FUNDAMENTALS OF STATISTICS (4ECON013C)

TUTORIAL 3

Discrete Probability Distributions

Learning outcomes:

- understand the concept of discrete random variable.
- be able to find expected value and variance of discrete distributions.
- learn about probability mass function of Binomial distribution.
- understand Geometric distribution.

Discrete random variable and its probability distribution

A variable is said to be **discrete** and **random** if it can take only certain values that occur by chance.

Table 3.1 Examples of discrete random variable

Experiment	***Random variable***	***Sample space***
Coin is tossed 4 times	Number of heads	{0, 1, 2, 3, 4}
Ten people are sampled	Number of brown eyes	{0, 1, 2, …, 9, 10}
Make 50 calls to sell a product	Number of purchases	{0, 1, 2, …, 49, 50}

The **probability distribution of a discrete random variable** is graph, table, or formula that specifies the probability associated with each possible value that the random variable can assume. The following requirements are set for the probability distribution of a discrete random variable:

1) $p(x) \geq 0$ for all values of x.
2) $\sum p(x) = 1$ where the summation of p(x) is over all possible values of x.

Example 3.1. A fair square spinner with sides labelled 1, 2, 3 and 4 is spun twice. The two scores obtained are added together to give the total, X. Draw up the probability distribution table for X.

Solution. Let's create the joint table of the outcomes from two spins:

1st spin \ 2nd spin	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

X can take the following values (outcomes): {2, 3, 4, 5, 6, 7, 8}. To create the probability distribution table, we need to count the occurrence of each outcome and then divide it by the total number of trials, 16.

Based on the table above, we can create the probability distribution table for X:

x	2	3	4	5	6	7	8
P(X = x)	1/16	2/16	3/16	4/16	3/16	2/16	1/16

Expected value and variance of a discrete random variable are given below:

The **variance** of a random variable x is

$$\sigma^2 = E[(x-\mu)^2] = \Sigma(x-\mu)^2 p(x) = \Sigma x^2 p(x) - \mu^2$$

The **mean, or expected value,** of a discrete random variable x is

$$\mu = E(x) = \Sigma x p(x)$$

Example 3.2. A portfolio manager calculated the following chances of returns for his portfolio: 30% chance of \$400, 25% chance of \$600, 15% chance of \$1000 and 30% chance of loss of \$200 (i.e. -\$200). Compute the E(x), Var(x) and SD(x), where x is the return of the portfolio.

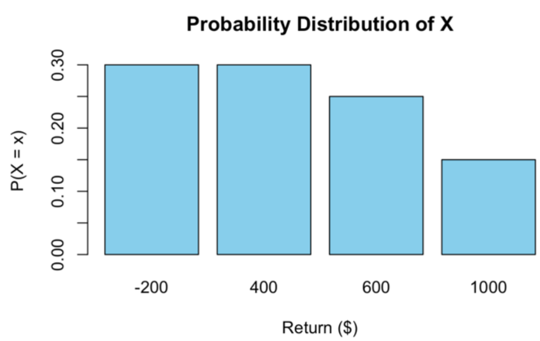
Solution. The probability distribution is given below:

x	-200	400	600	1000
P(X = x)	0.3	0.3	0.25	0.15

E(x) = -200*0.3 + 400*0.3 + 600*0.25 + 1000*0.15 = \$360

Var(x) $= \sigma^2 = (-200-360)^2 * 0.3 + (400-360)^2 * 0.3 + (600-360)^2 * 0.25 + (1000-360)^2 * 0.15 = 170{,}400$

SD(x) $= \sigma = \sqrt{170{,}400}$ = \$412.8



Binomial distribution

A **binomial distribution** is an example for discrete distribution and can be used to model the number of successes in a fixed number of independent trials.

Characteristics of a Binomial Random Variable

1. The experiment consists of n identical trials.
2. There are only two possible outcomes on each trial. We will denote one outcome by S (for Success) and the other by F (for Failure).
3. The probability of S remains the same from trial to trial. This probability is denoted by p, and the probability of F is denoted by $q = 1 - p$.
4. The trials are independent.
5. The binomial random variable x is the number of S's in n trials.

The probability mass function (pmf) of Binomial distribution

If a random variable X is given, where X ~ Bin(n, p), then

$$p(x) = \binom{n}{x} p^x q^{n-x} = {}_nC_x p^x q^{n-x} = {}_nC_x p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \ldots, n)$$

where

p = probability of a success on a single trial

q = probability of failure, 1-p

n = number of trials

x = number of successes in n trials

n-x = number of failures in n trials

$\binom{n}{x} = {}_nC_x = \frac{n!}{x! * (n-x)!}$

Example 3.3.

A company hires management trainees for entry level sales positions. Past experience indicates that only 10% will still be employed at the end of nine months. Assume the company recently hired six trainees.

a) What is the probability that three of the trainees will still be employed at the end of nine months?
b) What is the probability that at least two of the trainees will still be employed at the end of nine months?

Solution.

a) Let X - number of trainees employed at the end of nine months, then X ~ Bin(6, 0.10)

$P(X = 3) = {}_6C_3 * 0.10^3 * (1-0.10)^{6-3} = \frac{6!}{3!(6-3)!} * (0.1 * 0.9)^3 = 0.015$

b) $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - [{}_6C_0 * 0.10^0 * (1-0.10)^{6-0} + {}_6C_1 * 0.10^1 * (1-0.10)^{6-1}] =$
$= 1 - [0.531 + 0.354] = 0.115$

Mean, Variance, and Standard Deviation for a Binomial Random Variable

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard deviation: $\sigma = \sqrt{npq}$

Example 3.4 Given that $X \sim Bin(8, 0.25)$, calculate the mean and standard deviation of X.

Solution. Since n = 8 and p = 0.25,

$E(x) = \mu = 8*0.25 = 2$, $SD(x) = \sigma = \sqrt{8 * 0.25 * 0.75} = 1.225$

Geometric Probability Distribution

A random variable X that has a geometric distribution is denoted by $X \sim Geo(p)$, and the probability that first success occurs on the k^{th} trial is

$$\mathbf{P(X = k) = (1-p)^{k-1}p}, \quad \text{where } k = 1, 2, \ldots$$

Note, $P(X \leq k) = 1 - q^k$ and $P(X > k) = q^k$, where $q = 1-p$

Mean, Variance and Standard Deviation of Geometric distribution:

$$\mu = \frac{1}{p} \quad \sigma^2 = \frac{1-p}{p^2} \quad \sigma = \sqrt{\frac{1-p}{p^2}}$$

Example 3.5 You are rolling a die in an experiment. Find the probability that you get 3:

a) on the third trial.
b) on or before the second trial.
c) after the third trial.

Solution. The probability of getting 3 is 1/6 on each trial and the trials are independent. Let X be the number trials until you get first 3.

a) $P(X = 3) = (\frac{5}{6})^{3-1} * \frac{1}{6} = 0.116$
b) $P(X \leq 2) = P(X = 1) + P(X = 2) = p + (1-p)p = \frac{1}{6} + \frac{5}{6} * \frac{1}{6} = 0.306$
c) $P(X > 3) = (1-p)^3 = (\frac{5}{6})^3 = 0.579$

TASKS

1. For the following exercises, decide whether the statement makes sense (or is clearly true) or does not make sense (or is clearly false).

a. The probability mass function (pmf) of a discrete random variable must sum to 1.

b. The probability of any single outcome in a discrete random variable is always greater than zero.

c. The expected value of a discrete random variable is always one of its possible outcomes.

d. The binomial distribution can be used to model the number of times a coin lands on heads in a sequence of 10 flips.

e. The geometric distribution can be used to model the height of a student.

f. The probability of getting at least 1 head in 3 coin flips is 0.5.

g. The binomial distribution is a memoryless distribution.

h. The larger the success probability p, the longer we expect to wait for the first success.

2. An investment company has produced the following table, which shows the probabilities of various percentage profits on money invested over a period of 1 year.

Profit (%), x	-3	0	1	5	8	12
Probability, p(x)	0.10	0.10	0.20	0.30	2k	k

a. Find the value of k.

b. What is the probability of completing the year with a positive profit?

c. Calculate the expected profit on an investment of $50,000.

d. If $\sum x^2 p(x) = 35.8$, then compute the standard deviation of profit rate.

3. A small business sells three products: Product A, Product B and Product C. The probabilities of selling each product on any given day are as follows:

- P(A) = 0.4
- P(B) = 0.35
- P(C) = k

The profits from each product sale are:

- Product A: $50
- Product B: $40
- Product C: $C

Given that mean profit per product is $45, do the following:

a. Construct the discrete probability distribution table for the profits from selling a product.

b. Determine the standard deviation of the profit distribution.

4. Two 4-sided dice, labelled 1, 2, 3, 4 are rolled. The product and the sum of the two numbers obtained are calculated. The score awarded, S, is equal to the absolute (i.e. non-negative) difference between the product and the sum. For example, if 4 and 2 are rolled, then $S = |(4 \times 2) - (4 + 2)| = 2$.

a. Provide the probability distribution of S.

b. Calculate E(S).

5. On average, 15% of the WIUT graduates are awarded a first-class degree. Groups of 20 graduates are selected at random.

a. How many candidates in each group are not expected to be awarded a first-class?

b. Calculate the variance of the number of first classes in the groups of 20.

c. Find the probability that:

i. four graduates in a group of 20 are awarded first-class.

ii. at least three graduates in a group of 20 are awarded first-class.

6. In a particular country, 90% of both females and males drink tea. Of those who drink tea, 40% of the females and 60% of the males drink it with sugar. Find the probability that in a random selection of two females and two males:

a. all four people drink tea.

b. at most one male and at most one female drink tea.

c. an equal number of females and males drink tea with sugar (exclude the case of 0).

7. A computer generates random numbers using any of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The numbers appear on the screen in blocks of five digits, such as $\boxed{50119}$ $\boxed{16317}$ $\boxed{40068}$ ....

Find the probability that:

a. there are no 7s in the first block.

b. the first zero appears in the first block.

c. the first 9 appears in the second block.

8. A microfinance bank issues 4 independent business loans in one week. Each loan has a default probability of 0.2. If a loan defaults, the bank loses \$1,000. If the loan is repaid, the bank earns \$300. Let X = "number of defaults among the 4 loans" and Y="net profit from the 4 loans".

a. Identify the probability distribution of X and write its probability mass function (pmf).

b. Construct the discrete probability distribution of Y.

HOMEWORK

9. John Doe sells new cars for Pelican Ford. John usually sells the largest number of cars on Saturday. He has developed the following probability distribution for the number of cars he expects to sell on a particular Saturday.

Number of Cars Sold, *x*	Probability, *P*(*x*)
0	.1
1	.2
2	.3
3	.3
4	.1
	1.0

a. On a typical Saturday, how many cars does John expect to sell?

b. What is the probability of selling at least 2 cars?

c. What is the probability of selling no cars or at most 2?

d. What is the standard deviation of the distribution?

10. The probability distribution table for the random variable X is given.

x	0	1	2	3
P(X = x)	1-2k	2-4k	3-6k	4-8k

a. Find the value of the constant k.

b. ShowthatE(X)=2.

c. Find Var (X).

11. The Census Bureau reports that 27% of California residents are foreign-born. Suppose that you choose three Californians at random, so that each has probability 0.27 of being foreign-born and the three are independent of each other. Let the random variable W be the number of foreign-born people you chose.

a. What are the possible values of W?

b. What is the probability of all three being foreign born?

c. What is the probability that at least one of them is foreign born?

d. What is the probability that at most one of them is foreign born?

12. The following notice appeared in the golf shop at a Myrtle Beach, South Carolina, golf course.

> **Blackmoor Golf Club Members**
>
> The golf shop is holding a raffle to win a
> TaylorMade M1 10.5° Regular Flex Driver ($300 value).
> Tickets are $5.00 each.
> Only 80 tickets will be sold.
> Please see the golf shop to get your tickets!

What is the expected gain from playing this raffle?

13. An FBI survey shows that about 80% of all property crimes go unsolved. Suppose that in your town 4 such crimes are committed and they are each deemed independent of each other.

a. What is the probability that 1 of 4 of these crimes will be solved?

b. What is the probability that at least 1 of 4 of these crimes will be solved?

c. Suppose total of 5985 crimes were reported to FBI during the month of September 2019. Find the mean and standard deviation of solved crimes.

14. A biased 4-sided die is numbered 1, 3, 5 and 7. The probability of obtaining each score is proportional to that score.

a. Find the expected number of times that the die will be rolled, up to and including the roll on which the first non-prime number is obtained.

b. Find the probability that the first prime number is obtained on the third roll of the die.