Summary---Topic 3: Discrete & Continuous Probability Distributions

Random Variable

- The random variable can take different values and the value is subject to chance (probability)
- Can be subdivided into discrete random variables and continuous random variables

✓ **Discrete random variable**: has **countable** number of values

✓ **Continuous random variable**: has **uncountable** number of values

Expected Value of Discrete Random Variable

- ❖ The value of a random variable that would be "expected" to obtain
- ❖ Calculated by the average of the random variable values which resulted from **infinite** experiments
- ❖ For a discrete random variable X, its expected value is equal to the weighted average of all possible values of X, where the weight is the corresponding probability of the possible values of X

$$\mu = E(X) = \sum_{i=1}^{N} X_i P(X_i)$$

- ❖ The expected value only tells you the value that you can obtain **on average**, but does not guarantee you will get that value in the next experiment
- ❖ The expected value may not be a possible value of the random variable

Variance and Standard Deviation of Discrete Random Variable

- Variance of random variable X: The expected value of the squared deviation of X from its expected value (mean)

- For a discrete random variable **X**, its variance is equal to

$$\sigma^2 = \sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i)$$

- Standard deviation of random variable X: Square root of the variance of random variable X

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i)}$$

Binomial Distribution

- Consider **"n" independent identical trials** with **2 possible mutually exclusive outcomes** (i.e. "success" and "failure") and **constant probability of success "π" in each trial**. If X represents the number of successes that occur in n trials, then X is a binomial random variable.

- A discrete random variable is said to be a binomial random variable if its probability is given by

$$\mathrm{P}(X = x) = \frac{n!}{x!\,(n-x)!}\pi^{x}(1-\pi)^{(n-x)}, \qquad for\ x = 0,1,2,\dots,n-1,n,$$

where

✓ π: Probability of "success" at each trial
✓ n: Number of independent identical trials
✓ x: Number of "successes" in n trials (possible value of x = 0,1,…,n)
✓ $\mathrm{P}(X = x)$: Probability of getting x times of successes out of n trials
✓ $\frac{n!}{x!(n-x)!} = {}_{n}C_{x}$: Number of combinations of x times of successes out of n trials
✓ π^{x}: Total probability of getting x times of successes
✓ $(1-\pi)^{(n-x)}$: Total probability of getting (n - x) times of failures

Binomial Distribution

- A binomial distribution depends upon two parameters, “n” and “π”
- X ~ B(n, π)

- Expected value of binomial random variable:

$$\mu = \mathrm{E}[X] = n\pi$$

- Variance of binomial random variable:

$$\sigma^2 = n\pi(1-\pi)$$

- Standard deviation of binomial random variable:

$$\sigma = \sqrt{n\pi(1-\pi)}$$

Exercises and Solutions

Q1. In a recent survey concerning the age (to the nearest year) and weight (to the nearest 10 lb) of first-year university students, the following probability distribution was obtained:

Age	**Weight**				
	100	**110**	**120**	**130**	**140**
19	0.02	0.09	0.09	0.01	0.02
20	0.06	0.15	α	0.05	0.03
21	0.02	0.06	0.11	0.04	0.05

a) Find the value of α

b) Construct the probability distribution of the weights of these students.

c) A student is selected at random. What do you expect his/her weight to be?

d) What is the standard deviation of the distribution in (b)?

e) Are "age" and "weight" independent? Why or why not?

Solution:

a) $\alpha = 1 - 0.02 - 0.06 - 0.02 - 0.09 - 0.15 - 0.06 - 0.09 - 0.11 - 0.01 - 0.05 - 0.04 - 0.02 - 0.03 - 0.05 = 0.2$

Q1.

Age	Weight				
	100	**110**	**120**	**130**	**140**
19	0.02	0.09	0.09	0.01	0.02
20	0.06	0.15	α=0.2	0.05	0.03
21	0.02	0.06	0.11	0.04	0.05

b) Construct the probability distribution of the weights of these students.

c) A student is selected at random. What do you expect his/her weight to be?

d) What is the standard deviation of the distribution in (b)?

e) Are “age” and “weight” independent? Why or why not?

Solution:

b) Probability distribution of weight

Weight X_i	**100**	**110**	**120**	**130**	**140**
Probability $P(X_i)$	0.1	0.3	0.4	0.1	0.1

c) Expected Weight: $\mathrm{E[X]} = \sum_{i=1}^{n} X_i \mathrm{P}(X_i) =$

$$= 100 \times 0.1 + 110 \times 0.3 + 120 \times 0.4 + 130 \times 0.1 + 140 \times 0.1 = 118lb$$

d) standard deviation : $\sigma = \sqrt{\sum_{i=1}^{n} (X_i - \mathrm{E}[X])^2 \mathrm{P}(X_i)} =$

$$\sqrt{(100-118)^2 \times 0.1 + \cdots + (140-118)^2 \times 0.1} = 10.77lb$$

Q1.

Age	Weight				
	100	**110**	**120**	**130**	**140**
19	0.02	0.09	0.09	0.01	0.02
20	0.06	0.15	α=0.2	0.05	0.03
21	0.02	0.06	0.11	0.04	0.05

e) Are "age" and "weight" independent? Why or why not?

Solution:

e) Take Age=21 and Weight=110 lb as example,

$$\mathrm{P}(Age = 21 \ and \ Weight = 110) = 0.06$$

$$\mathrm{P}(Age = 21\ ) = 0.02 + 0.06 + 0.11 + 0.04 + 0.05 = 0.28$$
$$\mathrm{P}(Weight = 110\ ) = 0.3$$

$$\mathrm{P}(Age = 21\ ) \times \mathrm{P}(Weight = 110) = 0.3 \times 0.28 = 0.084 \neq 0.06$$

Hence, age and weight are not independent.

Q2. An airline wants to overbook flights in order to reduce the numbers of vacant seats. For a certain flight, it is known that the probabilities of 0, 1, 2 and 3 vacant seats are 0.70, 0.15, 0.10 and 0.05 respectively.

a) Find the mean and standard deviation for the number of vacant seats.

b) What is the expected total number of vacant seats on 100 such flights?

Solution:

a) Mean:

$$\mathrm{E[X]} = \sum_{i=1}^{n} X_i \mathrm{P}(X_i) = 0 \times 0.7 + 1 \times 0.15 + 2 \times 0.1 + 3 \times 0.05 = 0.5$$

Standard Deviation :

$$\sigma = \sqrt{\sum_{i=1}^{n} (X_i - E[X])^2 \mathrm{P}(X_i)}$$
$$= \sqrt{(0-0.5)^2 \times 0.7 + (1-0.5)^2 \times 0.15 + (2-0.5)^2 \times 0.1 + (3-0.5)^2 \times 0.05}$$
$$= 0.8660$$

b) Expected total number $= \mathrm{E}[100\mathrm{X}] = 100 \times \mathrm{E}[X] = 100 \times 0.5 = 50.$

Q3*. Given the following probability distributions:

Distribution A		Distribution B	
X	**P(X)**	**X**	**P(X)**
0	0.50	0	0.05
1	0.20	1	0.10
2	0.15	2	0.15
3	0.10	3	0.20
4	0.05	4	0.50

a) Compute the expected value for each distribution.

b) Compute the standard deviation for each distribution.

c) Compare the results of distributions A and B.

Solution:

a) $\mu_A = \mathrm{E}[X] = \sum_{i=1}^{n} X_i \mathrm{P}(X_i) = 0 \times 0.5 + 1 \times 0.2 + 2 \times 0.15 + 3 \times 0.1 + 4 \times 0.05 = 1$

$\mu_B = \mathrm{E}[X] = \sum_{i=1}^{n} X_i \mathrm{P}(X_i) = 0 \times 0.05 + 1 \times 0.1 + 2 \times 0.15 + 3 \times 0.2 + 4 \times 0.5 = 3$

Q3*.

Distribution A		Distribution B	
X	**P(X)**	**X**	**P(X)**
0	0.50	0	0.05
1	0.20	1	0.10
2	0.15	2	0.15
3	0.10	3	0.20
4	0.05	4	0.50

b) Compute the standard deviation for each distribution.

c) Compare the results of distributions A and B.

Solution:

b) $\sigma_A = \sqrt{\sum_{i=1}^{n}(X_i - E[X])^2 \mathrm{P}(X_i)} = \sqrt{(0-1)^2 \times 0.5 + (1-1)^2 \times 0.2 + (2-1)^2 \times 0.15 + (3-1)^2 \times 0.1 + (4-1)^2 \times 0.05} = 1.2247$

$\sigma_B = \sqrt{\sum_{i=1}^{n}(X_i - E[X])^2 \mathrm{P}(X_i)}$
$= \sqrt{(0-3)^2 \times 0.05 + (1-3)^2 \times 0.1 + (2-3)^2 \times 0.15 + (3-3)^2 \times 0.2 + (4-3)^2 \times 0.5}$
$= 1.2247$

c) Distribution A and B has the same spread but locate at different position. Distribution A is on the left-hand-side of Distribution B.

Q4*. You are trying to develop a strategy for investing in two different stocks. The anticipated annual return for a $1,000 investment in each stock has the following probability distribution:

Returns		
Stock X	**Stock Y**	**Probability**
-$50	-$100	0.1
20	50	0.3
100	130	0.4
150	200	0.2

a) For each stock, compute the expected return and the standard deviation of return.

b) Do you think that you will invest in stock X or stock Y? Explain.

Solution:

a) $\mu_{StockX} = \sum_{i=1}^{n} X_i \mathrm{P}(X_i) = -50 \times 0.1 + 20 \times 0.3 + 100 \times 0.4 + 150 \times 0.2 = 71,$

$\mu_{StockY} = \sum_{i=1}^{n} Y_i \mathrm{P}(Y_i) = -100 \times 0.1 + 50 \times 0.3 + 130 \times 0.4 + 200 \times 0.2 = 97,$

$\sigma_{StockX} = \sqrt{(-50-71)^2 \times 0.1 + \cdots + (150-71)^2 \times 0.2} = 61.88$

$\sigma_{StockY} = \sqrt{(-100-97)^2 \times 0.1 + \cdots + (150-97)^2 \times 0.2} = 84.27$

b) Stock Y gives investor higher expected return than stock X., but also a higher standard deviation. Thus, a risk-averse investor should invest in stock X, while investor who is willing to take a higher risk can expect a higher return from stock Y.

Q5. When a customer places an order with Rudy's On-Line Office Supplies, a computerized accounting information system (AIS) automatically checks to see if the customer has exceeded his or her credit limit. Past records indicate that the probability of customers exceeding their credit limit is 0.05. Suppose that, on a given day, 20 customers place orders. Assume that the number of customers that the AIS detects as having exceeded their credit limit is distributed as a binomial random variable.

a) What are the mean and standard deviation of the number of customers exceeding their credit limits?

b) What is the probability that 0 customers will exceed their limits?

c) What is the probability that 1 customer will exceed his or her limit?

d) What is the probability that 2 or more customers will exceed their limits?

Solution:

a) X = no. of customers that the AIS detects as having exceeded their credit limit

π = success probability = 0.05

n = 20

Since X is binomial distribution X ~B(n=20, π =0.05),

$$mean = n \times \pi = 20 \times 0.05 = 1$$

$$\text{std} = \sqrt{n \times \pi \times (1-\pi)} = \sqrt{20 \times 0.05 \times 0.95} = 0.9747.$$

Q5.

- X = no. of customers that the AIS detects as having exceeded their credit limit,
- X is binomial distribution X ~B(n=20, π =0.05).

b) What is the probability that 0 customers will exceed their limits?

c) What is the probability that 1 customer will exceed his or her limit?

d) What is the probability that 2 or more customers will exceed their limits?

For a discrete random variable X~B(n, π), its probability is given by

$$\mathrm{P}(X = x) = \frac{n!}{x!\,(n-x)!}\pi^{x}(1-\pi)^{(n-x)}, \qquad for\ x = 0,1,2,\dots,n-1,n$$

Solution:

b) $\mathrm{P(X = 0)} = \frac{20!}{0!(20-0)!} \times 0.05^{0} \times 0.95^{20} = 0.3585$

c) $\mathrm{P(X = 1)} = \frac{20!}{1!(20-1)!} \times 0.05^{1} \times 0.95^{19} = 0.3774$

d) $\mathrm{P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)} = 1 - 0.3585 - 0.3774 = 0.2642$

Q6*. For e-commerce merchants, getting a customer to visit a Web site isn't enough. Merchants must also persuade online shoppers to spend money by completing a purchase. Experts at Consumer Consulting estimate that 88% of Web shoppers abandon their virtual shopping carts before completing their transaction. Consider a sample of 20 customers who visit an e-commerce Web site, and assume that the probability that a customer will leave the site before completing the transaction is 0.88. What is the probability that all 20 of the customers will leave the site without completing a transaction?

Solution:

- Let X be the number of customers who will leave the site without completing a transaction.
- $X \sim B(n=20, \pi=0.88)$

$$\mathrm{P}(X = 20) = \frac{20!}{20! * (20-20)!} \times 0.88^{20} \times (1 - 0.88)^{0} = 0.0776.$$

Q7. A task force of CityU sampled 200 students after the mid-term test to ask them whether they went shopping the weekend before the mid-term test or spent the weekend studying, and whether they did well or poorly on the mid-term test. The following result was obtained.

	Did Well on Mid-Term Test	**Did Poorly on Mid-Term Test**
Studied for Mid-Term Test	90	10
Went Shopping	30	70

a) What is the probability that a randomly selected student did well on the mid-term test <u>or</u> went shopping the weekend before the mid-term test?

*Addition rule for compound events A or B (A ∪ B) :

$$\mathrm{P}(\mathrm{A}\ or\ B) = P(A) + P(B) - P(A\ and\ B)$$

Solution:

a) $\mathrm{P}(did\ well\ or\ went\ shopping)$

$$= \mathrm{P}(did\ well) + \mathrm{P}(went\ shopping) - \mathrm{P}(did\ well\ and\ went\ shopping)$$

$$= \frac{90+30}{200} + \frac{30+70}{200} - \frac{30}{200} = \frac{190}{200} = 0.95.$$

Q7. total sample size=200.

	Did Well on Mid-Term Test	Did Poorly on Mid-Term Test
Studied for Mid-Term Test	90	10
Went Shopping	30	70

b) A random sample of 10 students is selected. What is the probability that 2 of them did well on mid-term test and studied for mid-term test the weekend before the mid-term test? What distribution are you using? Why can you use such distribution?

Solution:

1. no. of trials is fixed
2. independent trials
3. two mutually exclusive outcomes
4. probability of success is constant

- Therefore, we can use Binomial distribution.
- Let X be the no. of students did well on mid-term test and studied for mid-term test the weekend before the mid-term test out of the selected 10 students

$$X \sim B(n = 10, \pi = 90/200 = 0.45)$$

$$P(X = 2) = \frac{10!}{2!(10-2)!} \times 0.45^2 \times (1 - 0.45)^{10-2} = 0.0763.$$

Q8. Suppose that 1,000 patrons of a restaurant were asked whether they preferred beer or wine. 70% said that they preferred beer. 60% of patrons were male. 80% of the males preferred beer.

a) What is the probability a randomly selected patron prefers wine?

b) What is the probability a randomly selected patron is female and prefers wine?

c) Suppose a randomly selected patron prefers wine, what is the probability that the patron is a male?

d) Suppose 5 patrons were selected, what is the probability that at least four of them prefer beer?

Solution:

$\mathrm{P(beer)} = 0.7;\ \ \mathrm{P}(male) = 0.6;\ \ \mathrm{P}(beer|male) = 0.8$

a) $\mathrm{P}(wine) = 1 - \mathrm{P}(beer) = 1 - 0.7 = 0.3$

b) $\mathrm{P}(female\ and\ wine)$

$= \mathrm{P}(wine) - \mathrm{P}(wine\ and\ male) = 0.3 - \mathrm{P}(wine\ and\ male).$

$= 0.3 - [\mathrm{P}(male) - \mathrm{P}(beer\ and\ male)] = 0.3 - [0.6 - P(beer\ and\ male)]$

$= 0.3 - [0.6 - \mathrm{P}(beer|male) \times \mathrm{P}(male)]$

$= 0.3 - [0.6 - 0.8 \times 0.6] = 0.18.$

Q8. Suppose that 1,000 patrons of a restaurant were asked whether they preferred beer or wine. 70% said that they preferred beer. 60% of patrons were male. 80% of the males preferred beer.

c) Suppose a randomly selected patron prefers wine, what is the probability that the patron is a male?

d) Suppose 5 patrons were selected, what is the probability that at least four of them prefer beer?

Solution:

$$\mathrm{P}(\text{beer}) = 0.7; \;\; \mathrm{P}(male) = 0.6; \;\; \mathrm{P}(beer|male) = 0.8$$

c) $\mathrm{P}(male|wine) = \frac{\mathrm{P}(male\ and\ wine)}{\mathrm{P}(wine)} = \frac{0.12}{1-0.7} = 0.4$

d) Define X be the number of patrons prefer beer in the 5 selected patrons, then X~B(n=5,π=0.7)

$$\mathrm{P}(at\ least\ 4\ patrons) = \mathrm{P}(X=4) + \mathrm{P}(X=5)$$
$$= \frac{5!}{4! \times (5-4)!} \times 0.7^4 \times (1-0.7)^{5-4} + \frac{5!}{5! \times (5-5)!} \times 0.7^5 \times (1-0.7)^{5-5}$$
$$= 0.36015 + 0.16807 = 0.52822.$$

Q9*. MTR Corporation has to conduct surveys regularly to evaluate its service quality. According to previous studies, 87% of the passengers refuse to take part in such surveys.

- If 15 passengers are selected randomly, what is the probability that at least 2 of them will respond to the survey?

Solution:

- X is the number of passengers responded to the survey
- π is the population proportion of passenger responded to the survey
- X~B(n=15, π=1-0.87=0.13)

$$\mathrm{P}(at\ least\ 2\ responors)$$
$$= \mathrm{P}(X \geq 2) = 1 - \mathrm{P}(X = 0) - \mathrm{P}(X = 1)$$
$$= 1 - \frac{15!}{0! \times (15-0)!} \times 0.13^0 \times (1 - 0.13)^{15-0} - \frac{15!}{1! \times (15-1)!} \times 0.13^1$$
$$\times (1 - 0.13)^{15-1}$$
$$= 1 - 0.1238 - 0.2275 = 0.5987$$

Q10. According to Dental Association, 60% of all dentists use nitrous oxide ("laughing gas") in their practice. Let x be the number of dentists who use laughing gas in practice in a random sample of five dentists. The probability distribution of x is as follows:

x	0	1	2	3	4	5
P(x)	0.0102	0.0768	0.2304	0.3456	0.2592	0.0778

a) Find the probability that less than 2 dentists use laughing gas in a sample of five.

b) Find E(x). Interpret the result.

c) Find the standard deviation of x.

d) Based on the results of (b) and (c), show that the distribution of x is binomial with n = 5 and π = 0.6.

Solution:

a) $\mathrm{P}(X < 2) = \mathrm{P}(X = 0) + \mathrm{P}(X = 1) = 0.0102 + 0.0768 = 0.087$

b) $\mathrm{E}[X] = \sum_{i=1}^{n} X_i \mathrm{P}(X_i) = 0 \times 0.0102 + 1 \times 0.0768 + \cdots + 5 \times 0.0778 = 3.0002$

It means, on average, among 5 dentists 3 of them will use "laughing gas".

c) $std\colon \sigma = \sqrt{\sum_{i=1}^{n} (X_i - E[X])^2 \mathrm{P}(X_i)} =$

$\sqrt{(0 - 3.0002)^2 \times 0.0102 + \cdots + (5 - 3.0002)^2 \times 0.0778} = 1.0954$

Q10. According to Dental Association, 60% of all dentists use nitrous oxide ("laughing gas") in their practice. Let x be the number of dentists who use laughing gas in practice in a random sample of five dentists. The probability distribution of x is as follows:

x	0	1	2	3	4	5
P(x)	0.0102	0.0768	0.2304	0.3456	0.2592	0.0778

a) Find the probability that less than 2 dentists use laughing gas in a sample of five.

b) Find E(x). Interpret the result. $\mathrm{E}[X] = 3.0002$

c) Find the standard deviation of x. $\sigma = 1.0954$

d) Based on the results of (b) and (c), show that the distribution of x is binomial with n = 5 and π = 0.6.

Solution:

d) Define "success" = use laughing gas, "failure" = not use laughing gas,
Let X represents the number of "success" in n independent trials and the probability of success in each trial is π.

According to results of b) and c), we have

$$\begin{cases} \mathrm{E}[X] = n\pi = 3.0002 \\ \sigma^2 = n\pi(1-\pi) = 1.0954^2 \end{cases}$$

Solve the equations, we have n = 5 and π = 0.6.