UNIT 11 DISCRETE UNIFORM AND HYPERGEOMETRIC DISTRIBUTIONS

Structure

11.1 INTRODUCTION

In the previous two units, we have discussed binomial distribution and its limiting form i.e. Poisson distribution. Continuing the study of discrete distributions, in the present unit, two more discrete distributions – Discrete uniform and Hypergeometric distributions are discussed.

Discrete uniform distribution is applicable to those experiments where the different values of random variable are equally likely. If the population is finite and the sampling is done without replacement i.e. if the events are random but not independent, then we use Hypergemetric distribution.

In this unit, discrete uniform distribution and hypergeometric distribution are discussed in Secs. 11.2 and 11.3, respectively. We shall be discussing their properties and applications also in these sections.

Objectives

After studing this unit, you should be able to:

- define the discrete uniform and hypergeometric distributions;
- compute their means and variances;
- compute probabilities of events associated with these distributions; and
- know the situations where these distributions are applicable.

11.2 DISCRETE UNIFORM DISTRIBUTION

Discrete uniform distribution can be conceived in practice if under the given experimental conditions, the different values of the random variable are equally likely. For example, the number on an unbiased die when thrown may be 1 or 2 or 3 or 4 or 5 or 6. These values of random variable, “the number on an unbiased die when thrown” are equally likely and for such an experiment, the discrete uniform distribution is appropriate.

Definition: A random variable X is said to have a discrete uniform (rectangular) distribution if it takes any positive integer value from 1 to n, and its probability mass function is given by

$$P[X = x] = \begin{cases} \frac{1}{n} & \text{for } x = 1, 2, ..., n \\ 0, & \text{otherwise.} \end{cases}$$

where n is called the parameter of the distribution.

For example, the random variable X, "the number on the unbiased die when thrown", takes on the positive integer values from 1 to 6 follows discrete uniform distribution having the probability mass function.

$$P[X = x] = \begin{cases} \frac{1}{6}, & \text{for } x = 1, 2, 3, 4, 5, 6. \\ 0, & \text{otherwise.} \end{cases}$$

Mean and Variance of the Distribution

$$\text{Mean} = E(X) = \sum_{x=1}^{n} x\,p(x) = \sum_{x=1}^{n} x.\left(\frac{1}{n}\right) = \frac{1}{n}\sum_{x=1}^{n} x$$

$$= \frac{1}{n}[1+2+3+...+n]$$

$$= \frac{1}{n}.\frac{n(n+1)}{2} \left[\begin{array}{l} \because \text{sum of first n natural numbers} = \frac{n(n+1)}{2} \\ \text{(see Unit 3 of Course MST} - 001) \end{array}\right]$$

$$= \frac{n+1}{2}.$$

$$\text{Variance} = E(X^2) - [E(X)]^2 \qquad [\because \mu_2 = \mu_2' - (\mu_1')^2]$$

where

$$E(X) = \frac{n+1}{2} \qquad \text{[Obtained above]}$$

$$E(X^2) = \sum_{x=1}^{n} x^2.p(x)$$

and $E(X^2) = \sum_{x=1}^{n} x^2.\frac{1}{n}$

$$= \frac{1}{n}[1^2 + 2^2 + 3^2 + ... + n^2]$$

$$= \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] \left[\begin{array}{l} \because \text{sum of squares of first n} \\ \text{natural numbers} = \frac{n(n+1)(2n+1)}{6} \\ \text{(see Unit 3 of Course MST} - 001) \end{array}\right]$$

$$= \frac{(n+1)(2n+1)}{6}$$

$$\therefore \text{ Variance} = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)}{12}\left[2(2n+1) - 3(n+1)\right]$$

$$= \frac{n+1}{12}\left[4n+2-3n-3\right] = \frac{(n+1)}{12}(n-1) = \frac{n^2-1}{12}$$

Example 1: Find the mean and variance of a number on an unbiased die when thrown.

Solution: Let X be the number on an unbiased die when thrown,

$\therefore$ X can take the values 1, 2, 3, 4, 5, 6 with

$$P[X = x] = \frac{1}{6}; x = 1, 2, 3, 4, 5, 6.$$

Hence, by uniform distribution, we have

$$\text{Mean} = \frac{n+1}{2} = \frac{6+1}{2} = \frac{7}{2}, \text{ and}$$

$$\text{Variance} = \frac{n^2-1}{12} = \frac{(6)^2-1}{12} = \frac{35}{12}.$$

Uniform Frequency Distribution

If an experiment, satisfying the requirements of discrete uniform distribution, is repeated N times, then expected frequency of a value of random variable is given by

$$f(x) = N.P[X = x]; x = 1, 2, ..., n$$

$$= N.\frac{1}{n}; x = 1, 2, 3, ..., n.$$

Example 2: If an unbiased die is thrown 120 times, find the expected frequency of appearing 1, 2, 3, 4, 5, 6 on the die.

Solution: Let X be the uniform discrete random variable, "the number on the unbiased die when thrown".

$$\therefore P[X = x] = \frac{1}{6}; x = 1, 2, ..., 6$$

Hence, the expected frequencies of the value of random variable are given as computed in the following table:

X	$P[X = x]$	Expected/Theoretical frequencies $f(x) = N.P[X = x] = 120.P[X = x]$
1	$\frac{1}{6}$	$120 \times \frac{1}{6} = 20$
2	$\frac{1}{6}$	$120 \times \frac{1}{6} = 20$
3	$\frac{1}{6}$	$120 \times \frac{1}{6} = 20$
4	$\frac{1}{6}$	$120 \times \frac{1}{6} = 20$
5	$\frac{1}{6}$	$120 \times \frac{1}{6} = 20$
6	$\frac{1}{6}$	$120 \times \frac{1}{6} = 20$

Now, you can try the following exercise:

E1) Obtain the mean, variance of the discrete uniform distribution for the random variable, "the number on a ticket drawn randomly from an urn containing 10 tickets numbered from 1 to 10". Also obtain the expected frequencies if the experiment is repeated 150 times.

11.3 HYPERGEOMETRIC DISTRIBUTION

In the last section of this unit, we have studied discrete uniform probability distribution wherein the probability distribution is obtained for the possible outcomes in a single trial like drawing a ticket from an urn containing 10 tickets as mentioned in exercise **E1).** But, if there are more than one but finite trials with only two possible outcomes in each trial, we apply some other distribution. One such distribution which is applicable in such a situation is binomial distribution which you have studied in Unit 9. The binomial distribution deals with finite and independent trials, each of which has exactly two possible outcomes (Success or Failure) with constant probability of success in each trial. For example, if we again consider the example of drawing ticket randomly from an urn containing 10 tickets bearing numbers from 1 to 10. Then, the probability that the drawn ticket bears an odd number is $\frac{5}{10} = \frac{1}{2}$. If we replace the ticket back, then the probability of drawing a ticket bearing an odd number is again $\frac{5}{10} = \frac{1}{2}$. So, if we draw ticket again and again with replacement, trials become independent and probability of getting an odd number is same in each trial. Suppose, it is asked that what is the probability of getting 2 tickets bearing odd number in 3 draws then we apply binomial distribution as follows:

Let X be the number of times on odd number appears in 3 draws, then by binomial distribution,

$$P[X=2] = {}^{3}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{3-2} = (3)\left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = \frac{3}{8}.$$

But, if in the example discussed above, we do not replace the ticket after any draw the probability of getting an odd number gets changed in each trial and the trials remain no more independent and hence in this case binomial distribution is not applicable. Suppose, in this case also, we are interested in finding the probability of getting ticket bearing odd number twice in 3 draws, then it is computed as follows:

Let A_i be the event that i^{th} ticket drawn bears odd number and $\bar{A}_i$ be the event that i^{th} ticket drawn does not bear odd number.

$\therefore$ Probability of getting ticket bearing odd number twice in 3 draws

$$= P\left[A_1 \cap A_2 \cap \bar{A}_3\right] + P\left[A_1 \cap \bar{A}_2 \cap A_3\right] + P\left[\bar{A}_1 \cap A_2 \cap A_3\right]$$

[As done in Unit 3 of this Course]

$$= P[A_1]P[A_2 \mid A_1]P\left[\bar{A}_3 \mid A_1 \cap A_2\right] + P[A_1]P\left[\bar{A}_2 \mid A_1\right]P\left[A_3 \mid A_1 \cap \bar{A}_2\right]$$
$$+ P\left[\bar{A}_1\right]P\left[A_2 \mid \bar{A}_1\right]P\left[A_3 \mid \bar{A}_1 \cap A_2\right]$$

[Multiplication theorem for dependent events (See Unit 3 of this Course)]

$$= \frac{5}{10}.\frac{4}{9}.\frac{5}{8} + \frac{5}{10}.\frac{5}{9}.\frac{4}{8} + \frac{5}{10}.\frac{5}{9}.\frac{4}{8}$$

$$= 3 \times \frac{5 \times 5 \times 4}{10 \times 9 \times 8}$$

This result can be written in the following form also:

$$= \frac{5 \times 4 \times 5 \times 3 \times 2}{2 \times 10 \times 9 \times 8} \qquad \text{[Multiplying and Dividing by 2]}$$

$$= \frac{5 \times 4}{2} \times 5 \times \frac{1}{\frac{10 \times 9 \times 8}{3 \times 2}} = {}^{5}C_2 \times {}^{5}C_1 \times \frac{1}{{}^{10}C_3} = \frac{{}^{5}C_2 \times {}^{5}C_1}{{}^{10}C_3}$$

In the above result, ${}^{5}C_2$ is representing the number of ways of selecting 2 out of 5 tickets bearing odd number, ${}^{5}C_1$ is representing the number of ways of selecting 1 out of 5 tickets bearing even number i.e. not bearing odd number, and ${}^{10}C_3$ is representing the number of ways of selecting 3 out of total 10 tickets.

Let us consider another similar example of a bag containing 20 balls out of which 5 are white and 15 are black. Suppose 10 balls are drawn at random one by one without replacement, then as discussed in the above example, the probability that in these 10 draws, there are 2 white and 8 black balls is

$\frac{{}^{5}C_2 \times {}^{15}C_8}{{}^{20}C_{10}}$.

Note: The result remains exactly same whether the items are drawn one by one without replacement or drawn at once.

Let us now generalize the above argument for N balls, of which M are white and $N-M$ are black. Of these, n balls are chosen at random without replacement. Let X be a random variable that denote the number of white balls drawn. Then, the probability of $X = x$ white balls among the n balls drawn is given by

$$P[X = x] = \frac{{}^{M}C_x . {}^{N-M}C_{n-x}}{{}^{N}C_n}$$

[For $x = 0, 1, 2, ..., n(n \le M)$ or $x = 0, 1, 2, ..., M(n > M)$]

The above probability function of discrete random variable X is called the Hypergeometric distribution.

Remark 1: We have a hypergeometric distribution under the following conditions:

i) There are finite number of dependent trials

ii) A single trial results in one of the two possible outcomes-Success or Failure

iii) Probability of success and hence that of failure is not same in each trial i.e. sampling is done without replacement

Remark 2: If number (n) of balls drawn is greater than the number (M) of white balls in the bag, then if $n \le M$, the number (x) of white balls drawn cannot be greater than n and if $n > M$, then number of white balls drawn cannot be greater than M. So, x can take the values upto n $(\text{if } n \le M)$ and M(if $n > M$) i.e. x can take the value upto n or M, whichever is less, i.e. $x =$ min {n, M}.

The discussion leads to the following definition

Definition: A random variable X is said to follow the hypergeometric distribution with parameters N, M and n if it assumes only non-negative integer values and its probability mass function is given by

$$P[X = x] = \begin{cases} \frac{{}^{M}C_x . {}^{N-M}C_{n-x}}{{}^{N}C_n} & \text{for } x = 0, 1, 2, ..., \min\{n, M\} \\ 0, & \text{otherwise} \end{cases}$$

where n, M, N are positive integers such that $n \le N$, $M \le N$.

Mean and Variance

$$\text{Mean} = E(X) = \sum_{x=0}^{n} x.p[X = x]$$

$$= \sum_{x=1}^{n} x. \frac{{}^{M}C_x . {}^{N-M}C_{n-x}}{{}^{N}C_n}$$

$$= \sum_{x=1}^{n} x.\frac{M}{x}.\frac{{}^{M-1}C_{x-1}.{}^{N-M}C_{n-x}}{{}^{N}C_{n}}$$

$$= \frac{M}{{}^{N}C_{n}}\sum_{x=1}^{n}\left({}^{M-1}C_{x-1}.{}^{N-M}C_{n-x}\right)$$

$$= \frac{M}{{}^{N}C_{n}}\left[{}^{M-1}C_{0}.{}^{N-M}C_{n-1} + {}^{M-1}C_{1}.{}^{N-M}C_{n-2} + \ldots + {}^{M-1}C_{n-1}.{}^{N-M}C_{0}\right]$$

$$= \frac{M}{{}^{N}C_{n}}\left({}^{N-1}C_{n-1}\right)$$

[This result is obtained using properties of binomial coefficients and involves lot of calculations and hence its derivation may be skipped. It may be noticed that in this result the left upper suffix and also the right lower suffix is the sum of the corresponding suffices of the binomial coefficients involved in each product term. However, the result used in the above expression is enrectangled below for the interesting learners.]

We know that

$$(1+x)^{m+n} = (1+x)^{m}.(1+x)^{n} \qquad \text{[By the method of indices]}$$

Expanding using binomial theorem as explained in Unit 9 of this course, we have

$${}^{m+n}C_{0}.\,x^{m+n} + {}^{m+n}C_{1}.x^{m+n-1} + {}^{m+n}C_{2}.x^{m+n-2} + \ldots + {}^{m+n}C_{m+n}$$

$$= \left({}^{m}C_{0}x^{m} + {}^{m}C_{1}\,x^{m-1} + {}^{m}C_{2}\,x^{m-2} + \ldots + {}^{m}C_{m}\right)$$

$$.\left({}^{n}C_{0}x^{n} + {}^{n}C_{1}\,x^{n-1} + {}^{n}C_{2}\,x^{n-2} + \ldots + {}^{n}C_{n}\right)$$

Comparing coefficients of X^{m+n-r}, we have

$${}^{m+n}C_{r} = \left({}^{m}C_{0}.{}^{n}C_{r} + {}^{m}C_{1}.{}^{n}C_{r-1} + \ldots + {}^{m}C_{r}.{}^{n}C_{0}\right)$$

$$= \frac{M\lfloor n \lfloor N-n}{\lfloor N}.\frac{\lfloor N-1}{\lfloor N-n \lfloor n-1}$$

$$= \frac{M.n\lfloor n-1}{N.\lfloor N-1}.\frac{\lfloor N-1}{\lfloor n-1} = \frac{nM}{N}.$$

$$E\left(X^{2}\right) = E\left[X\left(X-1\right)+X\right]$$

$$= E\left[X\left(X-1\right)\right] + E\left(X\right)$$

$$= \left[\sum_{x=0}^{n} x\left(x-1\right).\frac{{}^{M}C_{x}.{}^{N-M}C_{n-x}}{{}^{N}C_{n}}\right] + \left(\frac{nM}{N}\right)$$

$$= \sum_{x=0}^{n}\left[x\left(x-1\right).\frac{M}{x}.\frac{M-1}{x-1}.\frac{{}^{M-2}C_{x-2}.{}^{N-M}C_{n-x}}{{}^{N}C_{n}}\right] + \left(\frac{nM}{N}\right)$$

$$= \frac{M(M-1)}{{}^{N}C_n}\left[\sum_{x=0}^{n}\left({}^{M-2}C_{x-2}.{}^{N-M}C_{n-x}\right)\right] + \left(\frac{nM}{N}\right)$$

$$= \frac{M(M-1)}{{}^{N}C_n}\left({}^{N-2}C_{n-2}\right) + \left(\frac{nM}{N}\right)$$

[The result in the first term has been obtained using a property of binomial coefficients as done above for finding E(X).]

$$= \frac{M(M-1)\lfloor N-n \lfloor n}{\lfloor N}.\frac{\lfloor N-2}{\lfloor n-2 \lfloor N-n} + \frac{nM}{N}$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N}$$

Thus,

$$V(X) = E(X^2) - [E(X)]^2 = \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} - \left(\frac{nM}{N}\right)^2$$

$$= \frac{NM(N-M)(N-n)}{N^2(N-1)} \qquad \text{[On simplification]}$$

Example 2: A jury of 5 members is drawn at random from a voters' list of 100 persons, out of which 60 are non-graduates and 40 are graduates. What is the probability that the jury will consist of 3 graduates?

Solution: The computation of the actual probability is hypergeometric, which is shown as follows:

P **[**2 non-graduates and 3 graduates**]** $= \frac{{}^{60}C_2.{}^{40}C_3}{{}^{100}C_5}$

$$= \frac{60\times59\times40\times39\times38\times5\times4\times3\times2}{2\times6\times100\times99\times98\times97\times96}$$

$$= 0.2323$$

Example 3: Let us suppose that in a lake there are N fish. A catch of 500 fish (all at the same time) is made and these fish are returned alive into the lake after making each with a red spot. After two days, assuming that during this time these 'marked' fish have been distributed themselves 'at random' in the lake and there is no change in the total number of fish, a fresh catch of 400 fish (again, all at once) is made. What is the probability that of these 400 fish, 100 will be having red spots.

Solution: The computation of the probability is hypergeometric and is shown as follows: As marked fish in the lake are 500 and other are $N-500$,

$\therefore$ P[100 marked fish and 300 others] = $\frac{{}^{500}C_{100}.{}^{N-500}C_{300}}{{}^{N}C_{400}}$.

We cannot numerically evaluate this if N is not given. Though N can be estimated using method of Maximum likelihood estimation which you will read in Unit 2 of MST-004 We are not going to estimate it. You may try it as an exercise after reading Unit 2 of MST-004.

Here, let us take an assumed value of N say 5000.

Then,

$$P[X = 100] = \frac{{}^{500}C_{100} \cdot {}^{4500}C_{300}}{{}^{5000}C_{400}}$$

You will agree that the exact computation of this probability is complicated. Such problem is normally there with the use of hypergeometric distribution, especially, if N and M are large. However, if n is small compared to N i.e. if n is such that $\frac{n}{N} < 0.05$, say then there is not much difference between sampling with and without replacement and hence in such cases, the probability obtained by binomial distribution comes out to be approximately equal to that obtained using hypergeometric distribution.

You may now try the following exercise.

E2) A lot of 25 units contains 10 defective units. An engineer inspects 2 randomly selected units from the lot. He/She accepts the lot if both the units are found in good condition, otherwise all the remaining units are inspected. Find the probability that the lot is accepted without further inspection.

We now conclude this unit by giving a summary of what we have covered in it.

11.4 SUMMARY

The following main points have been covered in this unit:

1) A random variable X is said to have a **discrete uniform (rectangular)** distribution if it takes any positive integer value from 1 to n, and its probability mass function is given by

$$P[X = x] = \begin{cases} \frac{1}{n} & \text{for } x = 1, 2, \ldots, n \\ 0, & \text{otherwise.} \end{cases}$$

where n is called the parameter of the distribution.

2) For **discrete uniform** distribution, **mean** = $\frac{n+1}{2}$ and **variance** = $\frac{n^2 - 1}{12}$.

3) A random variable X is said to follow the **hypergeometic distribution** with parameters N, M and n if it assumes only non-negative integer values and its probability mass function is given by

$$P(X = x) = \begin{cases} \frac{{}^{M}C_{x} \cdot {}^{N-M}C_{n-x}}{{}^{N}C_{n}} & \text{for } x = 0, 1, 2, \ldots, \min\{n, M\} \\ 0, & \text{otherwise} \end{cases}$$

where n, M, N are positive integers such that $n \le N$, $M \le N$.

4) For **hypergeometric** distribution, **mean** $= \frac{nM}{N}$ and

variance $= \frac{NM(N-M)(N-n)}{N^2(N-1)}$.

11.5 SOLUTIONS/ANSWERS

E1) Let X be the number on the ticket drawn randomly from an urn containing tickets numbered from 1 to 10.

$\therefore$ X is a discrete uniform random variable having the values

1, 2, 3, 4, …, 10 with probability of each of these values equal to $\frac{1}{10}$.

Thus, the expected frequencies for the values of X are obtained as in the following table:

X	$P(X=x)$	Expected/Theoretical frequency $f(x) = N.P[X = x]$ $= 150.P[X = x]$
1	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
2	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
3	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
4	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
5	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
6	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
7	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
8	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
9	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$
10	$\frac{1}{10}$	$150 \times \frac{1}{10} = 15$

E2) Here N = 25, M = 10 and n = 2.

The desired probability = $P\left[\begin{array}{l}\text{none of the 2 randomly selected}\\ \text{units is found defective}\end{array}\right]$

$$= \frac{{}^{10}C_0 . {}^{25-10}C_2}{{}^{25}C_2} = \frac{(1).{}^{15}C_2}{{}^{25}C_2} = \frac{15 \times 14}{25 \times 24} = 0.35.$$