2.21 Find mean, variance and standard deviation for the random variable Y whose p.d.f is $f(y) = \begin{cases} \frac{2}{3}(2-y) , & 0 \le y \le 1 \\ 0 , & elsewhere \end{cases}$

2.22 The joint distribution of two independent random variables X and Y is

X \ Y	0	1
1	1/6	1/6
2	1/6	1/6
3	1/6	1/6

Find $E(X)$, $E(Y)$, $E(X+Y)$, $E(XY)$

2.23 Suppose that X and Y have the following joint distribution.

X \ Y	10	20	40	80
20	0.04	0.08	0.08	0.05
40	0.12	0.24	0.24	0.15

(i) Are X and Y independent?

(ii) Find $E(X)$, $E(Y)$, $E(X+Y)$, $E(XY)$

2.24 For the following joint distribution:

X \ Y	1	2
0	.06	.04
1	.30	.20
2	.24	.16

Find (i) $V(X)$ (ii) $V(Y)$

(iii) $V(X+Y)$ (iv) $V(X-Y)$

Unit -03

Special Discrete Probability Distributions

After studying this unit, the students will be able to

- Define discrete uniform random variable, discrete uniform probability distribution and discrete uniform probability mass function.
- Calculate mean, variance and standard deviation of discrete uniform probability distribution
- Define random digits/numbers and to know, how the random digits/numbers are generated.
- Solve real life problems using discrete uniform probability distribution.
- Define Bernoulli trials, Bernoulli probability distribution and Bernoulli mass function.
- Calculate mean, variance and standard deviation of Bernoulli probability distribution.
- Solve real life problems using Bernoulli probability distribution.
- Define Binomial experiment, Binomial random variable, Binomial probability distribution, Binomial probability mass function and Binomial frequency distribution.
- Calculate mean, variance and standard deviation of Binomial probability distribution
- Solve real life problems using Binomial probability distribution.
- Define hypergeometric experiment, hypergeometric random variable, hypergeometric probability distribution and hypergeometric probability mass function.
- Calculate mean, variance and standard deviation of hypergeometric probability distribution
- Solve real life problems using hypergeometric probability distribution.

3.1 When to use discrete probability distributions

Discrete probability distributions are used to compute probabilities for all possible values of discrete random variables, by an easy way. In this unit a few simple and commonly used distributions are given that cover many areas in our surrounding. It is important to know and remember that which probability distribution is suitable for a particular situation.

3.1.1 Introduction to discrete uniform distribution

This is the simplest probability distribution among the discrete probability distributions. It is used in the experiments/situations where probability at every point remains the same e.g. the outcomes of rolling a fair die, drawing cards from a well shuffled deck of cards, drawing of prize bond number etc. follow uniform distribution. A variable denoting the outcomes of such uniform experiments is called discrete uniform random variable and its probability distribution is called discrete uniform probability distribution. The most important application of the uniform distribution is in the generation of random numbers.

3.1.2 Definition

A probability distribution of the type:

X	x_1	x_2	x_3	...	x_N	Total
$p(x)$	$\frac{1}{N}$	$\frac{1}{N}$	$\frac{1}{N}$	$\frac{1}{N}$	$\frac{1}{N}$	1

is called discrete uniform probability distribution.

Example 3.1

Find probability distribution for the outcomes of a fair die when it is rolled once.

Solution:

When a fair die is rolled once, then all outcomes are equally likely, therefore, its probability distribution is given by

X	1	2	3	4	5	6	Total
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Note that X follows discrete uniform distribution.

3.1.3 Definition of discrete uniform probability function

A probability function defined as
$$p(x)=\begin{cases}\frac{1}{N} & , \quad x=1,2,\cdots,N \\ 0 & , \quad otherwise\end{cases}$$

is called discrete uniform probability function and the variable X is called discrete uniform random variable. This distribution has only one parameter "N", the total number of results /items of a uniform experiment.

Example 3.2

Write discrete uniform probability function for the results of a fair die when it is rolled once.

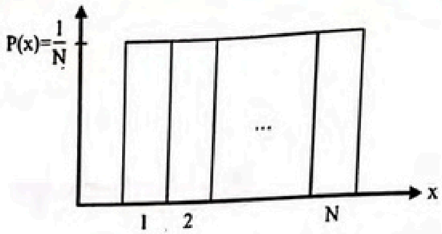
Solution:

Here random variable X takes all of its values with equal probability $\frac{1}{6}$.

Thus, discrete uniform probability function will be of the form:

$$p(x)=\begin{bmatrix}\frac{1}{6}, & x=1,2,3,4,5,6 \\ 0\ , & otherwise\end{bmatrix}$$

- Graphical representation of discrete uniform distribution by means of a histogram is a set of adjacent rectangles with equal heights i.e.



Due to rectangular shape, uniform distribution is also known as rectangular distribution

3.1.4 Properties of discrete uniform probability distribution

(i) Mean of discrete uniform distribution

By definition mean $= E(X) = \sum_x x\, p(x)$

Put range and formula of discrete uniform distribution, we get

$$E(X) = \sum_{x=1}^{N} x\frac{1}{N} = \frac{1}{N}\sum_{x=1}^{N} x = \frac{1}{N}[1+2+3+\cdots+N]$$

$$= \frac{1}{N}\left[\frac{N(N+1)}{2}\right] \quad as\ 1+2+3+\cdots+N = \frac{N(N+1)}{2}$$

$$= \frac{N+1}{2}$$

(ii) Variance of discrete uniform distribution

By definition variance: $V(X) = E(X^2) - [E(X)]^2$ (1)

We know that $E(X) = \frac{N+1}{2}$ (2)

Now $E(X^2) = \sum_{x=1}^{N} x^2 \frac{1}{N} = \frac{1}{N}\sum_{x=1}^{N} x^2 = \frac{1}{N}[1^2+2^2+\cdots+N^2]$

$$= \frac{1}{N}\left[\frac{N(N+1)(2N+1)}{6}\right], \quad as\ 1^2+2^2+\ \ +N^2 = \frac{N(N+1)(2N+1)}{6}$$

$$= \frac{(N+1)(2N+1)}{6} \qquad (3)$$

Put equation (2) and (3) in equation (1) we get,

$$V(X) = \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 = \frac{N+1}{2}\left[\frac{2N+1}{3} - \frac{N+1}{2}\right]$$

$$= \frac{N+1}{2}\left[\frac{4N+2-3N-3}{6}\right] = \frac{N+1}{2}\left[\frac{N-1}{6}\right]$$

$$= \frac{N^2-1}{12}$$

(iii) Standard deviation of uniform distribution

By definition $S.D(X) = \sqrt{V(X)} = \sqrt{\frac{N^2-1}{12}}$

Example 3.3

If X is a discrete uniform random variable taking values 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 each with probability equal to $\frac{1}{10}$. Find its mean, variance and standard deviation.

Solution:

Here total number of observations $= N = 10$

$$\therefore \text{Mean} = \frac{N+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5$$

$$\text{Variance} = \frac{N^2-1}{12} = \frac{10^2-1}{12} = \frac{100-1}{12} = \frac{99}{12} = 8.25$$

Standard deviation $= \sqrt{8.25} = 2.87$

3.1.5 Random digits/numbers

There are ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Involving equal chance $\frac{1}{10}$ of selection for each digit, then they are called random digits because now personal liking or disliking is impossible or one cannot predict in advance that which digit will come out during the selection process. Similarly pair of digits make a number e.g. 00, 01, 02,..., 99 each having probability $\frac{1}{100}$ or 001, 002, 003, ..., 1000 each having probability $\frac{1}{1000}$ and so on, are called random numbers. Thus, random digits/numbers can simply be defined as the digits or numbers which have equal chances of occurrence or their selection cannot be predicted in advance are called random digits/numbers.

3.1.6 Random number generator

Random number generator is a process (computational or physical device) designed to generate a sequence of numbers that cannot be accurately predicted. One of the simple procedures is that: write each digit separately on a slip of paper of equal size, fold it and put them in a basket or a bowl etc. and mix them thoroughly. Then draw out the digits blindly one by one with replacement and write in the order they occur and group them for convenience in blocks of two, three, four etc. The resulting set of random digits is called random number table. Nowadays several random number tables exist constructed by Tippet, Kendall and Smith, Fisher and Yates, RAND Corporation etc. Random numbers are also available in calculators, computer packages and on the Internet. In the real life, these random digits/numbers are used to make the selection process fair and free from personal biases. These are extensively used in the random sampling for selection of a random sample, which everywhere is necessary for precise statistical inference.

Table 3.1 Random digits (blocked merely for convenience)

3 6 7 4 4	1 9 6 6 6	2 2 3 3 6	5 2 9 1 9	9 0 7 9 0
9 5 6 5 5	9 0 9 4 1	0 6 6 1 2	8 4 7 4 7	5 6 0 9 0
7 7 2 7 9	6 9 6 1 1	3 5 3 1 1	4 8 7 5 3	3 6 3 2 4
3 1 3 0 0	5 7 0 2 1	1 6 4 9 9	7 3 5 1 3	0 2 8 0 6
7 2 4 3 8	5 6 4 9 1	9 4 0 1 6	2 0 0 5 8	5 7 1 5 4
2 0 7 3 4	1 7 7 7 9	8 0 7 7 6	3 5 9 1 0	9 8 1 2 9
7 1 2 6 1	1 9 2 7 5	2 3 4 7 4	6 8 8 0 2	9 2 7 4 8
5 7 5 9 7	7 5 1 8 8	4 3 5 7 8	9 1 4 9 9	3 2 0 5 0
3 4 7 8 7	6 5 0 3 3	0 3 6 1 2	3 0 7 7 5	2 9 5 0 4
7 8 9 8 4	3 2 6 4 0	1 1 0 0 7	5 7 9 1 3	8 9 2 1 1
0 8 0 1 9	8 5 8 2 9	5 2 0 2 4	2 7 5 0 8	7 6 8 7 6
2 9 8 6 4	5 3 3 9 5	6 7 9 4 3	1 8 5 9 2	2 1 8 3 1
8 1 1 7 0	3 7 4 1 3	4 1 9 5 9	4 9 4 7 5	0 1 2 5 9
7 8 5 0 7	7 9 9 2 8	8 3 3 5 6	1 2 5 1 1	9 8 5 8 4
9 8 7 7 1	8 0 3 5 2	0 0 4 4 2	1 3 5 2 4	5 4 1 2 8
8 3 1 0 5	9 9 9 9 4	0 6 1 1 0	4 6 9 5 7	4 5 7 4 9
1 0 5 0 0	7 6 1 4 6	5 7 1 1 8	5 5 1 5 2	6 0 1 7 8
0 8 8 7 4	6 2 6 8 8	8 6 7 4 6	9 3 1 2 1	6 4 8 2 7
4 9 5 3 3	7 8 1 5 4	2 5 1 2 9	2 0 6 5 4	9 2 0 8 4
7 0 6 7 1	1 2 3 0 1	7 5 0 4 2	8 4 7 3 4	5 3 0 4 7
9 0 3 0 0	3 1 0 5 3	7 1 3 3 2	6 9 2 9 3	3 6 7 3 3
3 5 1 3 7	4 8 4 2 5	4 3 9 5 2	8 5 3 2 5	6 3 0 5 3
2 8 1 4 9	1 2 9 4 1	7 5 9 8 8	1 7 9 3 5	9 1 0 4 6
1 9 1 7 9	1 0 9 5 8	6 1 4 4 6	3 9 3 7 5	8 6 4 1 7
9 1 9 0 9	2 3 8 2 7	7 5 9 1 9	0 6 8 2 2	5 2 1 3 4
5 8 4 6 7	7 7 3 8 1	9 7 5 3 1	9 1 7 5 1	6 0 1 2 4
9 0 3 5 7	1 6 8 9 5	7 3 6 9 4	2 4 9 3 6	2 9 7 7 6
7 8 1 5 3	0 5 1 2 9	7 1 1 5 6	0 4 0 2 4	6 9 6 5 3
6 2 8 5 1	5 7 7 1 3	4 3 7 9 2	7 3 3 0 0	6 8 2 2 2
9 6 8 6 3	9 1 4 7 2	8 8 5 3 9	3 7 2 4 5	0 2 9 0 5
4 4 1 2 0	8 6 9 9 1	8 8 8 0 4	7 4 8 5 5	7 8 0 9 3
1 7 0 5 3	7 3 3 5 7	1 3 3 4 9	7 5 5 0 1	9 8 1 7 0
9 9 1 6 6	0 6 7 5 6	9 7 4 8 7	2 1 9 5 2	1 8 1 4 2
1 4 4 3 2	8 1 4 1 9	2 9 3 9 9	9 8 4 1 3	4 5 1 7 3
8 0 5 1 4	0 5 8 0 4	4 4 3 9 2	7 6 7 0 8	5 0 4 9 3
0 6 4 8 7	8 2 5 8 0	8 0 5 4 2	2 5 1 8 6	0 3 2 9 6
6 7 7 6 9	8 8 8 4 0	6 7 0 2 6	8 3 2 7 0	6 6 7 2 1
7 2 7 3 9	9 5 4 6 6	4 1 6 1 6	9 7 0 0 1	1 5 0 2 2
5 4 2 1 7	4 2 7 2 1	0 9 7 1 0	8 3 1 3 1	8 8 1 1 3
9 0 4 3 5	2 9 2 3 9	6 4 6 0 8	1 0 5 9 4	1 3 7 6 3

66034	75309	96046	22168	99444
32641	93625	41932	01486	45865
74259	79936	38500	84077	19884
87330	93683	45259	53123	43710
86614	94764	51718	59595	26243
78253	78268	36131	97092	42021
46917	29075	84305	76423	77102
70207	67531	86838	28991	50684
58854	08374	68653	80828	88533
68791	67788	58989	82473	51698

21668	80255	66894	12098	43384
27854	72271	89054	19150	94567
13254	37370	75935	87381	05937
96796	97279	97269	94925	88444
17207	20229	24481	88663	07972
67327	86453	25441	70151	91145
74776	96696	84698	57182	53692
83611	77391	05732	93953	33571
02220	63100	55479	65467	45214
38868	90695	38707	66559	56097

03125	37918	27326	57569	66926
41716	33997	62977	37783	01721
60909	81749	40695	46300	99454
16831	74736	30580	50317	89607
26320	95962	57043	56068	99178
38560	93318	20584	65516	02000
13279	98576	13502	28204	89920
59509	36275	53958	27677	66896
57932	14232	81924	41074	71320
81604	86347	53944	98999	46239

92563	01673	80078	01185	53576
32349	71365	73498	83351	54729
77557	93799	50774	86124	66910
86844	28062	26995	23874	96299
68398	05265	47617	56161	09565
11162	07509	62087	56221	89547
48010	29662	42683	28326	19331
28705	48531	71182	70010	02562
79303	21902	17791	49598	98519
74059	80907	23392	99742	68771

3.1.7 Procedure of selecting random sample using random numbers

(i) If a population has $N = 10$ elements, allot digits 0, 1, 2....9 to elements of the population. If $N = 100$, then numbers 00, 01, 02... 99 are allotted. If $N = 1000$, then numbers 001, 002, 003... 1000 are allotted and so on.

(ii) Take any table of random numbers and take a random start from anywhere either from the top or bottom, vertically, horizontally or diagonally.

(iii) Draw a digit or number of two digits, number of three digits and so on according to your population size, ignore the number which is greater than the population size N or sampling is done without replacement and go further.

(iv) Continue the process till the desired number of sample units is obtained.

Example 3.4

Select a random sample of six students from a statistics class of 40 students by using a random number table.

Solution:

(i) Allot serial number 00, 01, 02, ..., 39 to students of the class.

(ii) Take a random number table and take numbers of two digits from anywhere, ignore the number which is greater than 39, and go further. Let the random numbers be 39, 37, 02, 10, 21, and 22.

(iii) The students whose serial numbers are 02, 10, 21, 22, 37, 39 are included in our sample. This is random sample because our personal judgment is not involved here and all the students were given equal opportunity through random number table for being included in the sample.

3.2 Bernoulli trail

A random experiment which has two possible outcomes classified as success and failure is called a Bernoulli trial. For example, gender of a child (male or female), performance of a student in an examination (pass or fail), result of tossing a coin (head or tail) etc. The random variable is taking only two values i.e. "0" for failure and "1" for success. Further, probability of success is denoted by p and failure by q such that $p + q = 1$.

3.2.1 Bernoulli probability distribution

A probability distribution of the type:

X	p(x)
0	q
1	p
Total	1

is called Bernoulli probability distribution and the variable X which denotes the results of a Bernoulli trail is called Bernoulli random variable. This distribution was introduced by a Swiss mathematician Jacob Bernoulli (1654–1705). It is also known as point Binomial distribution because it has only two classes of events. Mathematically

$$p(x)=\begin{cases} p^x q^{1-x} & , \ x=0,1 \\ 0 \ , & otherwise \end{cases}$$

is called Bernoulli probability mass function. This distribution has only one parameter i.e. p.

Example 3.5

A fair coin is tossed o ce. What is the probability that (i) no head occurs? (ii) one head occurs?

Solution:

Tossing a coin is Bernoulli experiment. Let us define a random variable as

X: number of heads

$= 0, 1$

$p = \text{probability of head} = \frac{1}{2}$ (as coin is fair).

$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$

The Bernoulli probability function is given by

$$p(x)=\begin{cases} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{1-x} & , \ x=0,1 \\ 0 \ , & otherwise \end{cases}$$

(i) probability of no head $= P(X=0) = \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{1-0} = (1)\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)$

(ii) Probability of one head $= P(X=1) = \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{1-1} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^0 = \left(\frac{1}{2}\right)$

3.2.3 Properties of Bernoulli distribution

- Mean of Bernoulli distribution

  By definition mean $= E(X) = \sum_x x\, p(x)$

  $$= \sum_{x=0}^{1} x\, p^x\, q^{1-x}$$

  $$= 0p^0q^{1-0} + 1p^1q^{1-1} = 0 + p = p$$

- Variance of Bernoulli distribution

  By definition variance $= E(X^2) - [E(X)]^2$

  $$= \sum_x x^2 p(x) - (p)^2$$

$$= \sum_{x=0}^{1} x^2 p^x q^{1-x} - p^2$$

$$= 0p^0q^{1-0} + 1p^1q^{1-1} - p^2$$

$$= 0 + p - p^2 = p\,(1-p) = pq,\ (\because q = 1-p)$$

- Standard deviation $= \sqrt{pq}$

Example 3.6

For a Bernoulli random variable X, the probability of success is equal to 0.6. Find the mean, variance and standard deviation for this Bernoulli probability distribution.

Solution:

(i) By calculation:

X	$p(x)$	$Xp(x)$	$X^2p(x)$
0	0.4	0	0
1	0.6	0.6	0.60
Total	1	0.6	0.60

Mean $= E(X) = \Sigma x\, p(x) = 0.6$

Variance $= V(X) = E(X^2) - [E(X)]^2 = \Sigma x^2\, p(x) - (0.6)^2 = 0.6 - 0.36 = 0.24$

(ii) By properties:

Given that $p = 0.60$, $q = 1 - p = 1 - 0.6 = 0.40$

Mean of Bernoulli distribution $= p = 0.60$

Variance of Bernoulli distribution $= pq = (0.60)\,(0.40) = 0.24$

$S.D(X) = \sqrt{0.24} = 0.49$

3.3 Binomial experiment

If a Bernoulli trail is repeated a fixed number of times, say n, then such an experiment is called Binomial experiment. It has the following four properties:

i. Each trial has only two possible outcomes i.e. success and failure.
ii. The probability of success remains constant for all trails.
iii. The successive trails are all independent.
iv. The Bernoulli trail is repeated a fixed number of times, say n.

The variable denoting the number of successes of a Binomial experiment is called Binomial random variable i.e. $X = 0, 1, 2, \ldots, n$.

3.3.1 Binomial probability distribution

The binomial probability distribution is given as:

X	0	1	2	...	n	Total
$p(x)$	${}^{n}C_{0}\, p^0 q^{n-0}$	${}^{n}C_{1}\, p^1 q^{n-1}$	${}^{n}C_{2}\, p^2 q^{n-2}$	...	${}^{n}C_{n}\, p^n q^{n-n}$	1

3.3.2 Binomial probability mass function

The Binomial probability mass function is given by the formula:

$$p(x) = \begin{cases} {}^{n}C_{x}\, p^x q^{n-x}, & x = 0,1,2,\ldots,n \\ 0, & otherwise \end{cases}$$

This distribution has two parameters n, p whereas n denotes the number of independent Bernoulli trails and p denotes the probability of success on a single Bernoulli trail. Note that if $n = 1$, the Binomial distribution reduces to Bernoulli distribution.

Example 3.7

A fair coin is tossed four times. Find the probability distribution for obtaining various numbers of heads.

Solution:

Here $n = 4$ (four Bernoulli trails)

X: the number of successes in four trials

$X = 0, 1, 2, 3, 4$

$p = \frac{1}{2}$ (probability of head on a single coin)

$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$

The Binomial probability function is:

$$p(x) = \begin{cases} {}^{4}C_{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{4-x}, & x = 0,1,2,3,4 \\ 0, & otherwise \end{cases}$$

The Binomial probability distribution is

X	$p(x) = {}^{4}C_{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{4-x}$
0	${}^{4}C_{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{4-0} = \frac{1}{16}$
1	${}^{4}C_{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{4-1} = \frac{4}{16}$
2	${}^{4}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{4-2} = \frac{6}{16}$
3	${}^{4}C_{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{4-3} = \frac{4}{16}$
4	${}^{4}C_{4}\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{4-4} = \frac{1}{16}$
Total	1

Example 3.8

Suppose X has a Binomial distribution with $n = 6$ and $p = 0.75$. Find (i) $P(X = 4)$ (ii) $P(X \geq 5)$ and (iii) $P(X < 4)$.

Solution:

Here $n = 6$

X: number of successes in 6 trials

$X = 0, 1, 2, 3, 4, 5, 6$

$p = 0.75$

$q = 1 - p = -0.75 = 0.25$

The Binomial probability function is

$$p(x) = \begin{cases} {}^{6}C_{x}(0.75)^{x}(0.25)^{6-x}, & x = 0,1,2,3,4,5,6 \\ 0, & otherwise \end{cases}$$

(i) $P(X = 4) = {}^{6}C_{4}(0.75)^{4}(0.25)^{6-4} = 0.297$

(ii) $P(X \geq 5) = {}^{6}C_{5}(0.75)^{5}(0.25)^{6-5} + {}^{6}C_{6}(0.75)^{6}(0.25)^{6-6} = 0.534$

(iii) $P(X < 4) = {}^{6}C_{0}(0.75)^{0}(0.25)^{6-0} + {}^{6}C_{1}(0.75)^{1}(0.25)^{6-1} + {}^{6}C_{2}(0.75)^{2}(0.25)^{6-2} +$
${}^{6}C_{3}(0.75)^{3}(0.25)^{6-3} = 0.169$

Example 3.9

A certain drug treatment cures 90% of cases of hookworm in children. Suppose that 20 children suffering from hookworm are to be treated and that the children can be regarded as a sample from the population. Find the probability that (i) all 20 children will be cured (ii) exactly 18 will be cured (iii) at least one will be cured.

Here $n = 20$

X: number of children to be cured

$X = 0, 1, 2, 3... 20$

$p = 0.90, q = 0.1$

$$p(x) = \begin{cases} \overset{20}{\underset{x}{C}}(0.90)^x (0.10)^{20-x}, & x = 0,1,2,3,...,20 \\ 0, & otherwise \end{cases}$$

(i) $P(\text{all 20 will be cured}) = P(X = 20) = \overset{20}{\underset{20}{C}}(0.90)^{20}(0.10)^{20-20} = 0.12158$

(ii) $P(\text{exactly 18 will be cured}) = P(X = 18) = \overset{20}{\underset{18}{C}}(0.90)^{18}(0.10)^{20-18} = 0.28518$

(iii) $P(\text{at least one will be cured}) = 1 - P(\text{no one will be cured})$

$$= 1 - P(X = 0)$$

$$= 1 - \overset{20}{\underset{0}{C}}(0.90)^0 (0.10)^{20-0}$$

$$= 1 - 0 \text{ (approximately)}$$

$$= 1$$

Example 3.10

In Peshawar 42 % of the population has type-A blood. Consider, taking a sample of size 4. Let Y denotes the number of persons in the sample with type-A blood. Find (i) $P(Y = 0)$ (ii) $P(Y = 1)$ (iii) $P(0 \le Y \le 2)$ and (iv) $P(0 < Y \le 2)$

Solution:

Given that $p = 0.42, q = 0.58, n = 4$

Y: number of persons in the sample with type-A blood

$Y = 0, 1, 2, 3, 4$

$$\therefore p(y) = \begin{cases} \overset{4}{\underset{y}{C}}(0.42)^y (0.58)^{4-y}, & y = 0,1,2,3,4 \\ 0, & otherwise \end{cases}$$

(i) $P(Y = 0) = \overset{4}{\underset{0}{C}}(0.42)^0 (0.58)^{4-0} = 0.11316$

(ii) $P(Y = 1) = \overset{4}{\underset{1}{C}}(0.42)^1 (0.58)^{4-1} = 0.32779$

(iii) $P(0 \le Y \le 2) = P(Y = 0) + P(Y = 1) + P(Y = 2)$

$$= 0.11316 + 0.32779 + \overset{4}{\underset{2}{C}}(0.42)^2 (0.58)^{4-2}$$

$$= 0.44095 + 0.35609$$

$$= 0.79700$$

(iv) $P(0 < Y \le 2) = P(Y = 1) + P(Y = 2)$

$$= 0.32779 + 0.35605 = 0.68384$$

Example 3.11

A machine produces 10 per cent defective items. Ten items are selected at random. Find the probability of not more than two items being defective.

Solution:

We have $p = 10\% = \frac{10}{100} = \frac{1}{10}$

$\therefore q = \frac{9}{10}$

$n = 10$

X: number of defective items

$X = 0, 1, 2, 3... 10$

$$p(x)=\begin{cases}{}^{10}C_x\left(\frac{1}{10}\right)^x\left(\frac{9}{10}\right)^{10-x}, & x=0,1,2,...,10\\ 0\quad, & otherwise\end{cases}$$

Now

P[not more than two items being defective] $= P[X \le 2]$

$$={}^{10}C_0\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^{10-0}+{}^{10}C_1\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^{10-1}+{}^{10}C_2\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^{10-2}$$

$$=\left(\frac{9}{10}\right)^{10}+10\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)^9+45\left(\frac{1}{100}\right)\left(\frac{9}{10}\right)^8$$

$$= 0.3487 + 0.3874 + 0.1937 = 0.9298$$

Example 3.12

The chances of a bomber hitting the target and missing the target are 3:2. Calculate the probability that the target will be hit at least once in five sorties.

Solution:

Given that $p=\frac{3}{5}, q=\frac{2}{5}, n=5$

X: denotes the number of hits.

$X = 0, 1, 2, 3, 4, 5$

So $$p(x)=\begin{cases}{}^{5}C_x\left(\frac{3}{5}\right)^x\left(\frac{2}{5}\right)^{5-x}, & x=0,1,2,3,4,5\\ 0\quad, & otherwise\end{cases}$$

Now, P[at least one will hit the target] $= 1 - P$[no one will hit the target]

$$P(X \ge 1) = 1 - P[X = 0]$$

$$= 1-{}^{5}C_0\left(\frac{3}{5}\right)^0\left(\frac{2}{5}\right)^{5-0}$$

$$= 1-\left(\frac{2}{5}\right)^5 = 1-0.01024 = 0.98976$$

3.3.3 Properties of Binomial distribution

(i) Mean of Binomial distribution

By definition

$$\text{Mean} = E(X) = \sum_x x\,p(x)$$

$$= \sum_{x=0}^{n} x\,{}^{n}C_x\, p^x q^{n-x}$$

$$= 0\,{}^{n}C_0\, p^0 q^{n-0} + 1\,{}^{n}C_1\, p^1 q^{n-1} + 2\,{}^{n}C_2\, p^2 q^{n-2} + \ldots + n\,{}^{n}C_n\, p^n q^{n-n}$$

$$= 0 + npq^{n-1} + \ldots + np^n$$

$$= np\,[q^{n-1} + \ldots + p^{n-1}]$$

$$= np\,[q+p]^{n-1}$$

$$= np\,(1)^{n-1} \qquad \because p+q=1$$

$$= np$$

(ii) Variance of Binomial distribution

By definition

$$\text{Variance} = V(X) = E(X^2) - [E(X)]^2 \qquad (A)$$

$$\text{As } E(X) = np \qquad (1)$$

Now $E(X^2) = \Sigma x^2\, p(x)$

$$= \sum_x [x(x-1) + x]\, p(x) \qquad \because x^2 = x(x-1) + x$$

$$= \sum_x x(x-1)\, p(x) + \sum_x x\, p(x)$$

$$= \sum_{x=0}^{n} x(x-1) \overset{n}{\underset{x}{C}} p^x q^{n-x} + \sum_{x=0}^{n} x \overset{n}{\underset{x}{C}} p^x q^{n-x}$$

$$= \left[0(0-1) \overset{n}{\underset{0}{C}} p^0 q^{n-0} + 1(1-1) \overset{n}{\underset{1}{C}} p^1 q^{n-1} + 2(2-1) \overset{n}{\underset{2}{C}} p^2 q^{n-2} + \ldots + n(n-1) p^n q^{n-n} \right] + [np]$$

$$= [0 + 0 + n\,(n-1)\, p^2 q^{n-2} + \ldots + n\,(n-1)\, p^n] + (np)$$

$$= n\,(n-1)\, p^2\, [q^{n-2} + \ldots + p^{n-2}] + (np)$$

$$= n\,(n-1)\, p^2\, [q + p]^{n-2} + (np)$$

$$= n\,(n-1)\, p^2\, (1)^{n-2} + (np) \qquad \text{As } p + q = 1$$

$$= n^2p^2 - np^2 + np \qquad (2)$$

Putting equation (1) and equation (2) in equation (A) we get

$$V(X) = n^2p^2 - np^2 + np - n^2p^2$$

$$= np - np^2$$

$$= np\,(1-p)$$

$$= npq \qquad (\because q = 1 - p)$$

(iii) Standard deviation of Binomial distribution

By definition

$$S.D(X) = \sqrt{V(X)}$$

$$= \sqrt{npq}$$

Example 3.13

Find the mean, variance and standard deviation of the Binomial distribution whose parameters are $n = 20$ and $p = \frac{3}{5}$.

Solution:

We have $n = 20$, $p = \frac{3}{5}$, $q = \frac{2}{5}$, therefore,

$$\text{Mean} = n\,p = 20 \times \frac{3}{5} = 12$$

$$\text{Variance} = npq = 20 \times \frac{3}{5} \times \frac{2}{5} = 4.8$$

$$S.D(X) = \sqrt{npq} = \sqrt{4.8} = 2.19$$

Example 3.14

If $n = 4$, $p = \frac{1}{3}$, find (i) the complete Binomial probability distribution (ii) mean and variance of this distribution. (iii) calculate the mean and variance using properties and compare the results.

Solution:

(i) Here $p = \frac{1}{3}$, $q = \frac{2}{3}$, $n = 4$, $X = 0, 1, 2, 3, 4$

$$p(x) = \begin{cases} \overset{4}{\underset{x}{C}} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{4-x}, & x = 0,1,2,3,4 \\ 0, & \textit{otherwise} \end{cases}$$

X	$p(x)=\binom{4}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{4-x}$	$Xp(x)$	$X^2p(x)$
0	$\binom{4}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{4-0}=\frac{16}{81}$	0	0
1	$\binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{4-1}=\frac{32}{81}$	$\frac{32}{81}$	$\frac{32}{81}$
2	$\binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{4-2}=\frac{24}{81}$	$\frac{48}{81}$	$\frac{96}{81}$
3	$\binom{4}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{4-3}=\frac{8}{81}$	$\frac{24}{81}$	$\frac{72}{81}$
4	$\binom{4}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^{4-4}=\frac{1}{81}$	$\frac{4}{81}$	$\frac{16}{81}$
Total	1	$\frac{108}{81}$	$\frac{216}{81}$

(ii) Mean $= E(X) = \Sigma x\,p(x) = \frac{108}{81} = 1.333$

Variance $= E(X^2) - [E(X)]^2$

$= \Sigma x^2 p(x) - (1.333)^2$

$= \frac{216}{81} - 1.778 = 2.667 - 1.778 = 0.889$

$\text{S.D}(X) = \sqrt{0.887} = 0.943$

(iii) Mean and variance by properties:

Mean of Binomial distribution $= n\,p = 4\times\frac{1}{3} = 1.333$

Variance of Binomial distribution $= n\,p\,q = 4\times\frac{1}{3}\times\frac{2}{3} = 0.889$

$\text{S.D}(X) = 0.943$

Note that both by calculation and properties, we have the same results.

Example 3.15

Is it possible to have a Binomial distribution with mean = 5 and S.D = 4?

Solution:

Given that $n\,p = 5$

Squaring both sides

$\sqrt{npq} = 4$

Or $n\,p\,q = 16$

$5q = 16$

$q = \frac{16}{5} = 3.2 > 1$

p or q should not be greater than one because $p + q = 1$

Hence it is not possible to have a Binomial distribution with mean 5 and S.D 4.

Example 3.16

A random variable X is binomially distributed with mean 38 and S.D 2.94. Find n and p.

Solution:

Given $n\,p = 38$ (i)

$\sqrt{npq} = 2.94$

Or $npq = 8.64$ (ii)

Putting value of $n\,p$ in equation (ii), we get

$38q = 8.64$

$$q = \frac{8.64}{38} = 0.23$$

$$\therefore p = 1 - q = 1 - 0.23 = 0.77$$

Put in equation (i), the value of p, to have

$$n(0.77) = 38$$

$$n = \frac{38}{0.77} = 49.351 \cong 50$$

Hence the required parameters are $[n = 49, p = 0.77]$

3.3.4 Binomial frequency distribution

If the Binomial probability distribution is multiplied by N, the number of Binomial experiments, i.e. $Np(x) = N\left[\overset{n}{\underset{x}{C}} p^x q^{n-x}\right]$, the resulting distribution is known as Binomial frequency distribution. It is used to find the expected frequency (E_f) of X successes in N Binomial experiments.

Example 3.17

Suppose five fair dice are rolled 96 times. Find the expected frequencies when the number 4, 5 or 6 is regarded as success.

Solution:

We have $N = 96$, $n = 5$, X: number of dice showing 4, 5 or 6.

$$X = 0, 1, 2, 3, 4, 5$$

$$p = P[\text{resulting 4,5 or 6 on a single die}) = \frac{3}{6} = \frac{1}{2}, \quad q = \frac{1}{2}$$

$$p(x) = \overset{5}{\underset{x}{C}} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}, \quad x = 0,1,2,3,4,5$$

Now expected frequencies are computed as follows:

X	$p(x) = \overset{5}{\underset{x}{C}} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$	$E_f = N\,p(x)$
0	$\overset{5}{\underset{0}{C}} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{5-0} = \frac{1}{32}$	$96\left(\frac{1}{32}\right) = 03$
1	$\overset{5}{\underset{1}{C}} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{5-1} = \frac{5}{32}$	$96\left(\frac{5}{32}\right) = 15$
2	$\overset{5}{\underset{2}{C}} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{5-2} = \frac{10}{32}$	$= 30$
3	$\overset{5}{\underset{3}{C}} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{5-3} = \frac{10}{32}$	$= 30$
4	$\overset{5}{\underset{4}{C}} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{5-4} = \frac{5}{32}$	$= 15$
5	$\overset{5}{\underset{5}{C}} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{5-5} = \frac{1}{32}$	$= 03$
Total	1	96

Example 3.18

Fit a Binomial distribution to the following data and compute expected frequencies:

X	0	1	2	3	4
f	30	62	46	10	2

Solution:

First we compute mean of the given frequency distribution

X	0	1	2	3	4	Total
f	30	62	46	10	2	150
fx	0	62	92	30	8	192

Given that $n = 4$, $N = 150$. Now for Binomial distribution $\mu = np$ but here μ is unknown so we replace μ with its estimator $\overline{X}$.

$$\overline{X} = \frac{\sum fx}{\sum f} = \frac{192}{150} = 1.28$$

$$\therefore \overline{X} = np \quad \Rightarrow 1.28 = np$$

$$\text{Or } 1.28 = 4p \quad \Rightarrow p = \frac{1.28}{4} = 0.32 \qquad \therefore q = 0.68$$

The fitted Binomial distribution is:

$$p(x) = \begin{cases} \overset{4}{\underset{x}{C}}(0.32)^x (0.68)^{4-x} & , \; x = 0,1,2,3,4 \\ 0 & , \; otherwise \end{cases}$$

The expected frequencies are computed as:

X	$p(x) = \overset{4}{\underset{x}{C}}(0.32)^x (0.68)^{4-x}$	$E_f = N\,p(x)$
0	$\overset{4}{\underset{0}{C}}(0.32)^0 (0.68)^{4-0} = 0.213814$	$150 \times 0.213814 = 32$
1	$\overset{4}{\underset{1}{C}}(0.32)^1 (0.68)^{4-1} = 0.402478$	$150 \times 0.402478 = 60$
2	$\overset{4}{\underset{2}{C}}(0.32)^2 (0.68)^{4-2} = 0.284099$	$= 43$
3	$\overset{4}{\underset{3}{C}}(0.32)^3 (0.68)^{4-3} = 0.089129$	$= 13$
4	$\overset{4}{\underset{4}{C}}(0.32)^4 (0.68)^{4-4} = 0.010486$	$= 02$
Total	1	150

Example 3.19

Suppose that seven coins are tossed and the number of heads noted. This experiment is repeated 128 times (i) fit a Binomial distribution under the hypothesis that the coins are unbiased. (ii) compute the theoretical frequencies. (iii) find its mean and standard deviation.

Solution:

(i) We have the following information

$n = 7$, $X = 0, 1, 2, 3, 4, 5, 6, 7$ $\quad N = 128$, $p = \frac{1}{2}$ (coins are unbiased),

$q = \frac{1}{2}$, therefore, the fitted Binomial distribution is

$$p(x)=\begin{cases}\overset{7}{\underset{x}{C}}(1/2)^x(1/2)^{7-x}, & x=0,1,\ldots,7\\ 0, & otherwise\end{cases}$$

(ii)

X	$p(x)=\overset{7}{\underset{x}{C}}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{7-x}$	$E_f = N\,p(x)$
0	$\overset{7}{\underset{0}{C}}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{7-0}=\frac{1}{128}$	$128\times\frac{1}{128}=01$
1	$\overset{7}{\underset{1}{C}}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{7-1}=\frac{7}{128}$	$128\times\frac{7}{128}=07$
2	$\overset{7}{\underset{2}{C}}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{7-2}=\frac{21}{128}$	$=21$
3	$\overset{7}{\underset{3}{C}}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{7-3}=\frac{35}{128}$	$=35$
4	$\overset{7}{\underset{4}{C}}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{7-4}=\frac{35}{128}$	$=35$
5	$\overset{7}{\underset{5}{C}}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{7-5}=\frac{21}{128}$	$=21$
6	$\overset{7}{\underset{6}{C}}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^{7-6}=\frac{7}{128}$	$=07$
7	$\overset{7}{\underset{7}{C}}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^{7-7}=\frac{1}{128}$	$=01$
Total	1	128

(iii) Mean of Binomial distribution= $E(X)=np=7\left(\frac{1}{2}\right)=3.5$

Standard deviation of Binomial distribution $=\sigma_x = \sqrt{npq}=\sqrt{7\times\frac{1}{2}\times\frac{1}{2}}$

$=\sqrt{\frac{7}{4}}=1.32$

3.4 Hypergeometric experiment

The Binomial distribution is based on the assumption that the successive trials are independent and the probability of success remains unchanged from trial to trial. These assumptions hold only for sampling with replacement from an infinite population but there are experiments in which the conditions of independence is violated and the probability of success does not remain the same/constant for all trails e.g. if sampling is without replacement from a finite population, the probability will change from trail to trail and the successive trails will be dependent. Such experiments are called hypergeometric experiments having the following four properties:

i. Each trail may have two possible results, success and failure.

ii. The probability of success changes on each trail.

iii. The successive trails are dependent.

iv. The experiment is repeated a fixed number of time, say n.

The random variable X denoting the number of successes in a hypergeometric experiment is called hypergeometric random variable and its probability distribution is called hypergeometric probability distribution.

3.4.1 Hypergeometric probability distribution

If values of a hypergeometric random variable along with their associated probabilities are shown in tabular form, then it is called hypergeometric probability distribution i.e.

X	0	1	2	…	n	Total
$p(x)$	$\frac{\binom{K}{0}\binom{N-K}{n-0}}{\binom{N}{n}}$	$\frac{\binom{K}{1}\binom{N-K}{n-1}}{\binom{N}{n}}$	$\frac{\binom{K}{2}\binom{N-K}{n-2}}{\binom{N}{n}}$	…	$\frac{\binom{K}{n}\binom{N-K}{n-n}}{\binom{N}{n}}$	1

3.4.2 Hypergeometric probability mass function

The probability mass function of hypergeometric random variable X is given as:

$$p(x)=\begin{cases}\dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, & x=0,1,2,\ldots,\min(n,K)\\ 0 & otherwise\end{cases}$$

This distribution has three parameters N, K, n, whereas

- N denotes total number of items in the population.
- K denotes the number of successes in the population.
- n denotes the number of elements in the sample drawn at random.
- X denotes the number of successes in the sample.

Example 3.20

A committee of size three is selected from 4 men and 2 women. Find the probability distribution for the number of women in the committee.

Solution: we have

$$\boxed{\begin{array}{l}4M\\2W\end{array}}\longleftarrow K$$

$$N\longrightarrow 6 \text{ persons}$$

As it has not been mentioned that an object selected is returned to the population before next draw so, we consider it without replacement sampling case and hence hypergeometric probability distribution is required.

Now $n=3$ (size of committee)

K=2 (number of successes)

$X=0, 1, 2$ (number of women in the committee)

Hence the fitted hypergeometric probability function is:

$$p(x)=\begin{cases}\dfrac{\binom{2}{x}\binom{6-2}{3-x}}{\binom{6}{3}}, & x=0,1,2\\ 0, & otherwise\end{cases}$$

The hypergeometric probability distribution is:

X	$p(x)=\dfrac{\binom{2}{x}\binom{4}{3-x}}{\binom{6}{3}}$
0	$\dfrac{\binom{2}{0}\binom{4}{3-0}}{\binom{6}{3}}=\dfrac{4}{20}$
1	$\dfrac{\binom{2}{1}\binom{4}{3-1}}{\binom{6}{3}}=\dfrac{12}{20}$
2	$\dfrac{\binom{2}{2}\binom{4}{3-2}}{\binom{6}{3}}=\dfrac{4}{20}$
Total	1

Example 3.21

In an international recitation competition of the Holy Quran, a panel of 11 judges is formed to judge the best recitation. Two recitations R_A and R_B were considered to be the best where the opinion of judges got divided. Six judges were in favour of R_A whereas five in favour of R_B. A random sample of five judges was drawn from the panel. Find the probability that out of five judges three are in favour of recitation R_A.

Solution:

In the given problem

$6R_A$, $5R_B$ $\longleftarrow K$

$N \longrightarrow 11$

$n = 5, \quad X = 0, 1, 2, 3, 4, 5 \ \{\because X=0,1,2,\ldots\min(n,k)\}$

$$p(x) = \begin{cases} \dfrac{\binom{6}{x}\binom{11-6}{5-x}}{\binom{N}{n}}, & x = 0,1,2,3,4,5 \\ 0 & otherwise \end{cases}$$

Now

$$P(X=3) = \frac{\binom{6}{3}\binom{5}{5-3}}{\binom{11}{5}} = \frac{200}{231}$$

3.4.3 Properties of hypergeometric probability distribution

- Mean of hypergeometric probability distribution by definition is given by

$$\text{Mean} = E(X) = \sum_x x\,p(x) = \sum_{x=0}^{n} x \frac{\binom{k}{x}\binom{N-k}{n-k}}{\binom{N}{n}}$$

$$= 0\frac{\binom{k}{0}\binom{N-k}{n-0}}{\binom{N}{n}} + \sum_{x=1}^{n} x \frac{\binom{k}{x}\binom{N-k}{n-k}}{\binom{N}{n}}$$

$$= \frac{1}{\binom{N}{n}} \sum_{x=1}^{n} x \frac{k!}{(k-x)!x!}\binom{N-k}{n-x}$$

$$= \frac{1}{\binom{N}{n}} \sum_{x=1}^{n} x \frac{k(k-1)!}{(k-x)!x(x-1)!}\binom{N-k}{n-x}$$

$$= \frac{k}{\binom{N}{n}} \sum_{x=1}^{n} \frac{(k-1)!}{(k-x)!(x-1)!}\binom{N-k}{n-x} = \frac{k}{\binom{N}{n}} \sum_{x=1}^{n} \binom{k-1}{x-1}\binom{N-k}{n-x}$$

Let $y = x-1 \Rightarrow x = y+1$

If $x = 1, y = 0$

If $x = n, y = n-1$

$$\therefore E(X) = \frac{k}{\binom{N}{n}} \sum_{y=0}^{n-1} \binom{k-1}{y}\binom{N-k}{n-y-1}$$

Apply the hypergeometric identity $\sum_{r=0}^{k}\binom{m}{r}\binom{n}{k-r} = \binom{m+n}{k}$

$$E(X)=\frac{k}{\binom{N}{n}}\binom{N-1}{n-1}$$

$$=\frac{K}{\frac{N!}{(N-n)!n!}}\frac{(N-1)!}{(N-n)!(n-1)!}=\frac{Kn(n-1)!}{N(N-1)!}\frac{(N-1)!}{(n-1)!}$$

$$=n\frac{K}{N}$$

- Variance of hypergeometric probability distribution

By definition

Variance $= V(X) = E(X^2) - [E(X)]^2$ (A)

As $E(X)=\frac{nK}{N}$ ....(1)

Now $E(X^2)=\sum_x x^2 p(x)$

$$=\sum_{x=0}^{n}x^2\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}=(0)^2\frac{\binom{k}{0}\binom{N-k}{n-0}}{\binom{N}{n}}+\sum_{x=1}^{n}x^2\frac{\binom{k}{x}\binom{N-k}{N-x}}{\binom{N}{n}}$$

$$=\sum_{x=1}^{n}\left[x(x-1)+x\right]\binom{k}{x}\binom{N-k}{n-x}\qquad \because x^2=x(x-1)+x$$

$$=\sum_{x=1}^{n}x(x-1)\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}+\sum_{x=1}^{n}x\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{x}}=\left[1(1-1)\frac{\binom{k}{1}\binom{N-k}{n-0}}{\binom{N}{n}}\right]+\sum_{x=2}^{n}x(x-1)\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}+\frac{nk}{N}$$

$$E(X^2)=\left[\frac{1}{\binom{N}{n}}\sum_{x=2}^{n}x(x-1)\frac{k!}{(k-x)!x!}\binom{N-k}{n-x}\right]+\left(\frac{nk}{N}\right)$$

$$=\frac{1}{\binom{N}{x}}\sum_{x=2}^{n}x(x-1)\frac{k(k-1)(k-2)!}{(k-x)!\,x(x-1)(x-2)!}\binom{N-k}{n-x}+\frac{nk}{N}$$

$$=\frac{k(k-1)}{\binom{N}{n}}\sum_{x=2}^{n}\frac{(k-2)!}{(k-x)!(x-2)!}\binom{N-k}{n-x}+\frac{nk}{N}$$

$$=\frac{k(k-1)}{\binom{N}{n}}\sum_{x=2}^{n}\binom{k-2}{x-2}\binom{N-k}{n-x}+\frac{nk}{N}$$

Let $y=x-2,\quad x=y+2$

If $x=2,\quad y=0$

If $x=n,\quad y=n-2$, therefore,

$$E(X^2)=\frac{k(k-1)}{\binom{N}{n}}\sum_{y=0}^{n-2}\binom{k-2}{y}\binom{N-k}{n-y-2}+\frac{nk}{N}$$

Using the hypergeometric identity $\sum_{r=0}^{k}\binom{m}{r}\binom{n}{k-r}=\binom{m+n}{k}$

$$E(X^2)=\frac{k(k-1)}{\binom{N}{n}}\binom{N-2}{n-2}+\frac{nk}{N}$$

$$=\frac{k(k-1)}{\frac{N!}{(N-n)!n!}}\frac{(N-2)!}{(N-n)!(n-2)!}+\frac{nk}{N}$$

$$=\frac{k(k-1)n(n-1)(n-2)!}{N(N-1)(N-2)!}\frac{(N-2)!}{(n-2)!}+\frac{nk}{N}$$

$$=\frac{n(n-1)k(k-1)}{N(N-1)}+\frac{nk}{N}\quad\text{.................(2)}$$

Putting equation (1) or equation (2) in equation (A)

$$V(X)=\frac{n(n-1)k(k-1)}{N(N-1)}+\frac{nk}{N}-\frac{n^2k^2}{N^2}$$

$$=\frac{nk}{N}\left[\frac{(n-1)(k-1)}{N-1}+1-\frac{nk}{N}\right]$$

$$V(X)=\frac{nk}{N}\left[\frac{N(n-1)(k-1)+N(N-1)-nk(N-1)}{N(N-1)}\right]$$

$$=\frac{nk}{N}\left[\frac{nNk-nN-Nk+N+N^2-N-nkN+nk}{N(N-1)}\right]$$

$$=\frac{nk}{N}\left[\frac{N^2-Nk+nk-nN}{N(N-1)}\right]$$

$$=\frac{nk}{N}\left[\frac{N(N-k)-n(N-k)}{N(N-1)}\right]$$

$$=\frac{nk}{N}\left[\frac{(N-k)(N-n)}{N(N-1)}\right]$$

$$=n\frac{k}{N}\frac{N-k}{N}\frac{N-n}{N-1}$$

$$=npq\frac{N-n}{N-1}\qquad\left(\because p=\frac{k}{N}, q=1-p=1-\frac{k}{N}=\frac{N-k}{N}\right)$$

- Standard deviation of hypergeometric probability distribution is

$$\text{S.D}(X)=\sqrt{npq\frac{N-n}{N-1}}$$

Example 3.22

A bowl contains 4 green and 7 red balls. A sample of 5 balls is selected from the bowl without replacement. Find the probability distribution for the number of green balls. Compute the mean and variance of this probability distribution and compare them with the theoretical mean and variance.

Solution:

The bowl contains

4 *G*	⟵ k
7 *R*	

$N \longrightarrow$ 11 balls

Here $n = 5$

X: number of green balls selected

$X = 0, 1, 2, 3, 4.$ [5 is not possible as there are only 4 green balls.].

$$\therefore p(x)=\begin{cases}\dfrac{\binom{4}{x}\binom{11-4}{5-x}}{\binom{11}{5}}, & x=0,1,2,3,4\\ 0 & otherwise\end{cases}$$

Now hypergeometric probability distribution of X is given as

X	$p(x)=\frac{\binom{4}{x}\binom{7}{5-x}}{\binom{11}{5}}$	$Xp(x)$	$X^2p(x)$
0	$\frac{\binom{4}{0}\binom{7}{5-0}}{\binom{11}{5}}=\frac{21}{462}$	0	0
1	$\frac{\binom{4}{1}\binom{7}{5-1}}{\binom{11}{5}}=\frac{140}{462}$	$\frac{140}{462}$	$\frac{140}{462}$
2	$\frac{\binom{4}{2}\binom{7}{5-2}}{\binom{11}{5}}=\frac{210}{462}$	$\frac{420}{462}$	$\frac{840}{462}$
3	$\frac{\binom{4}{3}\binom{7}{5-3}}{\binom{11}{5}}=\frac{84}{462}$	$\frac{252}{462}$	$\frac{756}{462}$
4	$\frac{\binom{4}{4}\binom{7}{5-4}}{\binom{11}{5}}=\frac{7}{462}$	$\frac{28}{462}$	$\frac{112}{462}$
Total	1	$\frac{840}{462}$	$\frac{1848}{462}$

Mean $= E(X) = \sum_{x=0}^{4} x\,p(x) = \frac{840}{462} = 1.8182$

Variance $= V(X) = E(X^2) - [E(X)]^2$

$$= \sum_{x=0}^{4} x^2\,p(x) - (1.8182)^2$$

$$= \frac{1848}{462} - (1.8182)^2 = 4 - 3.3059 = 0.694$$

- Verification

Mean of hypergeometric distribution $= E(X) = \frac{nk}{N} = \frac{5(4)}{11} = \frac{20}{11} = 1.8182$

Variance of hypergeometric distribution $V(X) = \frac{nk}{N}\,\frac{N-k}{N}\,\frac{N-n}{N-1}$

$$= 1.8182\frac{(11-4)}{11}\left(\frac{11-5}{11-1}\right)$$

$$= 0.694$$

Key points

- A probability function defined as

$$p(x)=\begin{cases}\frac{1}{N} & , \quad x=1,2,\cdots,N \\ 0 & , \quad otherwise\end{cases}$$

is called discrete uniform probability function

- A random experiment which has two possible outcomes, classified as success and failure, is called a Bernoulli trial.
- If a Bernoulli trail is repeated a fixed number of times, say n, then such an experiment is called Binomial experiment.
- The Binomial probability mass function is given by the formula:

$$p(x)=\begin{cases}{}^{n}C_{x}\, p^x q^{n-x} & , \quad x=0,1,2,...,n \\ 0 , & \quad otherwise\end{cases}$$

- Mean of Binomial distribution is np
- Binomial distribution has two parameters i.e. n and p. Its range is from 0 to n
- Variance of Binomial distribution is npq
- The hypergeometric probability distribution is

$$p(x)=\begin{cases}\dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} & , \quad x=0,1,2,...,\min(n,K) \\ 0 & \quad otherwise\end{cases}$$

- Hypergeometric distribution has three parameters and its range is from 0 to min(n,k)

Exercise

3.1 Write T for true and F for false in the following statement.

i) If

X	0	1
$p(x)$	$\frac{1}{2}$	$\frac{1}{2}$

then X is not uniform random variable.

ii. Bernoulli experiment will always give only two results.
iii. Tossing a fair coin a large number of times is a Binomial experiment
iv. Binomial probability distribution has only two parameters i.e. n and p.
v. A Binomial random variable is discrete.
vi. Mean and variance of binomial distribution are equal.
vii. If each digit from 0 to 9 has the same probability i.e. $\frac{1}{10}$, they are called random digits.
viii. The Binomial distribution will be symmetrical if $p=q=\frac{1}{2}$
ix. The hypergeometric random variable cannot assume the negative values.
x. Hypergeometric probability distribution has three parameters i.e. N, K, and n.

3.2 Fill in the blanks.

(i) Bernoulli trail has ________ possible outcomes.
(ii) Bernoulli distribution has ________ parameter.
(iii) Range of Binomial random variable is from 0 to ________.
(iv) Parameters of Binomial distribution are ________.
(v) Mean of Binomial distribution is ________ and variance is ________.
(vi) The number of trails in hypergeometric distribution is ________.

(vii) Bernoulli random variable takes only two values i.e. ________.

(viii) Pairs of random digits are called ________.

(ix) The successive trails in a Binomial experiment are ________.

(x) The probability of success ________ from trail to trail in a hypergeometric experiment.

3.3 Choose the correct answer.

i) The Binominal distribution was introduced by

a) Simon Denis Poisson b) Jacob Bernoulli
c) Abraham De Moirés d) R.A. Fisher

ii) In a Bernoulli trail the probability of success is denoted by

a) q b) $1-p$
c) p d) $1-q$

iii) The mean of the Binomial distribution is

a) $\sqrt{np}$ b) npq
c) np d) $\sqrt{npq}$

iv) If X has a Binomial distribution with $p=\frac{2}{3}$ and $n = 9$, then its mean will be equal to

a) 2 b) 3
c) 5 d) 6

v) Hypergeometric probability distribution has parameters

a) 2 b) 3
c) 4 d) 5

vi) The Binomial distribution is the

a) equal probability distribution
b) individual probability distribution
c) discrete probability distribution
d) continuous probability distribution

vii) If $n = 60$, $p = 0.7$ for a Binomial distribution then its standard deviation is equal to

a) 42 b) 6.48
c) 18 d) 3.55

viii) If a fair coin is tossed once, the value of p will be

a) $\frac{1}{2}$ b) $\frac{1}{3}$
c) $\frac{1}{4}$ d) $\frac{1}{1}$

ix) The number of parameters in a Binomial distribution are

a) one b) two
c) three d) four

x) The range of uniform distribution is equal to;

a) $-\infty$ to 0 b) $0,1,\ldots,N$
c) 0 to ∞ d) $1, 2, 3,\ldots,N$

3.4 Describe in brief the discrete uniform probability distribution.

3.5 What do you know about random digits, random numbers, random number generator and random number table?

3.6 Explain how you would select a random sample of 10 colleges form a list of 206 colleges by using a random number table.

3.7 Define i) Bernoulli experiment ii) Bernoulli random variable iii) Binomial experiment iv) Binomial random variable.

3.8 Define Bernoulli probability distribution. Find its mean, variance and standard deviation.

3.9 i) What is a Binomial experiment and what are its conditions/properties?

ii) Find the mean and variance of the Binomial distribution.

3.10 Suppose X has a Binomial probability distribution with $p = 0.4$ and $n = 3$ Find i) $P(X = -1)$, ii) $P(X = 2)$, iii) $P(X = 1.5)$, iv) $P(X \leq 2)$, v) $P(X \geq 2)$.

3.11 If $n = 5$ and $p = \frac{3}{8}$, Find the complete Binomial probability distribution.

3.12 A fair coin is tossed six times. What is the probability that i) Less than four heads occur ii) 2 or more heads occur.

3.13 If 40% of a consignment of eggs are bad. Estimate the chance that 5 eggs chosen at random contains i) None, ii) one and iii) at least one bad egg.

3.14 A certain drug causes kidney damage 1% of patients. Suppose the drug is to be tested on 50 patients. Find the probability i) none of the patients will experience kidney damage and ii) one or more of the patients will experience kidney damage.

3.15 If 60% of the students in a large college are day-scholar. Find the probability that in a random sample of 12 students from that college exactly 7 will be day-scholar.

3.16 If the probability of hitting a target is $\frac{1}{5}$ and ten shots are fired independently. What is the probability of the target being hit at least twice?

3.17 If the probability that a person entering a utility store will purchase sugar is 0.90. Compute the probability that exactly one person among the next five entering the utility store will purchase sugar?

3.18 a) Define binomial distribution and explain how it arises in practice?

b) Derive its mean and S.D.

3.19 Let the probability of a defective bolt is 0.10, Find a) the mean and b) the standard deviation far the distribution of defective bolts in a total of 400.

3.20 The mean and standard derivation of a Binomial distribution are 3 and 1.5 respectively. Find the two parameters of the Binomial distribution i.e. n and p.

3.21 In a Binomial distribution the mean and variance were found to be 12.38 and 8.64. Find n and p.

3.22 Is it possible to have a Binomial distribution with mean = 5 and standard deviation = 3?

3.23 a) If X is a Binomial random variable with $n = 10$ and $p = 0.6$ then find $E(3X - 2)$.

b) If a Binomial distribution has mean=3 and variance= 2. Find $P(X \leq 5)$

3.24 Find the mean, variance and S.D for the following binomial probability distribution:

X	0	1	2	4	5	6
$p(x)$	0.01521	0.08649	0.21785	0.30833	0.10973	0.02191

Also compare these results with the mean, variance and S.D of the Binomial distribution for $p = 0.5$

3.25 a) Define Binomial frequency distribution.

b) Four dice are thrown and the number of sixes in each throw is recorded, this is repeated 108 times. Find the expected frequencies of 0, 1,2,3,4 sixes.

3.26 Fit a Binomial distribution to the following table.

X	0	1	2	3	4	5	6	7	8	9	10
f	0	1	3	8	16	28	18	13	9	4	0