Chapter 5
Discrete Probability Distributions

5.1 – 5.3 Discrete Random Variables, Expected value, and Variance

Consider a random experiment with a sample space $S = \{s_1, s_2, ..., s_n\}$. A **random variable** is a rule (or a function) that assign a **numerical value** for each outcome of the experiment (describing the experiment outcome by a number)
That is, a **random variable is a numerical description** of the outcome of an experiment.

Random variables are denoted by X, Y, R,

Two types of random variables

Discrete Random Variable (DRV): A variable that may assume a finite number of values $x_1, x_2, ..., x_n$ or an infinite sequence of values $x_1, x_2, ..., x_n, ...$

Example 1

1. Number of students attend this online lecture: 0, 1, …, 80
2. Number of heads shown up on tossing a coin 10 times: 0,1, 2,.., 10
3. Number of children in a family : 0,1, 2, 3, 4, 5, 6
4. Number of questions answered correctly on 20-questions exam: 0, 1, 2,..., 20

Continuous Random Variable (CRV): A random variable that may assume a value in an interval or collection of intervals.

Example 2

1. Time X needed to finish a STAT 2361 final exam: $x \geq 0$
2. Weight of a new born baby Y: $2 \leq y \leq 5$
3. Height of a student in the class.
4. Distance traveled between classes

Discrete random Variables

Example 1

In the experiment of tossing a coin twice, the number of sample points is (2) (2) = 4, the sample space is;

S = {HH, HT, TH, TT}

Define the random variable **X** to be the number of heads observed, then

HH $\Rightarrow$2, HT $\Rightarrow$1, TH $\Rightarrow$1, TT $\Rightarrow$0.

X can assume the values 0, 1, and 2

X = {0, 1, 2}, X is a discrete random variable.

Example 3:

In the experiment of rolling a die twice, the sample space is given by the following 36 ordered pairs;

S = {(1, 1), …, (6, 6)}

- The outcomes are equally likely. That is, all outcomes have the same probability 1/36.

- Define the random variable **X** to be the **sum** of the two faces observed, then

(1, 1) $\Rightarrow$2, (1, 2) $\Rightarrow$3, (6, 6)$\Rightarrow$12.

X can assume the values: 2, 3, …, 12

X = {2, 3, …, 12}, X is a discrete random variable

- Define the random variable **Y** to be the **absolute difference** of the two faces observed, then

(1, 1) $\Rightarrow$0, (1, 2) $\Rightarrow$1, (1, 6)$\Rightarrow$5.

Y can assume the values 0, 1, …, 5

Y = {0, 1, …, 5}, Y is a discrete random variable

Example 4:

In the experiment the of tossing a coin three times, the sample space is

S = {HHH, HHT, …, TTT} (chapter 4)

The outcomes are equally likely. That is, all outcomes have the same probability 1/8.

Define the random variable **X** to be the number of heads observed, then

X = 0, 1, 2, 3

Now let us construct a table for values of X with the associated probability of occurring of each value

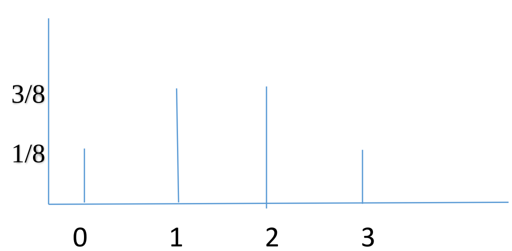
x	P(X = x)
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

This table is called the **probability distribution** for the random variable X (it describes how probabilities are distributed over the random variable).

The function f(x) = (P(X = x) is called the **probability function.** The probability function satisfied the two requirements:

$$1. f(x) \geq 0$$
$$2. \sum f(x) = 1$$

The following is a graphical representation of the random variable X,



Example 5

Consider the following probability distribution for the random variable X

X	f(x)
-4	0.05
-2	0.15
-1	0.2
0	
3	0.35

1. Find f(0), that is, find P(X = 0))
   f (0) = 1 – (0.05 + 0.15 + 0.2 + 0.35) = 0.25
2. Find $P(X \le 0)$
   $$P(X \le 0) = f(0) + f(-1) + f(-2) + f(-4)$$
   $$= 1 - P(X > 0) =$$
   $$= 1 - f(3) = 1 - 0.35$$
   $$= 0.65$$
3. Find $P(-4 < X \le 0)$
   $$P(-4 < X \le 0) = f(-2) + f(-1) + f(0)$$
   $$= 0.15 + 0.2 + 0.25 = 0.6$$

Expected value and Variance

Consider a discrete random variable $X = \{x_1, x_2, ..., x_n\}$ with probability function f(x) [$f(x) \ge 0, \sum f(x) = 1$], then

1. The **expected value** (the mean) of X, $E(X) = \mu$, is define by
   $$E(X) = \mu = \sum x f(x)$$
2. The **variance of X**, $Var(X) = \sigma^2$ is define by
   $$Var(X) = \sigma^2 = \sum (x - \mu)^2 f(x)$$
3. The standard deviation of X, $\sigma(X)$ is
   $$\sigma(X) = \sqrt{Var(X)} = \sqrt{\sum (x - \mu)^2 f(x)}$$

Example 5

Find the expected value, the variance and the standard deviation for the random variable of example 4.

X	P(X = x)	x*f(x)	$x - E(x)$	$(x - E(x))^2$	$(x - E(x))^2 f(x)$
0	1/8	0	-3/2	9/4	9/32
1	3/8	3/8	-1/2	1/4	3/32
2	3/8	6/8	½	1/4	3/32
3	1/8	3/8	3/2	9/4	9/32
Total		**12/8**	**0**		**24/32**

The expected value, $E(X) = \sum xf(x)$ =12/8 = 1.5

(The expected number of heads is 1.5).

The variance, $Var(X) = \sum (x - \mu)^2 f(x)$ =24/32 = 0.75

The standard deviation $\sigma = \sqrt{Var(x)} = \sqrt{0.75}$ = 0.8866

Example 6

A discrete random variable X has the following probability function:

$$f(x) = \frac{x+2}{10} \qquad x = 0, 1, 3$$

1. Is f(x) a valid probability function?
   f(0) = 2/10, f(1) = 3/10, f(3) = 5/10
   $\Rightarrow f(x) \geq 0, \; \sum f(x) = 0.2 + 0.3 + 0.5 = 1$
2. Find E(X)
   $E(X) = \sum xf(x) = (0)(0.2) + (1)(0.3) + (3)(0.5) = 1.8$
3. Find the variance of X
   $Var(X) = \sum (x - \mu)^2 f(x) =$
   $= (1.8)^2 (0.2) + (0.8)^2 (0.3) + (1.2)^2 (0.5) = 1.56$

Example 7

Let X be a discrete random variable with probability function

$f(x) = \frac{1}{n}$, n = number of values of the random variable X.

- X is called a discrete uniform probability distribution
- For example, rolling a die, S= {1, 2, 3, 4, 5, 6}, f(x) = 1/6
- $f(x) = \frac{1}{n} > 0$, $\sum f(x) = n.\frac{1}{n} = 1$. So, f(x) is a valid probability function.

Example 8 Exercise17 page 197

A volunteer ambulance service handles 0 t0 5 service calls on any given day. The probability distribution for the number of service calls is as follows.

Number of service Calls	Probability
0	0.1
1	0.15
2	0.30
3	0.20
4	0.15
5	0.10

1. What is the expected number of service calls?
2. What is the variance number of service calls? What is the standard deviation?

We will use calculator to find all of the above

- ✓ Shift mode 3 = =
- ✓ Mode 2
- ✓ 0 shift 0.1 M+, 1 shift, 0.15 M+, …, 5 shift, 0.1 M+
- ✓ 0 M+, 1 M+, 2M+, 3M+, 4M+, 5M+
- ✓ (REPLAY) ↓↓ 0.1 =, ↓↓ 0.15 =,..., ↓↓ 0.1
- ✓ Shift 2 1 = The expected value, $E(X) = 2.45$
- ✓ Shift 2 2 = The standard deviation $\sigma = 1.4310$
- ✓ The variance, $Var(X) = \sigma^2 = (1.431) \wedge 2 = 2.0475$

Example 9

The following distribution shows the number of smart phones soled per day on ABC store over a period of 200 days.

Number of units sold X	Number of days frequency	f(x)
0	40	0.2
1	70	0.35
2	60	0.30
3	20	0.10
4	10	0.05

- Let X = number of units sold per day
  X = 0, 1, 2, 3, 4
- Use relative frequency approach to find P(X = x) = f(x)

1. What is the probability that the store will not sell any phone?
2. What is the probability that the store will sell at least one phone?
3. What is the expected number of phones sold per day? What is the standard deviation?
4. If the profit on each phone is $25, what is the expected profit per day?

5.4 The Binomial Probability Distribution

The **binomial probability distribution** is discrete probability distribution used when there are **exactly two mutually exclusive outcomes** of a trial. These outcomes are called "**success**" and "**failure**".

The **binomial experiment** is an experiment satisfies the following four properties;

1. The experiment consists of a sequence of n identical trials.
2. Two outcomes, success and failure, are possible on each trial.
3. The probability of success p, and the probability of failure 1 – p do not change from trial to trial.
4. The trials are independent.

Example 1

1. Tossing a coin 15 times. Head means success.
2. Rolling a die 4 times, we are interested on the face shown up to be 6 (probability of success is 1/6 and probability of failure is 5/6).
3. Insurance salesperson visits 10 randomly selected families. A success means a family purchases an insurance policy (the salesperson knows the probability/ past experience).
4. Select three item for inspection from a production line. Defective means success (probability of defective is known).

The binomial distribution is used to obtain the probability of observing ***x* successes in *n* trials**, with the probability of success on a single trial denoted by *p*. The binomial distribution assumes that *p* is fixed for all trials.

Let us consider a binomial experiment with **n** trials and probability of success **p**, define the random variable X to be the number of successes in the n trials.
The random variable X can assume the values 0, 1, 2, …, n. That is,

$$X = \{0, 1, 2, \ldots, n\}.$$

X is a discrete random variable and is called a **binomial random variable**, it is denoted by

X = B (n, p).

For example, toss a coin 10 times and define X to be the number of heads observed, X is a binomial random variable with 10 trials and probability of success 0.5. That is X = B (10, 0.5)

Assume that **X = B (n, p),** then

- ❖ The probability function for X is given by the following formula

  $f(x) = (nCx)(p)^{x}(1-p)^{n-x}$

Where, f(x) probability of x success in n trials

$nCx = \frac{n!}{(n-x)!x!}$,

x = number of successes, x = 0, 1, 2, …. , n

p = P (success), 1 – p = P (failure)

- ❖ The expected value of X , E(X), is given by

$E(X) = np$

- ❖ The Variance of X, Var (x), is given

$Var(X) = np(1-p)$

The standard deviation is

$\sigma(X) = \sqrt{np(1-p)}$

<u>Example 2</u>

Let X = B (6, 0.4)

1. Write the probability function

   $f(x) = (6Cx)(0.4)^{x}(0.6)^{6-x}, \; x = 0, 1, \ldots, 6$

2. Find E(X), Var (X), and $\sigma(X)$

   E(X) = np = (6) (0.4) = 2.4,

   Var (X) = np (1 – p) = (6) (0.4) (0.6) = 1.44

$\sigma(X) = \sqrt{1,44} = 1.2$

3. Find the probability of exactly 4 successes.

   $f(4) = (6C4)(0.4)^4(0.6)^2$

   Use calculator: (6 nCr 4)*(0.4^4)(0.6^2) = 0.1382

4. Find the probability of at least 2 successes.

   $P(X \geq 2) = f(2) + f(3) + f(4) + f(5) + f(6)$

   $= 1 - P(X < 2)$

   $= f(1) + f(0)$

   $= 0.2333$

5. Find the probability of no more than (at most) 2 failures.

   Two methods to solve this problem:

   - Method one: At most 2 failures means at least 4 success.
   - Method two: Define a new binomial random variable with 6 trials and probability of success equals 0.6. That is, X = B (6, 0.6) an

     $f(x) = (6Cx)(0.6)^x(0.4)^{6-x}, x = 0, 1, \ldots, 6$

     $P(X \leq 2) = f(2) + f(1) + f(0)$

     $= 0.1382 + 0.369 + 0.0004$

     $= 0.5076$

Example 3

A STAT 2361 test has 12 multiple choice questions with four choices and one correct answer each. If you just randomly guess on each of the 12 questions.

1. Define the random variable X to be the number of correct answers. How X is distributed?

X is a binomial random variable with n = 12 and probability of success equals ¼. That is,

X = B (12, 0.25) x = 0, 1, …, 12

2. Find the expected number of correct answers. Find the variance.

E(X) = np = (12) (0.25) = 3

Var (X) = np (1 – p) = (12) (0.25) (0.75) = 2.25

3. What is the probability that you get exactly 3 correct questions?

$$f(3) = (12C3)(0.25)^3(0.75)^9$$
$$= 0.2581$$

4. What is the probability that you get no more than 2 correct questions?

$$P(X \le 2) = f(2) + f(1) + f(0)$$
$$= 0.2323 + 0.1267 + 0.0317$$
$$= 0.3907$$

5. What is the probability that you get at least 10 false questions?

$P(X \ge 10)$, x = B (12, 0.75)

$$P(X \ge 10) = f(12) + f(11) + f(10)$$
$$= 0.0317 + 0.1267 + 0.2323$$
$$= 0.3907$$

<u>Example 4</u>

According to a survey 70% of BZU students attend the online lectures (Corona Time).

1. In a randomly selected sample of 1000 BZU students, what is the expected number of students who attend the online lecture? What is the variance?

X = B (1000, 0.7)

$E(X) = np = 700$

$Var(X) = \sqrt{np(1-p)} = 14.49$

2. In a randomly selected sample of 400 BZU students, what is the expected number of students who do not attend the online lecture? What is the variance?

X = B (400, 0.3)

$E(X) = np = 120$

$Var(X) = \sqrt{np(1-p)} = 9.17$

3. In a randomly selected sample of 10 BZU students, what is the probability that exactly 5 students attend the online lecture?

X = B (10, 0.7)

$f(6) = 10C6(0.7)^6(0.3)^4$

$= 0.2$

4. In a randomly selected sample of 8 BZU students, what is the probability that at least 3 students attend the online lecture?

X = B (8, 0.7)

$P(X \geq 3) = 1 - P(X < 3)$

$= f(2) + f(1) + f(0)$

$=$

5. In a randomly selected sample of 12 BZU students, what is the probability that at most 10 students do not attend the online lecture?

X = B (12, 0.3)

$P(X \leq 10) = 1 - P(X > 10)$

$= f(12) + f(11)$

$=$