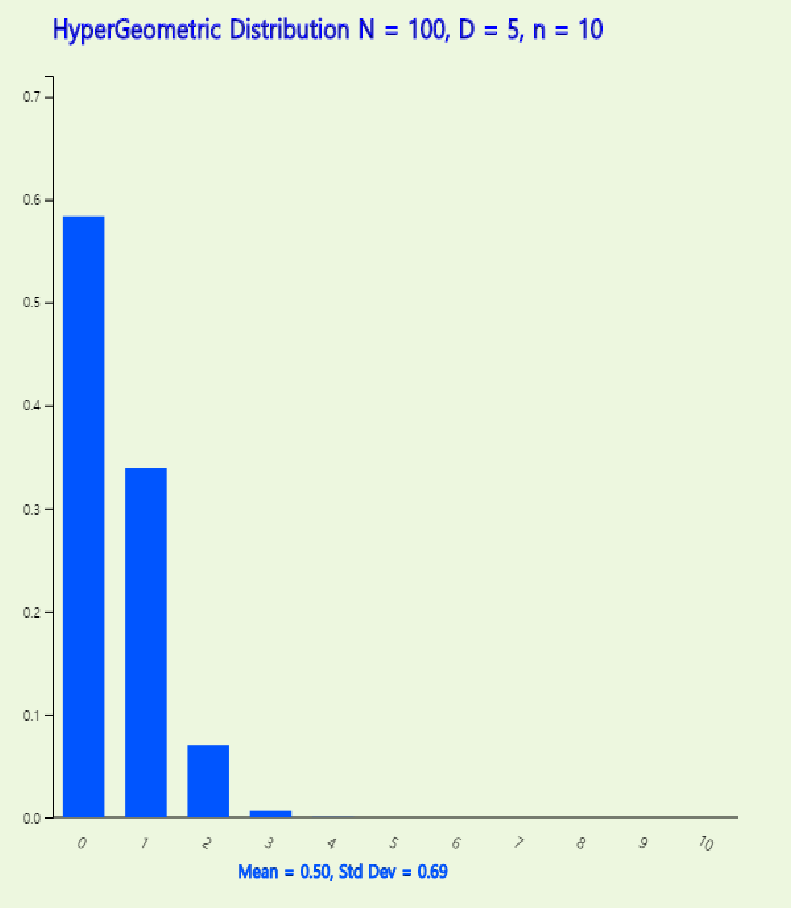
If we let $p = D/N$, the expectation and variance of the hypergeometric random variable are as follows:

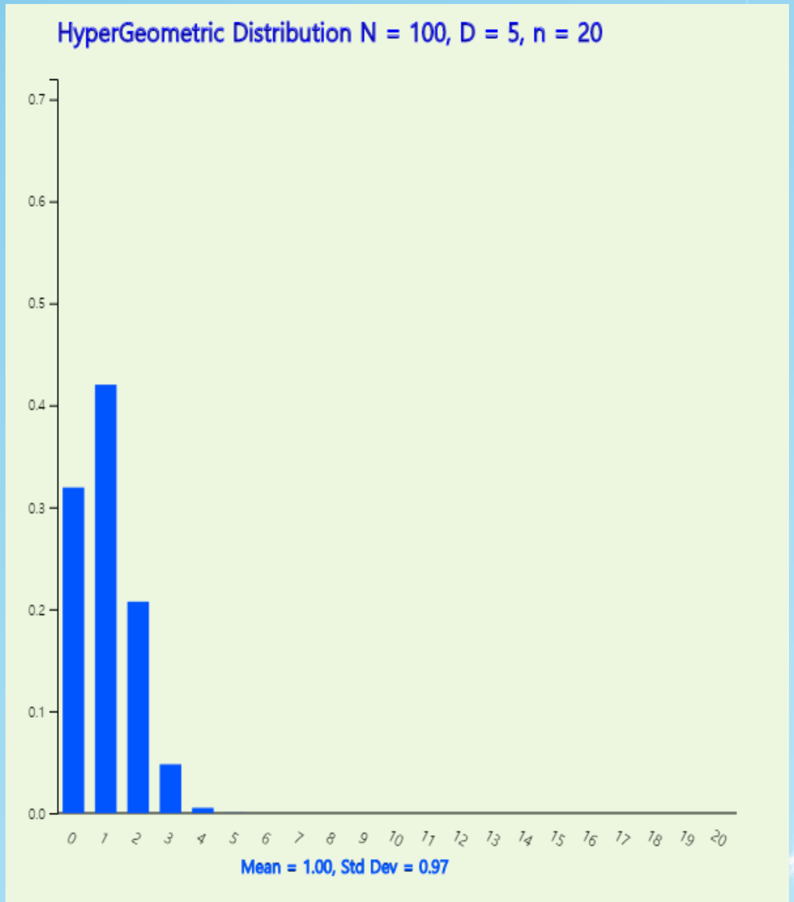
$$E(X) = np, \quad V(X) = np(1-p)\frac{N-n}{N-1}$$

5.3 Discrete Random Variable

5.3.4 HyperGeometric Distribution







5.3.4 HyperGeometric Distribution

[Ex 5.3.12] Sample of size 3 is selected from a box containing 20 tobacco products of which there are 15 normal products and 5 defective products. What is the probability of having one, two, or three defectives in the sample?

<Answer>

- These probability calculations have already been learned using combinations in section 5.1. This is the hypergeometric distribution with N = 20, D = 15, n = 3 , so it is as follows.

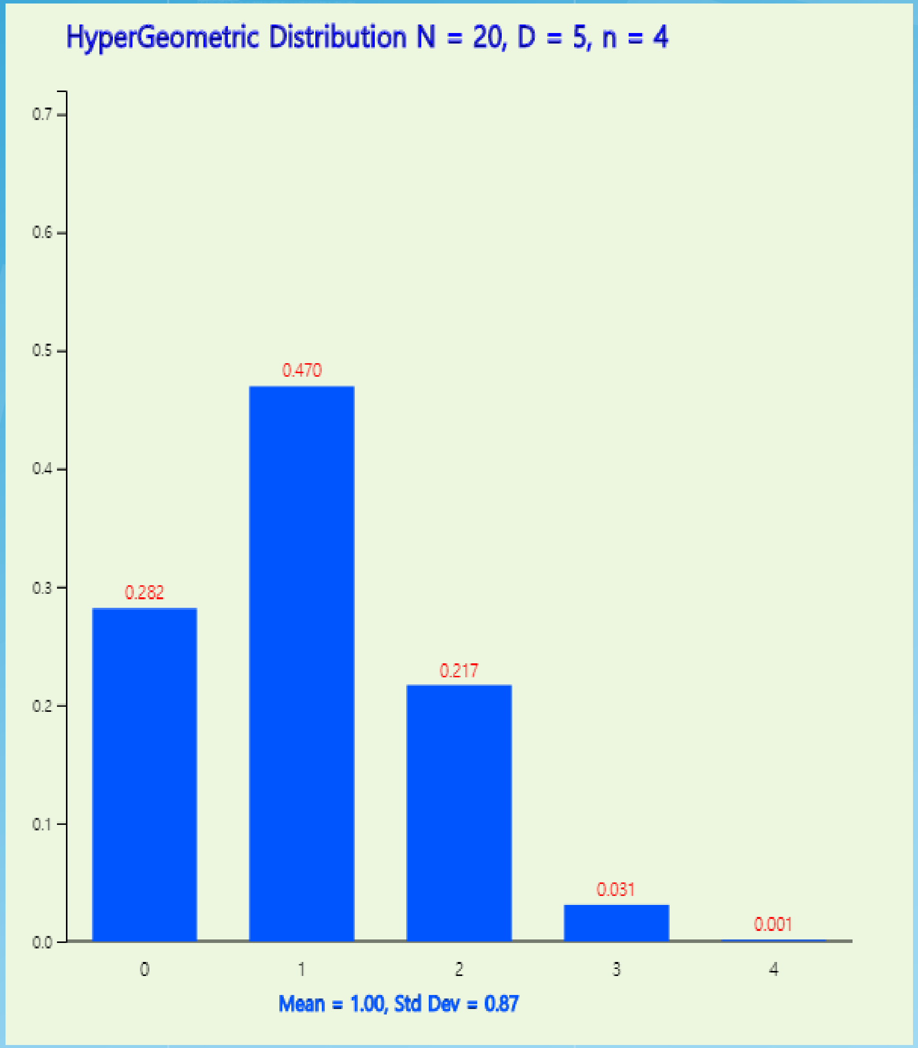
$$P(X=1) = \frac{{}_{15}C_2 \times {}_5C_1}{{}_{20}C_3} = \frac{15 \times 10}{1140} = 0.460$$

$$P(X=2) = \frac{{}_{15}C_1 \times {}_5C_2}{{}_{20}C_3} = \frac{105 \times 5}{1140} = 0.132$$

$$P(X=3) = \frac{{}_{15}C_0 \times {}_5C_3}{{}_{20}C_3} = \frac{455 \times 1}{1140} = 0.099$$

5.3 Discrete Random Variable

5.3.3 HyperGeometric Distribution



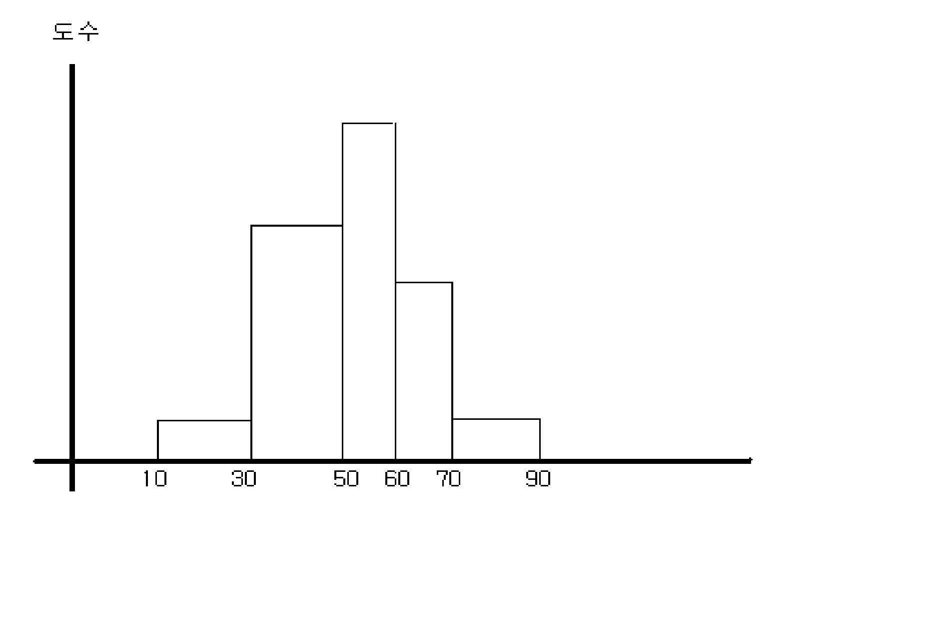
N = 20	D = 5	n = 3	
x	P(X = x)	P(X x)	P(X x)
0	0.3991	0.3991	1.0000
1	0.4605	0.8596	0.6009
2	0.1316	0.9912	0.1404
3	0.0088	1.0000	0.0088

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 place if the traffic is not congested.
- While the results of these experiments will usually be a real number near 30 minutes, define a random variable X as the 'time to work place'. As such, when there are an infinite number of possible values for a random variable and it cannot be counted, this is called a continuous random variable.
- Probability calculation at each point is meaningless because there are infinite values in the continuous random variable, and the probability at one point is considered zero.
- Instead of the probability of a single value, the probability of an interval is of interest, such as 'What is the 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

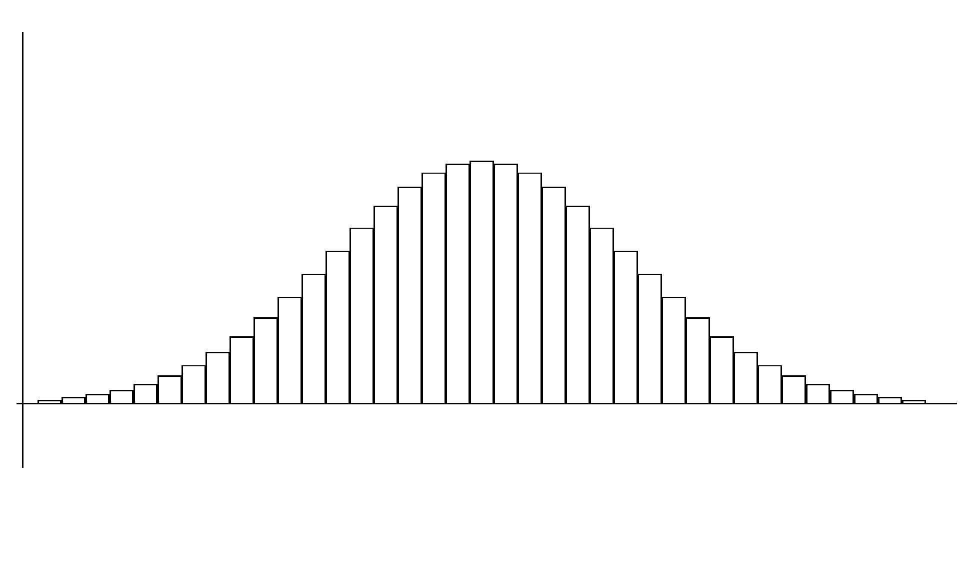


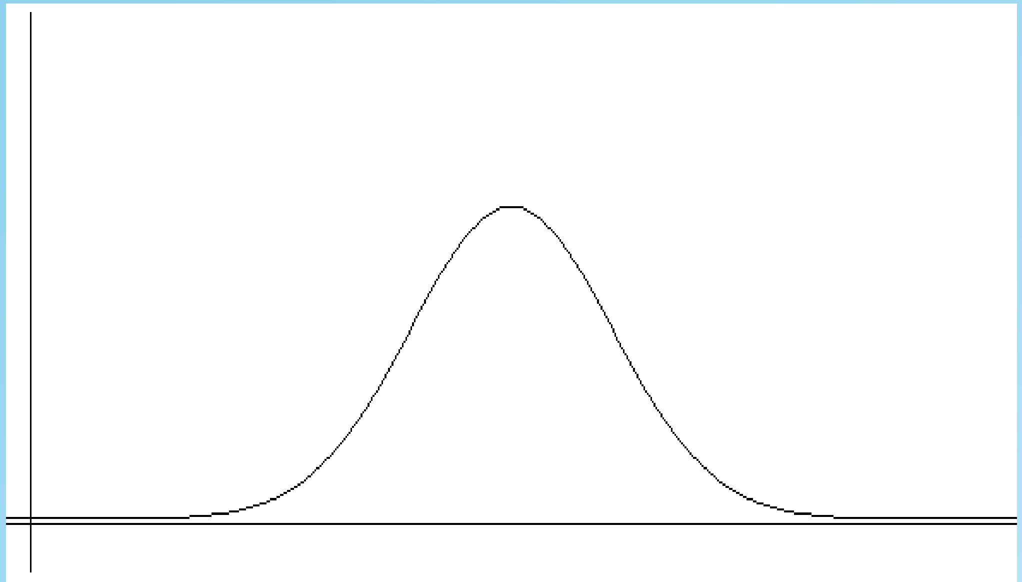
- **Using this frequency table, the calculation of the 'probability of commuting time between 30 and 60 minutes' is as follows.**

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

5.4 Continuous Random Variable

- However, if you use this table, you cannot calculate the probability of the commuting time between 25 and 35 minutes. In order to calculate this probability calculation 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 the width of the interval, this histogram will be approximated to the continuous. This function is called a probability distribution function of a continuous random variable.





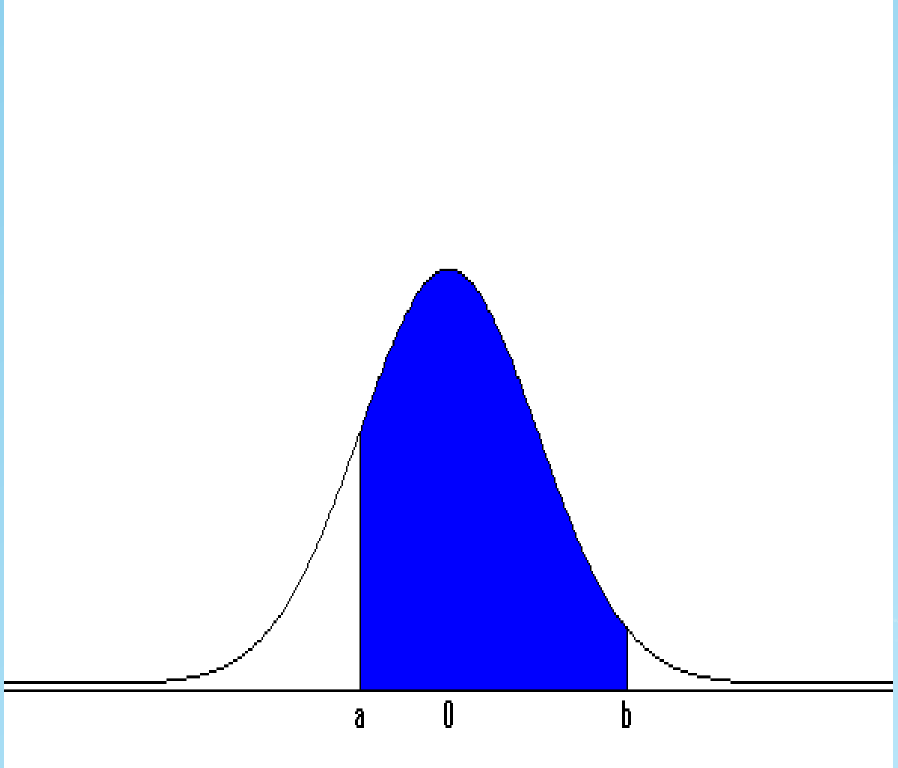
5.4 Continuous Random Variable

- If the probability distribution function of continuous random variable can be expressed as a mathematical 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

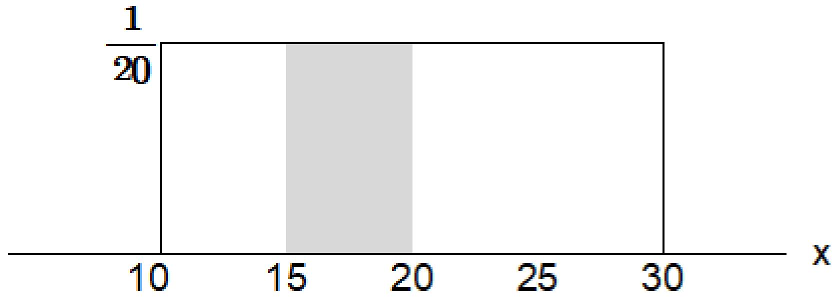
[Ex 5.4.1] The time it takes to order a pizza and get home has the same possibility as any time from 10 to 30 minutes. Let the random variable X be the time it takes to deliver a pizza. Find the probability distribution function of X and draw a picture. Find the probability of delivery between 15 and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

$$P(15 < X < 20) = (20 - 15) \times (1 / 20) = 0.25$$

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



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

5.4.1 Normal Distribution

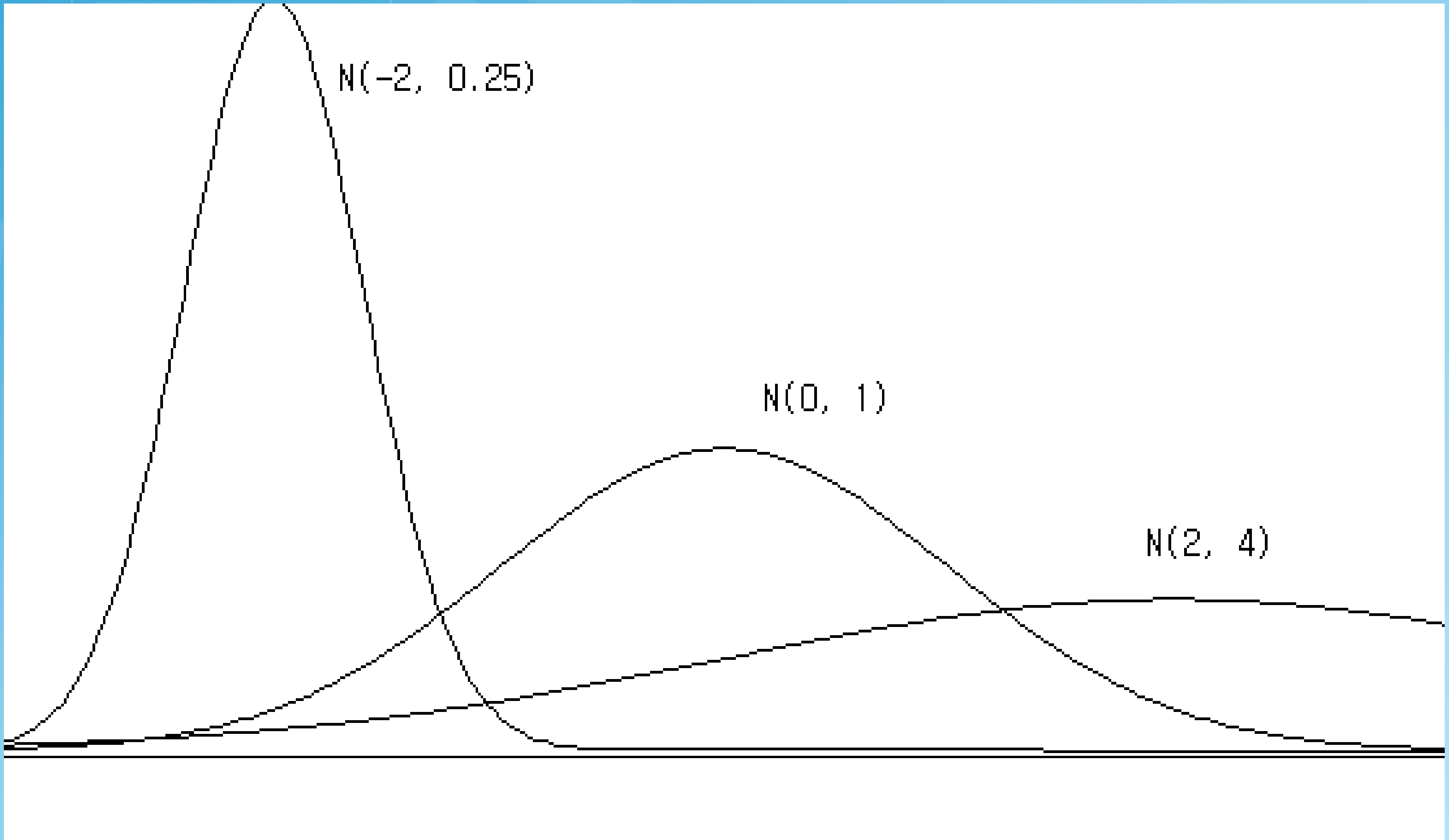
5.4.1 Normal Distribution

- In real life, the continuous data that appears more often in the form of a bell-shape which shows a large collection of data near the mean, with fewer data as it moves away from the mean, and is symmetrical around the mean.
- This type of data is called a normal distribution. Data obtained such as height, weight, and length of bolt are often in normal distribution.
- To make it easier to calculate probability for this type of data, mathematicians tried to find a function to describe this distribution type. Abraham de Moivre (1667-1754) was first discovered the function and then Carl Friedrich Gauss (1777-1855) extensively applied to physics and astronomy. This function is as follows and is called a normal distribution function or a Gaussian function

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} exp[-\frac{(x-\mu)^2}{2\sigma^2}] \qquad -\infty < x < \infty$$

5.4 Continuous Random Variable

5.4.1 Normal Distribution



5.4 Continuous Random Variable

5.4.1 Normal Distribution

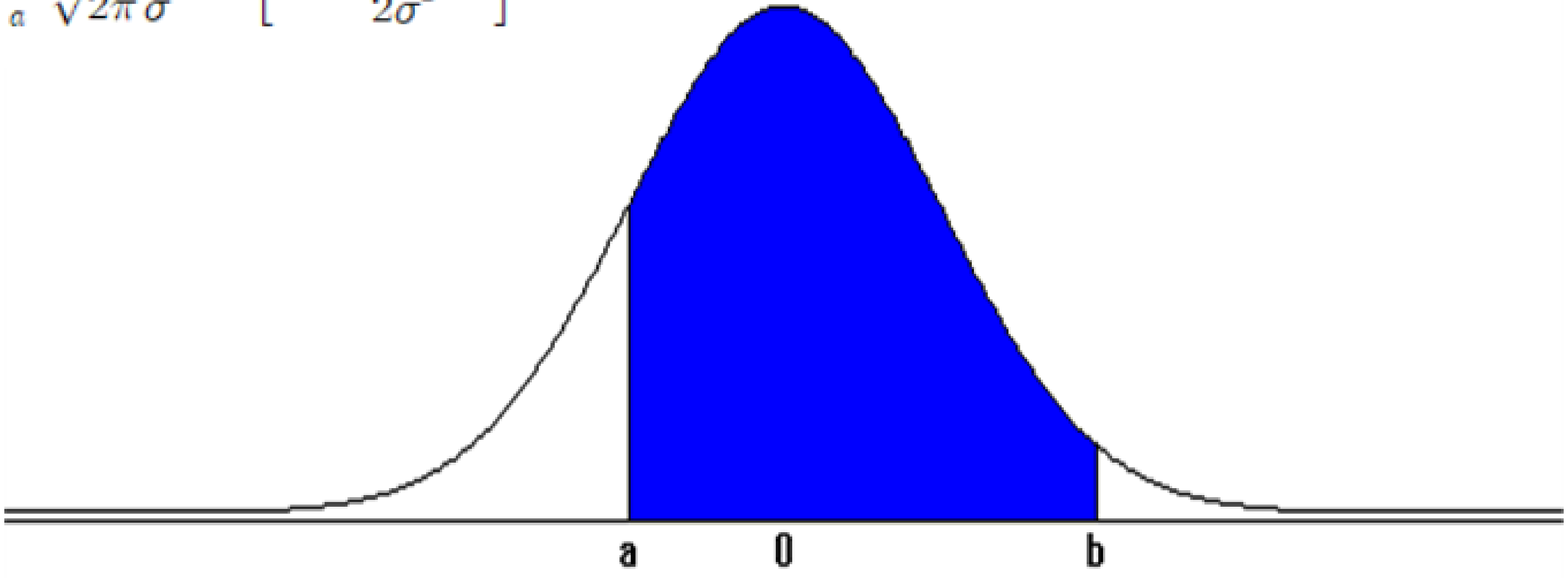
- The characteristics of the normal distribution can be summarized as follows.

1) It is a continuous function in the shape of a bell.
2) It is symmetrical with respect to the mean μ. So the probability of the left and right sides of the mean is 0.5 each.
3) There are an infinite number of normal distributions according to the value of μ and σ.
4) The probability of interval $[\mu - \sigma, \mu + \sigma]$ is 0.68, and the probability of interval $[\mu - 2\sigma, \mu + 2\sigma]$ is 0.95, and the probability of interval $[\mu - 3\sigma, \mu + 3\sigma]$ is 0.997. It implies that the Normal random variable have most of the values around the interval of $\mu \mp 3\sigma$ and there are few values outside of this interval

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Probability Calculation

$$P(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$



<Figure 5.4.8> probability of X on the interval $[a,b]$, $P(a \le X \le b)$

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Standard Normal Distribution

- If X is a normal random variable with mean μ and variance σ^2, the **standardized random variable** Z = (X - μ) / σ is a normal random variable with mean 0 and variance 1, I.e. Z ~ N(0,1). This fact implies that, if we can find the probabilities of all kinds of intervals with N(0,1) distribution, then we can also find the probabilities of all kinds of intervals with N(μ, σ^2). Therefore, N(0,1) is called the standard normal distribution or Z distribution in particular.

☞ If X is a Normal random variable with mean μ and variance σ^2, I.e. X ~ N(μ,σ^2), then the standardized random variable Z,

$$Z = \frac{X-\mu}{\sigma}$$

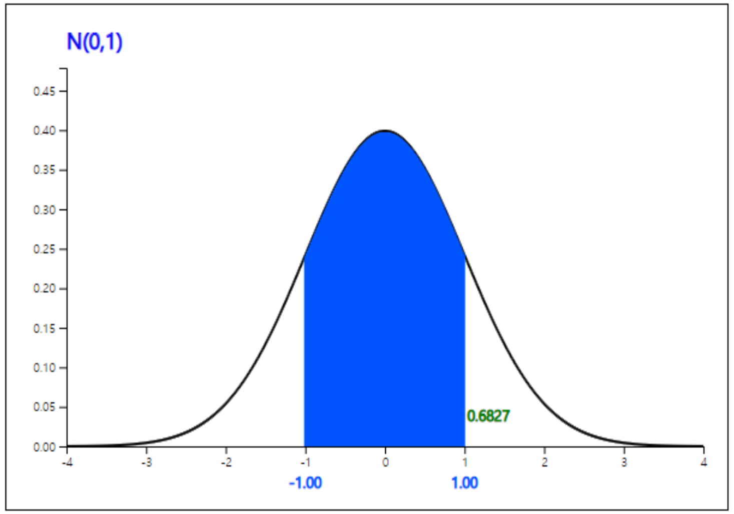
follows a Normal distribution with mean 0 and variance 1, I.e. Z ~ N(0,1).

5.4 Continuous Random Variable

5.4.1 Normal Distribution – 『eStatU』Probability Calculation

Normal Distribution

Mean μ= 0 Std Dev σ= 1 (If number is typed) Execute



P(-1 < Z < 1) = 0.6827

P(Z < 1.6449) = 0.95

P(-1.9600 < Z < 1.9600) = 0.95

Graph Save Reference Site : Wikipedia Wolfram StatTrek KhanAcademy

Normal Distribution Table Percentile Table Table Save

정규분포	μ = 0	σ = 1.000													
x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)
-3.99	0.0000	-2.99	0.0014	-1.99	0.0233	-0.99	0.1611	0.01	0.5040	1.01	0.8438	2.01	0.9778	3.01	0.9987
-3.98	0.0000	-2.98	0.0014	-1.98	0.0239	-0.98	0.1635	0.02	0.5080	1.02	0.8461	2.02	0.9783	3.02	0.9987
-3.97	0.0000	-2.97	0.0015	-1.97	0.0244	-0.97	0.1660	0.03	0.5120	1.03	0.8485	2.03	0.9788	3.03	0.9988
-3.96	0.0000	-2.96	0.0015	-1.96	0.0250	-0.96	0.1685	0.04	0.5160	1.04	0.8508	2.04	0.9793	3.04	0.9988
-3.95	0.0000	-2.95	0.0016	-1.95	0.0256	-0.95	0.1711	0.05	0.5199	1.05	0.8531	2.05	0.9798	3.05	0.9989
-3.94	0.0000	-2.94	0.0016	-1.94	0.0262	-0.94	0.1736	0.06	0.5239	1.06	0.8554	2.06	0.9803	3.06	0.9989
-3.93	0.0000	-2.93	0.0017	-1.93	0.0268	-0.93	0.1762	0.07	0.5279	1.07	0.8577	2.07	0.9808	3.07	0.9989
-3.92	0.0000	-2.92	0.0018	-1.92	0.0274	-0.92	0.1788	0.08	0.5319	1.08	0.8599	2.08	0.9812	3.08	0.9990
-3.91	0.0000	-2.91	0.0018	-1.91	0.0281	-0.91	0.1814	0.09	0.5359	1.09	0.8621	2.09	0.9817	3.09	0.9990
-3.90	0.0000	-2.90	0.0019	-1.90	0.0287	-0.90	0.1841	0.10	0.5398	1.10	0.8643	2.10	0.9821	3.10	0.9990
-3.89	0.0001	-2.89	0.0019	-1.89	0.0294	-0.89	0.1867	0.11	0.5438	1.11	0.8665	2.11	0.9826	3.11	0.9991
-3.88	0.0001	-2.88	0.0020	-1.88	0.0301	-0.88	0.1894	0.12	0.5478	1.12	0.8686	2.12	0.9830	3.12	0.9991
-3.87	0.0001	-2.87	0.0021	-1.87	0.0307	-0.87	0.1922	0.13	0.5517	1.13	0.8708	2.13	0.9834	3.13	0.9991
-3.86	0.0001	-2.86	0.0021	-1.86	0.0314	-0.86	0.1949	0.14	0.5557	1.14	0.8729	2.14	0.9838	3.14	0.9992
-3.85	0.0001	-2.85	0.0022	-1.85	0.0322	-0.85	0.1977	0.15	0.5596	1.15	0.8749	2.15	0.9842	3.15	0.9992
-3.84	0.0001	-2.84	0.0023	-1.84	0.0329	-0.84	0.2005	0.16	0.5636	1.16	0.8770	2.16	0.9846	3.16	0.9992
-3.83	0.0001	-2.83	0.0023	-1.83	0.0336	-0.83	0.2033	0.17	0.5675	1.17	0.8790	2.17	0.9850	3.17	0.9992
-3.82	0.0001	-2.82	0.0024	-1.82	0.0344	-0.82	0.2061	0.18	0.5714	1.18	0.8810	2.18	0.9854	3.18	0.9993
-3.81	0.0001	-2.81	0.0025	-1.81	0.0351	-0.81	0.2090	0.19	0.5753	1.19	0.8830	2.19	0.9857	3.19	0.9993
-3.80	0.0001	-2.80	0.0026	-1.80	0.0359	-0.80	0.2119	0.20	0.5793	1.20	0.8849	2.20	0.9861	3.20	0.9993

5.4 Continuous Random Variable

[Ex 5.4.2] When Z is a standard normal random variable, obtain the following probability using the standard normality distribution table. Then use 『eStatU』.
1) $P(Z < 1.96)$ 2) $P(-1.96 < Z < 1.96)$ 3) $P(Z > 1.96)$

<Answer>
1) By using the standard normal distribution table, $P(Z < 1.96) = 0.975$.
2) $P(-1.96 < Z < 1.96) = P(Z < 1.96) - P(Z < -1.96) = 0.975 - 0.025 = 0.95$
3) $P(Z > 1.96) = 1 - P(Z < 1.96) = 1 - 0.975 = 0.025$
- By using the normal distribution module of 『eStatU』,
1) enters the interval –4, 1.96 in the first of the options below the graph, then clicks the [Execute] button.

P(-4.00 < Z < 1.96) = 0.9750

2) The answer is calculated by entering interval of 1.96 and 1.96,
3) is calculated by entering interval of 1.96 and 4

5.4 Continuous Random Variable

[Example 5.4.3] When Z is a standard random variable, obtain x that satisfies the following formula. Then use 『eStatU』 to find this value x.

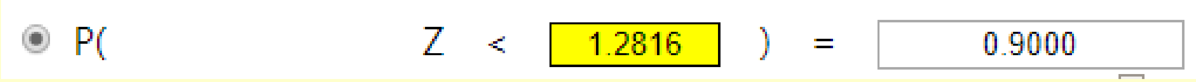
1) $P(Z < x) = 0.90$ 2) $P(-x < Z < x) = 0.99$ 3) $P(Z > x) = 0.05$

<Answer>

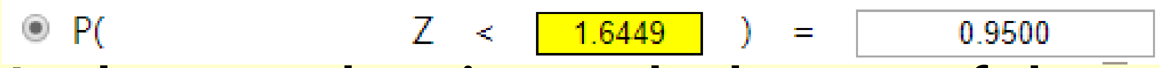
1) By using the standard normal distribution, the value of x is 1.2826
2) By using the standard normal distribution, the percentile of 0.995 is 2.575. .
3) By using the standard normal distribution, the value of x is 1.645.

- by using 『eStatU』,

1) Enter p = 0.90 in the right box in the second option at the bottom of the graph screen, then clicks the [Execute] button.



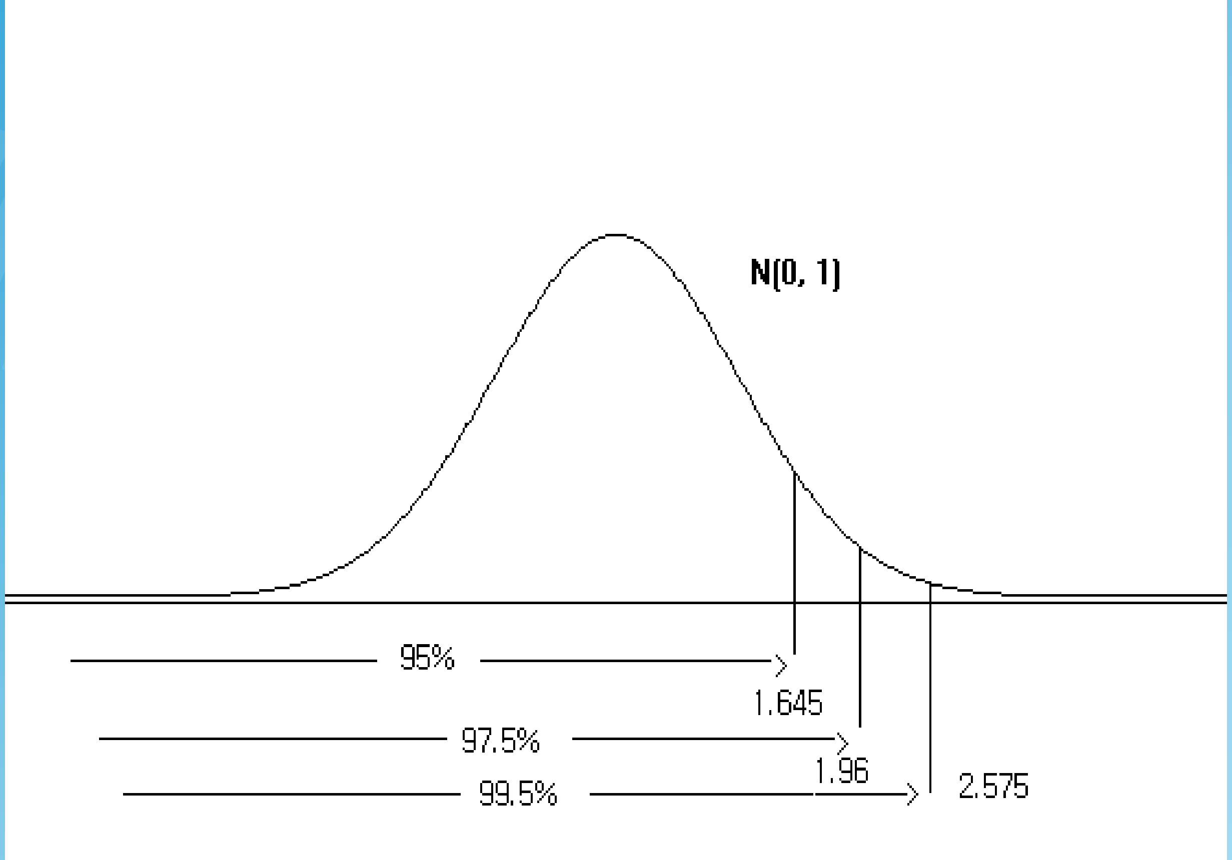
2) Enter p = 0.99 in the right box in the third option and click the [Execute] button.

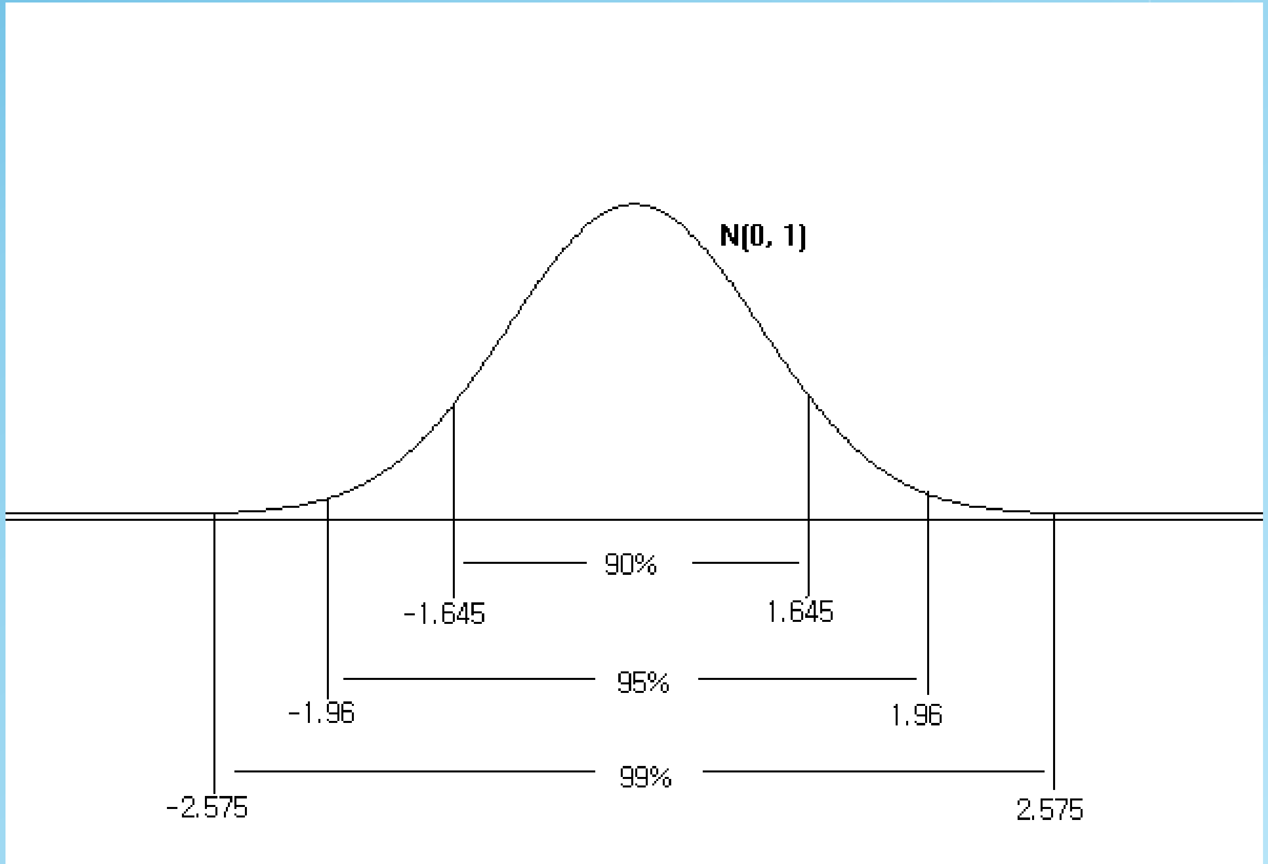


3) In the second option at the bottom of the graph screen, type p = 0.95 in the right box and click the [Execute] button.

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Frequently used percentile





5.4 Continuous Random Variable

5.4.1 Normal Distribution – Probability Calculation

☞ When X is a normal random variable with a mean μ and variance σ^2, $(X-\mu)/\sigma$ follows the standard normal distribution. Therefore, the probability P(a< X <b) of the interval $[a,b]$ of X is as follows.

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

5.4 Continuous Random Variable

[Example 5.4.4] If the mid-term scores (X) of the Statistics course follows a normal distribution with an average of 70 points and a standard deviation of 10 test results X, calculate the following probabilities. Check the calculated value by using 「eStatU.

1) P(X < 94.3)　　2) P(X > 57.7)　　3) P(57.7 < X < 94.3)

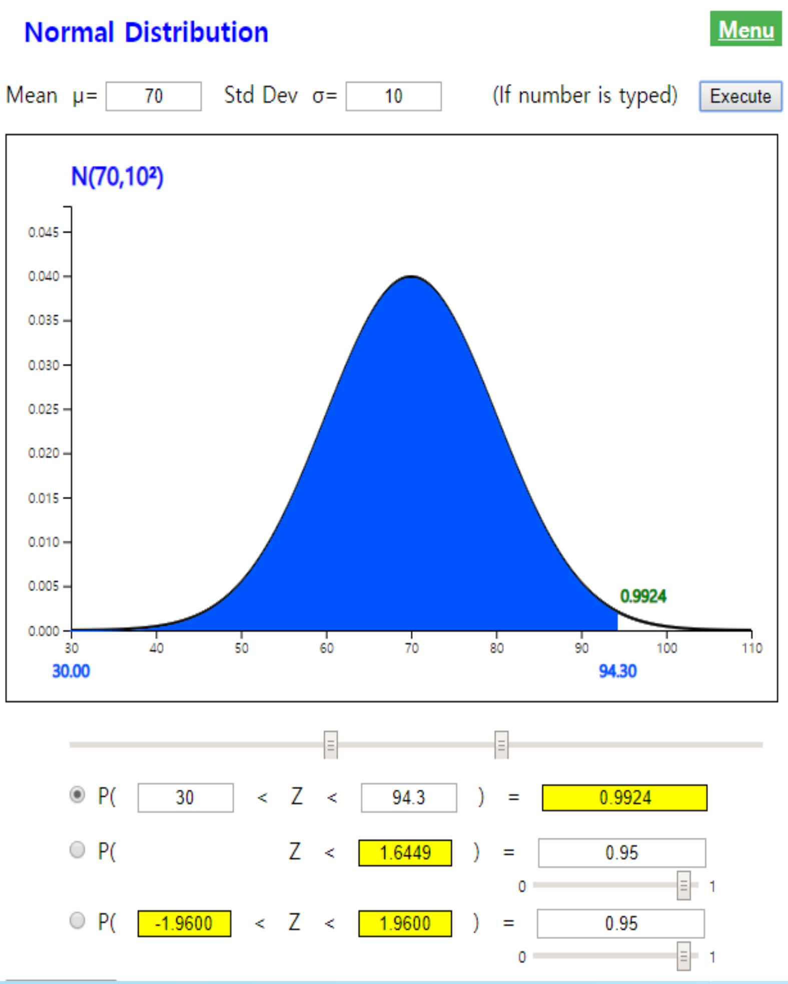
<Answer>

1) $P(X < 94.3) = P\left(\frac{X-70}{10} < \frac{94.3-70}{10}\right) = P(Z < 2.43) = 0.9925$

2) $P(X > 57.7) = P\left(\frac{X-70}{10} > \frac{57.7-70}{10}\right) = P(Z > -1.23) = 0.8907$

3) $P(57.7 < X < 94.3) = P\left(\frac{57.7-70}{10} < \frac{X-70}{10} < \frac{94.3-70}{10}\right)$

$= P(-1.23 < Z < 2.43) = 0.8832$



5.4 Continuous Random Variable

[Example 5.4.5] In [Example 5.4.4], obtain the following percentiles by using 「eStatU」.

1) What is the 95% percentile of the mid-term test scores?

2) What is the 95% percentile of two sided of the mid-term scores?

<Answer>

1) The 95 percentile $P(Z < ?) = 0.95$ in N(0,1) is 1.645, so the percentile in the $N(70, 10^2)$ is 70 + 1.645 × 10 = 86.45.

2) The 95 percentile of two-sided type $P(? < Z < ?) = 0.95$ in N(0,1), which can be calculated from $P(Z < ?) = 0.975$, is 1.96. So the two-sided 95% percentile interval is [70 − 1.96 × 10, 70 + 1.96 × 10], I.e., [50.4, 89.6].

- To obtain a percentile for a general normal distribution using 「eStatU, enter the mean as 70 and the standard deviation as 10 on the screen in <Fig. 5.4.13>. 1) enters 0.95 in the second right box of options under the graph screen and press the [Execute] button to display the 95% percentile 86.4485.

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Binomial probability calculation approximately

- In the case of large n in a binomial distribution, a direct probability calculation is not possible. In such cases, a normal distribution with an average of np and a variance of np (1-p) is used to calculate an approximated probability.

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Binomial probability calculation approximately

[Example 5.4.6] The defect rate of products produced in a factory is 5 per cent. One day, a sample of 100 products is collected
1) What is the probability that there are less than two defective products?
2) What is the probability that there are defectives between 3 and 7?

<Answer>
- X is a binomial distribution of n = 100, p = 0.05. The mean is np = 100 × 0.05 = 5, and the variance is np(1-p) = 100 × 0.05 x (1-0.05) = 4.75. The probability calculation using N(5, 4.75) is as follows.

1) $P(X \le 2) = P\left(Z \le \frac{(2-5)}{\sqrt{4.75}}\right) = P(Z \le -1.376) = 0.0845$

2) $P(3 \le X \le 7) = P\left(\frac{(3-5)}{\sqrt{4.75}} \le Z \le \frac{(7-5)}{\sqrt{4.75}}\right)$

$= P(-0.918 \le Z \le 0.918) = 0.642$

5.4 Continuous Random Variable

5.4.2 Exponential Distribution

- Most of the continuous data obtained in real life follows normal distribution, but sometimes it is not. Let's take a look at the following examples.
  - Investigate the time interval of coming calls between 9 a.m. and 10 a.m. in an office.
  - Investigate the time interval between defective products appearing in the factory production line.

- These examples are the data that appears when events occur at the same rate at a given time (e.g., three calls per hour, etc.). If the average number of events per unit hour is λ and X is the random variable of the time between events, then X can be considered as the exponential distribution. λ is the parameter of the exponential distribution

5.4 Continuous Random Variable

5.4.2 Exponential Distribution

☞ Exponential Distribution

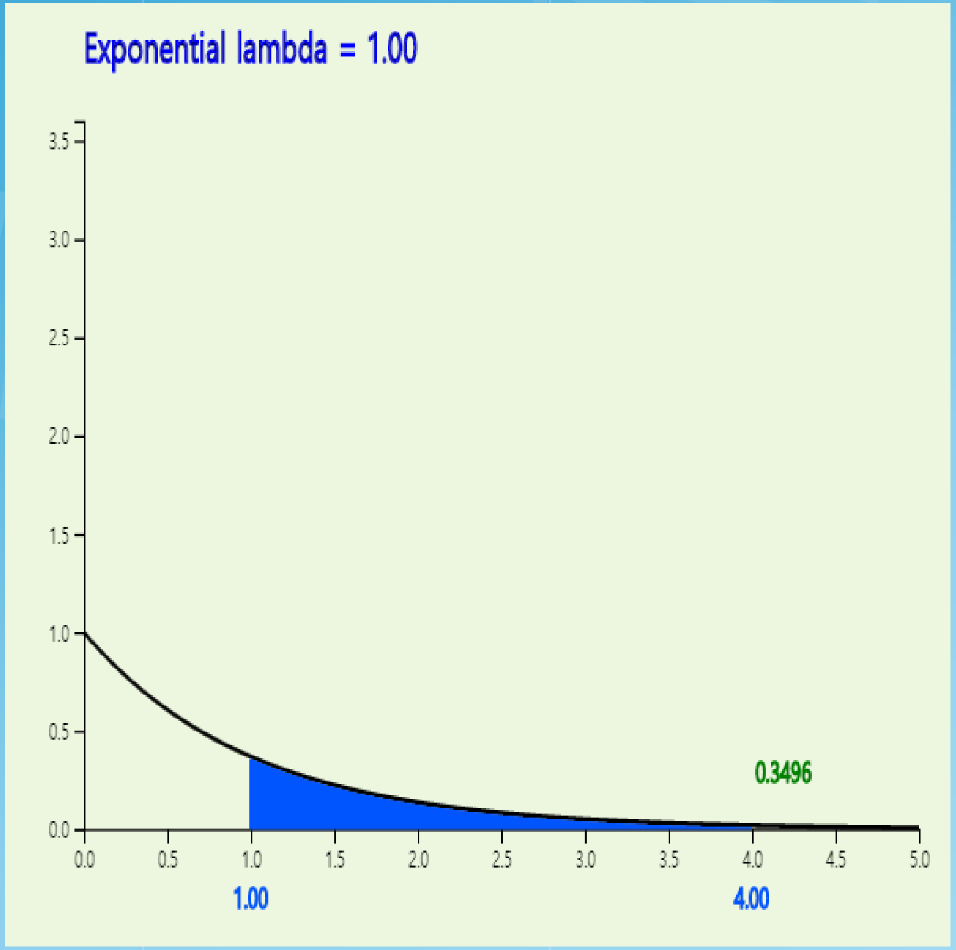
When the average number of events per unit hour is λ anx the random variable X is the time between events, the probability distribution function of X is as follows.

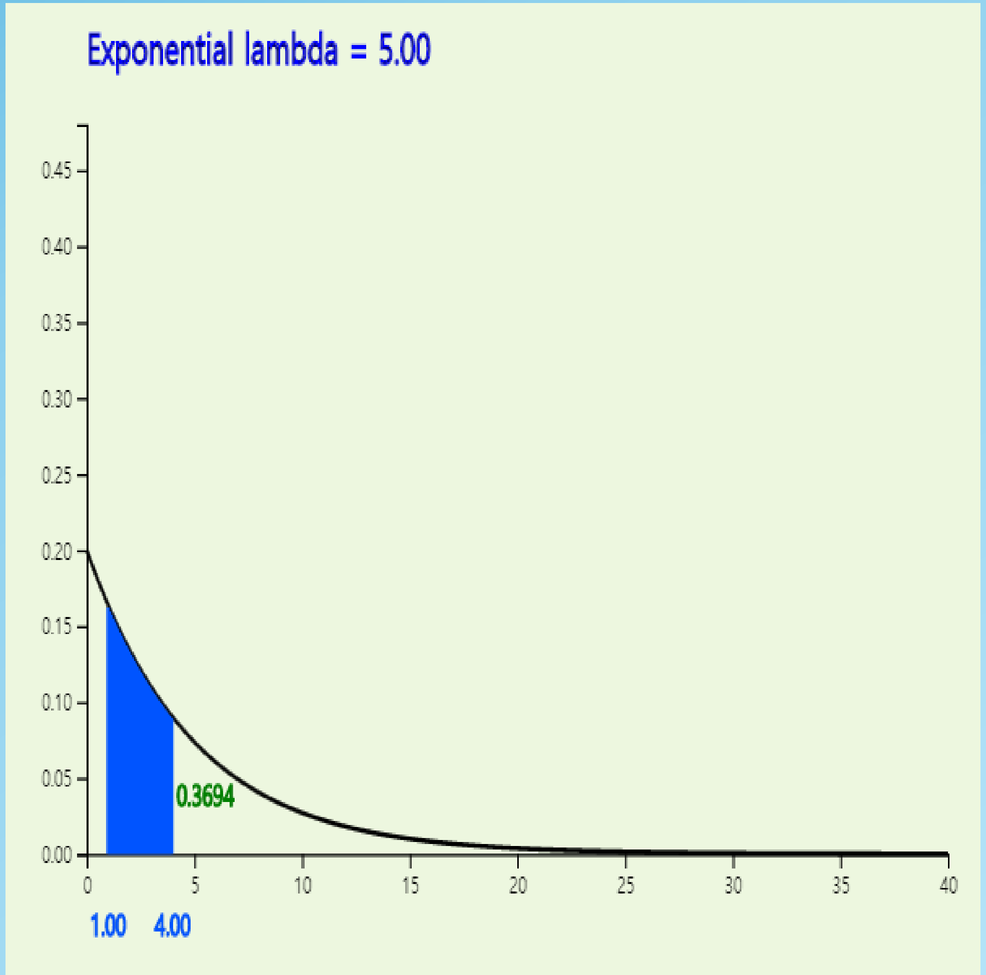
$$f(x) = \lambda \exp(-\lambda x), \qquad x = 1,2, \ldots$$

It is called the exponential distribution and its expectation and variance are $E(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$.

5.4 Continuous Random Variable

5.4.2 Exponential Distribution





5.4 Continuous Random Variable

[Ex 5.4.7] If the average life span of a product is 10 hours and follows the exponential distribution, obtain the following probabilities using 『eStatU』.
1) What is the probability of a product having a lifespan of less than 5 hours?
2) What is the probability of a product having a lifespan more than 10 hours?
<Answer>

- **Select = 10 in 'Exponential Distribution' of 『eStatU』**

1) Enter 0 and 5 at the boxes on the 1st line below the distribution graph as follows and click [Execute] button.

P(0.00 < X < 5) = 0.3297

2) Similarly, enter 10 and a large number 50 at the boxes on the 1st line below the distribution graph as follows and click [Execute] button.

P(10 < X < 50) = 0.3611

5.5 Summary

- Definition of probability
  - classical definition
  - relative frequency distribution
- Probability calculation
  - Additive rule and mutually exclusive event
  - Multiplicative rule and independent event
- Discrete Random Variable
  - Binomial, Poisson, Geometric, HyperGeometric
- Continuous Random Variable
  - Normal, Exponential

Thank you