DISCRETE PROBABILITY DISTRIBUTION

5.1 RANDOM VARIABLE

A variable that assumes a unique numerical value for each of the outcomes in the sample space of a probability experiment is called a *random variable*. In other words, a random variable is used to represent the outcome of a probability experiment. For example, if we toss four coins and the random variable x represents the number of tails that occur, then the only possible values it can assume are x = 0, 1, 2, 3, or 4. This is a discrete random variable. (it is called "random" because the value it assumes is the result of a chance, or random event). A discrete random variable is one that can assume any of a set of possible values that can be counted or listed.

Another example of discrete random variable is when we roll two dice and observe the sum that appears in both dice. The random variable x will be integer values from 2 to 12. When we deal with a discrete random variable and consider all the possibilities associated with it, we generate a *discrete distribution* or a probability distribution.

Mean of a Probability distribution is

$$\mu = \sum x P(x) \qquad (5.1)$$

The variance of a probability distribution is

$$\sigma^2 = \sum (x - \mu)^2 p(x) \text{ or } \sum x^2 p(x) - \mu^2 \qquad (5.2)$$

The standard deviation of a probability distribution

$$\sigma = \sqrt{\sum (x-\mu)^2 p(x)} \text{ or } \sqrt{\sum x^2 p(x) - \mu^2} \qquad (5.3)$$

Example 5.1: Toss a die once and give the possible outcomes. Find the mean and standard deviation.

Solution

The possible outcomes are 1, 2, 3, 4, 5, or 6, and so the random variable x giving the number on the top face is a discrete random variable. The associated probability distribution is as follows

x	1	2	3	4	5	6
P(x)	1/6	1/6	1/6	1/6	1/6	1/6

To find the mean and standard variance of x:

x	P(x)	x P(x)	x^2	x^2 P(x)
1	1/6	1/6	1	1/6
2	1/6	1/3	4	4/6
3	1/6	½	9	9/6
4	1/6	2/3	16	16/6
5	1/6	5/6	25	25/6
6	1/6	1	36	36/6
Total		7/2		91/6

$$\mu = \text{Mean} = \sum x\,P(x) = \frac{21}{6} = \frac{7}{2} = 3.5$$

$$\text{variance} = \sum x^2\,P(x) - \mu^2 = 91/6 - (3.5)^2$$

$$= 15.167 - 12.25 \approx 2.917$$

$$\text{standard deviation} = \sqrt{Variance} = \sqrt{2.917} = 1.708$$

Example 5.2: The number of calls x to arrive at a switchboard during any 1-minute period is a random variable and has the following probability distribution.

x	0	1	2	3	4
p(x)	0.2	0.1	0.3	0.3	0.1

Find the mean and standard deviation of x

Solution

x	p(x)	x p(x)	x^2	x^2 p(x)
0	0.2	0	0	0
1	0.1	0.1	1	0.1
2	0.3	0.6	4	1.2
3	0.3	0.9	9	2.7
4	0.1	0.4	16	1.6
Total	1.0	2.0		5.6

$$\mu = \text{Mean} = \sum x\,P(x) = 2.0$$

$$\sigma^2 = \text{variance} = \sum x^2\,P(x) - \mu^2 = 5.6 - 2^2$$

$$= 5.6 - 4 = 1.6$$

$$\sigma = \text{S.D.} = \sqrt{1.6} = 1.265$$

5.2 BERNOULLI DISTRIBUTION

A *Bernoulli experiment* is a random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, say, success or failure (e.g. true or false, male or female, good or bad, etc). We represent for example, the probability of success, say, p, and failure by 1 – p or q. *Bernoulli trial* occurs when a Bernoulli experiment is done a number of independent times.

The probability density function (pdf) is given as

$$f(x) = p^x (1 - p)^{1-x} \qquad x = 0, 1 \qquad (5.4)$$

where x is a random variable associated with a Bernoulli trial by defining

$$x(\text{success}) = 1 \text{ and } x(\text{failure}) = 0$$

We say x has a Bernoulli distribution with parameter p [denoted as X~ Bernoulli (p)]

The expected value or mean of x is

$$\mu = E(x) = \sum_{x=0}^{1} x P^x (1 - P)^{1-x} = 0(1 - p) + (1)(p) = p$$

and the variance of x is

$$\sigma^2 = \text{var}(x) = \sum_{x=0}^{1} (x - p)^2 p^x (1 - p)^{1-x}$$

$$= (0 - p)^2 (1-p) + (1 - p)^2 p.$$

$$= p(1 - p) = pq$$

as usual, the standard deviation of X will be

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

Example 5.3: If X~ Ber (0.7), find the mean and variance.

Solution

It can be deduced that p = 0.7 and q = 1 – p = 1 – 0.7 = 0.3

Therefore,

Mean = $\mu = E(x) = p = 0.7$

Variance = $\sigma^2 = V(x) = pq = p(1-p) = 0.7 \times 0.3$

$= 0.21$

5.2 BINOMIAL DISTRIBUTION

If we let the random variable X equal the number of observed successes in n Bernoulli trials, the possible values of x are 0, 1, 2,..., n. If x successes occur, where x = 0, 1, 2, ..., n, then n–x failures occur. The pdf of x, say f(x) is

$$f(x) = nC_x P^x (1 - P)^{n-x}, \quad x = 0, 1, 2, ... n \qquad (5.5)$$

where

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

The number of ways of selecting x positions for the x successes in the n trials. Where (5.5) is the binomial distribution of the random variable X.

The binomial experiment must possess the following properties:

1. Each trial has two possible outcomes (success, failure)
2. There are n repeated independent trial
3. The probability of success on each trial is a constant p; the probability of failure is q = 1 – p.
4. The random variable X equals the number of successes in the n trials.

The random variable x defined above is said to be a Binomial distribution with parameters n and p denoted as X~ Bi (n, p)

Example 5.4: If X is a binomial random variable, calculate the probability of x for

a. n = 3, x = 2, P = 0.3
b. n = 4, x = 0, P = 0.4

Solution

a. $P(X = x) = nC_x \; p^x (1 - p)^{n-x}$

$$P(X = 2) = 3C_2 (0.3)^2 (1-0.3)^{3-2} = \frac{3!}{1!\,2!} (0.3)^2 (0.7)^1$$

$$= 3 \times 0.09 \times 0.7 = 0.189$$

b. $P(x = 0) = 4C_0 \; (0.4)^0 (1 - 0.4)^{4-0} = \overline{4!} \; (0.6)^4$

0! 4!

$= (0.6)^4 = 0.1296$

Example 5.5: A coin is loaded in such a way that it lands with a head showing 70% of the time. If it is tossed five times, find the probability of getting.

i. 4 heads
ii. no head
iii. 3 heads
iv. at least 3 heads
v. at most 2 heads

Solution

Let x be the number of heads

then

$$f(x) = P(X = x) = nC_x \ (0.7)^x (0.3)^{n-x}$$

i. $P(X = 4) = 5C_4 (0.7)^4 (0.3)^{5-4} = 5(0.7)^4 (0.3)$

$= 0.36015$

ii. $P(X = 0) = 5C_0 (0.7)^0 (0.3)^{5-0} = (0.3)^5$

$= 0.00243$

iii. $P(X = 3) = 5C_3(0.7)^3 (0.3)^{5-3} = 5C_3 (0.7)^3 (0.3)^2$

$= 10(0.7)^3 (0.3)^2 = 0.3087$

iv. $P(X \geq 3) = P(x = 3) + P(x = 4) + P(x = 5)$

$P(X = 5) = 5C_5(0.7)^5 (0.3)^{5-5} = (0.7)^5 = 0.16807$

Therefore

$$P(x \geq 3) = 0.3087 + 0.36015 + 0.16807 = 0.83692$$

$P(x \leq 2) = 1 – P(x \geq 3) = 1 – 0.83692$

$= 0.16308$

Cumulative probabilities like those in the preceding example are often of interest. We call the function defined by

$$F(x) = P(X \leq x)$$

The C*umulative Distribution Function*, or more simply, the *distribution function* of the random variable X.

Values of the distribution function of a random variable X that is Bi (n, p) are given in Table I of the Appendix for selected values of n and p.

If $X \sim Bi(n, p)$ then

Mean = $E(x) = \mu = np$

Variance $= \sigma^2 = Var(x) = npq$ where $q = 1-p$

Example 5.6: Let $X \sim Bi(8, 0.6)$, use the cumulative Binomial table to find.

i. $P(x \leq 6)$
ii. $P(x \leq 3)$
iii. $P(x \leq 4)$
iv. $P(x = 0)$

Solution

We shall use the Cumulative Binomial Probability table in the Appendix.

i. Check for n = 8, p = 0.6 and c = 6
$P(x \leq 6) = 0.894$

ii. Checking similarly
$P(x \leq 3) = 0.174$

iii. $P(x \leq 4) = 0.406$

iv. $P(x = 0) = P(x \leq 0) = 0.001$

Example 5.7: It is observed that 30% of Nigerians do not have any health insurance. Suppose that this is true and let X equal the number with no health insurance in a random sample of n = 17 Nigerians. Find

i. $P(4 \leq X \leq 9)$
ii. $P(7 < X \leq 11)$
iii. $P(X > 7)$
iv. The coefficient of variation

Solution

$n = 17, \ p = 0.3$

i. $P(4 \le X \le 9)$ is the sum of probabilities 4, 5, 6, 7, 8 and 9, that is,

$P(4 \le X \le 9) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9)$

$= P(X \le 9) - P(X \le 3)$ from table, we have

$P(4 \le X \le 9) = 0.987 - 0.202 = 0.785$

ii. $P(7 < X \le 11)$ is the sum of probabilities from 8 to 11 excluding 7, That is,

$P(7 < X \le 11) = P(X \le 11) - P(X \le 7)$

$= 0.999 - 0.895 = 0.104$

iii. $P(X > 7) = 1 - P(X \le 7)$

$= 1 - 0.895 = 0.105$

iv. Coefficient of variation denoted by CV (X) is

$$= \frac{\text{Standard Deviation}}{\text{Mean}} = \frac{SD(X)}{E(X)} \times 100\%$$

Mean $= E(X) = np = 17 \times 0.3 = 5.1$

Variance $= \sigma^2 = np(1 - p) = 17 \times 0.3 \times 0.7 = 3.57$

S.D (X) $= \sqrt{Variance} = \sqrt{3.57} = 1.889$

$CV(X) = \frac{1.889}{5.1} \times 100\% \approx 37.04\%$